# SECURITY ENHANCEMENT AND PRIVACY PROTECTION FOR BIOMETRIC SYSTEMS

zur Erlangung des akademischen Grades

*Doktor der Naturwissenschaften (Dr. rer.nat.)*

vorgelegte Dissertation von

JASCHA KOLBERG, M.Sc.

geboren in Duisburg

1. Gutachten: Prof. Dr. Ralf Dörner - HOCHSCHULE RHEINMAIN
2. Gutachten: Prof. Dr. Raymond Veldhuis - UNIVERSITY OF TWENTE
3. Gutachten: Prof. Dr. Andreas Uhl - UNIVERSITY OF SALZBURG

Tag der Einreichung: 01.04.2021
Tag der Prüfung: 08.07.2021

Promotionszentrum Angewandte Informatik
HAW Hessen

PARTNER:
Frankfurt University of Applied Sciences
Hochschule Darmstadt
Hochschule Fulda
Hochschule RheinMain

Darmstadt, Juli 2021

SUPERVISORS:
Prof. Dr. Christoph Busch - HOCHSCHULE DARMSTADT
Prof. Dr. Andreas Heinemann - HOCHSCHULE DARMSTADT
Prof. Dr. Marta Gomez-Barrero - HOCHSCHULE ANSBACH

# ABSTRACT

Biometric recognition systems are part of our daily life. They enable a user-convenient authentication alternative to passwords or tokens as well as high security identity assessment for law enforcement and border control. However, with a rising usage in general, fraudulent use increases as well. One drawback of biometrics in general is the lack of renewable biometric characteristics. While it is possible to change a password or token, biometric characteristics (e. g. the fingerprint) stays the same throughout a lifespan. Hence, biometric systems are required to ensure privacy protection in order to prevent misuse of sensitive data. In this context, this Thesis evaluates cryptographic solutions that enable storage and real time comparison of biometric data in the encrypted domain. Furthermore, long-term security is achieved by post-quantum secure mechanisms.

In addition to those privacy concerns, presentation attacks targeting the capture device are threatening legit operations. Since no information about inner system modules are required to use a presentation attack instrument (PAI) at the capture device, also non-experts could attack the biometric system. Thus, presentation attack detection (PAD) modules are essential to distinguish between bona fide presentations and attack presentations. In this regard, different methods for fingerprint PAD are analysed in this Thesis, including benchmarks on several classifiers based on handcrafted features as well as deep learning techniques. The results show that the PAD performance depends on material properties of the used PAI species in combination with the captured data type. However, fusing multiple approaches enhances the detection rates for both convenient and secure application scenarios.

# ZUSAMMENFASSUNG

Die Nutzung biometrischer Authentisierungsverfahren als Alternative zu Passwörtern und Schlüsselkarten ist im letzten Jahrzehnt stetig gestiegen. Große Anwendungsgebiete sind auf der einen Seite die benutzerfreundliche Entsperrung von mobilen Endgeräten, sowie auf der anderen Seite sicherheitskritische Identifizierungsverfahren bei der Grenzkontrolle. Allerdings steigt auch die Anzahl der Angriffe auf biometrische Systeme mit deren Verbreitung. Die Möglichkeit biometrische Charakteristika zur Authentisierung nutzen zu können, ist zugleich aus IT-sicherheitstechnischer Sicht ebenso ein Nachteil bezüglich des Datenschutzes. Während Passwörter und Schlüssel ausgetauscht werden können, sind biometrische Charakteristika (z.B. der Fingerabdruck) permanent. Daher erfordert die Bereitstellung biometrischer Systeme weiterreichende Maßnahmen um den Datenschutz gewährleisten und Missbrauch verhindern zu können. In diesem Zusammenhang werden in dieser Dissertation kryptographische Lösungen untersucht, um biometrische Daten sicher zu speichern und zudem im verschlüsselten Raum zu vergleichen. Um dabei langfristigen Schutz zu garantieren, werden ausschließlich Verfahren genutzt, die selbst zukünftigen Quantencomputern standhalten.

Neben den Datenschutzbedenken wird die Sicherheit biometrischer Systeme durch Präsentationsangriffe während der Aufnahme gefährdet. Insbesondere da für diese Angriffe keine weitgehenden Kenntnisse notwendig sind, können Artefakte auch von Laien einem biometrischen Erfassungsgerät präsentiert werden. Daher sind Verfahren zur Präsentationsangriff Detektierung (PAD) erforderlich um zwischen bona fiden Aufnahmen und Angriffen unterscheiden zu können. Zu diesem Zweck werden in dieser Dissertation verschiedene PAD Methoden für Fingerabdrucksysteme analysiert. Dies beinhaltet eine Bewertung von unterschiedlichen Verfahren basierend auf maschinellem Lernen. Die Ergebnisse zeigen, dass die Erkennungsleistung von den benutzten Materialien sowie der Aufnahmetechnik abhängt und es generell sinnvoll ist, ergänzende Ansätze zu kombinieren.

# PUBLICATIONS

## JOURNALS


[1] J. Kolberg, M. Gomez-Barrero, and C. Busch. "On the Generalisation Capabilities of Fingerprint Presentation Attack Detection Methods in the Short Wave Infrared Domain." In: *IET Biometrics* (2021), pp. 1–15.

[2] J. Kolberg, M. Grimmer, M. Gomez-Barrero, and C. Busch. "Anomaly Detection with Convolutional Autoencoders for Fingerprint Presentation Attack Detection." In: *IEEE Trans. on Biometrics, Behavior, and Identity Science (T-BIOM)* 3.2 (2021), pp. 190–202.


## BOOKCHAPTER


[1] J. Kolberg, M. Gomez-Barrero, S. Venkatesh, R. Ramachandra, and C. Busch. "Presentation Attack Detection for Finger Recognition." In: *Handbook of Vascular Biometrics*. Springer, 2020, pp. 435–463.


## CONFERENCES


[1] J. Kolberg, P. Bauspieß, M. Gomez-Barrero, C. Rathgeb, M. Dürmuth, and C. Busch. "Template Protection based on Homomorphic Encryption: Computationally Efficient Application to Iris-Biometric Verification and Identification." In: *Proc. IEEE Workshop on Information Forensics and Security (WIFS)*. 2019, pp. 1–6.

[2] J. Kolberg, P. Drozdowski, M. Gomez-Barrero, C. Rathgeb, and C. Busch. "Efficiency Analysis of Post-quantum-secure Face Template Protection Schemes based on Homomorphic Encryption." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2020, pp. 1–4.

[3] J. Kolberg, M. Gomez-Barrero, and C. Busch. "Multi-algorithm Benchmark for Fingerprint Presentation Attack Detection with Laser Speckle Contrast Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2019, pp. 1–5.

[4] J. Kolberg, A. C. Vasile, M. Gomez-Barrero, and C. Busch. "Analysing the Performance of LSTMs and CNNs on 1310 nm Laser Data for Fingerprint Presentation Attack Detection." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. 2020, pp. 1–7.

# FURTHER CONTRIBUTIONS

## JOURNALS

[1] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Erkennung von Präsentationsangriffen auf Fingerabdruck Systemen." In: *Datenschutz und Datensicherheit* 44.1 (2020), pp. 26–31.

[2] L. J. González-Soler, M. Gomez-Barrero, J. Kolberg, L. Chang, A. Pérez-Suárez, and C. Busch. "Local Feature Encoding for Unknown Presentation Attack Detection: An Analysis of Different Local Feature Descriptors." In: *IET Biometrics* (2021), pp. 1–18.

[3] A. Nautsch, A. Jiménez, A. Treiber, J. Kolberg, C. Jasserand, et al. "Preserving Privacy in Speaker and Speech Characterisation." In: *Computer Speech & Language* 58 (2019), pp. 441–480.

[4] A. Treiber, A. Nautsch, J. Kolberg, T. Schneider, and C. Busch. "Privacy-preserving PLDA Speaker Verification Using Outsourced Secure Computation." In: *Speech Communication* 114 (2019), pp. 60–71.

## BOOKCHAPTER

[1] M. Gomez-Barrero, R. Tolosana, J. Kolberg, and C. Busch. "Multi-Spectral Short Wave Infrared Sensors and Convolutional Neural Networks for Biometric Presentation Attack Detection." In: *AI and Deep Learning in Biometric Security: Trends, Potential and Challenges*. CRC Press, 2021, pp. 105–132.

## CONFERENCES

[1] P. Bauspieß, J. Kolberg, D. Demmler, J. Krämer, and C. Busch. "Post-Quantum Secure Two-Party Computation for Iris Biometric Template Protection." In: *Proc. IEEE Workshop on Information Forensics and Security (WIFS)*. 2020, pp. 1–6.

[2] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Towards Fingerprint Presentation Attack Detection Based on Short Wave Infrared Imaging and Spectral Signatures." In: *Proc. Norwegian Information Security Conf. (NISK)*. 2018.

[3] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Towards Multi-Modal Finger Presentation Attack Detection." In: *Proc. Intl. Conf. on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2018, pp. 547–552.

[4] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Multi-Modal Fingerprint Presentation Attack Detection: Looking at the Surface and the Inside." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2019, pp. 1–8.

[5] P. Keilbach, J. Kolberg, M. Gomez-Barrero, C. Busch, and H. Langweg. "Fingerprint Presentation Attack Detection using Laser Speckle Contrast Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2018, pp. 1–6.

[6] A. Nautsch, S. Isadskiy, J. Kolberg, M. Gomez-Barrero, and C. Busch. "Homomorphic Encryption for Speaker Recognition: Protection of Biometric Templates and Vendor Model Parameters." In: *Proc. Speaker Odyssey* (2018), pp. 16–23.

[7] R. Tolosana, M. Gomez-Barrero, J. Kolberg, A. Morales, C. Busch, and J. Ortega. "Towards Fingerprint Presentation Attack Detection Based on Convolutional Neural Networks and Short Wave Infrared Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2018, pp. 1–5.

*Niemand kann mich irritieren.*
*Für mich zählen nur die Fakten.*
*Niemand kann mich irreführen,*
*denn ich trau nur dem Exakten.*
*Ich such' Wahrheit,*
*und die Wahrheit will immer Klarheit.*
*Mein Verstand ist unbestechlich.*
*Ich studier das Positive.*
*Ich bin niemals oberflächlich.*
*Ich seh immer in die Tiefe.*
*Denn die Wahrheit, will immer Klarheit.*
*Mein Wissensdrang kommt nicht zur Ruh,*
*solang noch Zweifel nagen.*
*Ich lasse kein Geheimnis zu,*
*ich hör nicht auf zu fragen:*
*Wie und was und wer und wo und wann.*

— Abronsius in Tanz der Vampire (Musical) [272]

## ACKNOWLEDGMENTS

Foremost, I want to thank my supervisors Christoph, Marta, and Andreas Heinemann for their guidance throughout the last years. Special thanks to Christoph in the first place for inviting me to do a master thesis in biometrics and encouraging me to follow up in this area with a PhD. In this context, additional thanks to Marta for the daily supervision and assistance across all aspects. I learned a lot! Furthermore, I was happy to be a part of the da/sec group. Thank you all for the cooperative and pleasant working atmosphere, where everyone was willing to help whether it was a technical issue or unrelated to work. In particular I would like to name Andreas, Christian, Dailé, Daniel, Hareesh, Janier, Jannis, Pawel, Sergey, and Ulrich from the biometric staff as well as Lorenz, Thomas, and Thomas from forensics. During the time of my PhD, I had the pleasure to work with a number of students. I would like to express my gratitude to Pascal, Henrik, Ahmed, Alexandru, Marcel, Tobias, Jonas, and especially Pia, with whom I enjoyed collaborating over a period of two years. Thank you all for your contributions.

Furthermore, parts of this Thesis are based on collaborations. Hence, further thanks go to Ruben (UAM), Amos, Daniel, Juliane, and Thomas (ATHENE, TUD), Raghu and Kiran (NTNU), Daniel and Jörg (Jenetric), as well as Ralph and Arndt (BSI). It is always valuable to discuss topics from different perspectives and I am glad that our paths crossed. In this context, I profited as well from the EAB in general, which provides

a constant platform to connect and discuss with others. Moreover, thanks go to Mohamed, Leonidas, Joe, and Hengameh (USC) for their effort within the BATL project and the joyful times during our joint data collection periods, as well as Lars and Simona (IARPA) for the coordination. Additional thanks to the participants of our data collections as biometric research is simply not possible without all the volunteers.

I would like to thank my family and friends for their support and encouragement. My parents always were available for a change of location, be it for work or relaxation, which both were necessary from time to time. Further thanks to Lukas, Jens, and Nina, who showed interest in my work and offered assistance for proofreading. Finally, I want to thank Anni for her endless strength and love during this challenging last years in order to keep my balance.

# CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

## ACRONYMS

ACER Average Classification Error Rate. 9

ADA AdaBoost Classifier. 26, 43, 44

AE Autoencoder. 56–58, 62, 66, 67, 74, 75, 77, 78, 82, 84, 86, 88, 91

APCE Attack Presentation Classification Error. 71, 73, 77, 79, 81, 82, 84, 86, 88

APCER Attack Presentation Classification Error Rate. 8, 35, 38, 40–46, 77, 82, 133, 134

ARL adversarial representation learning. 12

AS authentication server. 100, 101, 109–111

AUC Area Under Curve. xix, 74

BFV Brakerski/Fan-Vercauteren. 18, 96, 99–101, 103, 104

BGN Boneh-Goh-Nissim. 18

BIP Biometric Information Protection. 4–7, 16, 93, 94, 97–100, 102, 104, 105, 109, 111, 113–115, 119–124, 127, 128

BPCER Bona fide Presentation Classification Error Rate. 8, 35, 36, 38–47, 71, 91, 126, 133, 134

BRIEF Binary Robust Independent Elementary Features. xix, 91

BSIF Binarized Statistical Image Features. 24, 25, 30, 31, 35, 36, 42–45, 91

CKKS Cheon-Kim-Kim-Song. 18, 95, 96, 99–101, 103, 104
CNN Convolutional NN. 11, 12, 14, 15, 27, 29, 33, 34, 36, 44, 45, 47, 50, 51, 54–56, 59–62, 66–68, 70–73, 77, 79–82, 84, 87, 88, 90, 123, 126

D-EER Detection Equal Error Rate. 8, 43, 45, 46, 78, 80, 134
DB database. 1, 2, 100, 101, 103, 104, 108–111
DET Detection Error Trade-off. 43, 45, 68, 72, 74, 75, 78–81, 84, 102, 119
DGK Damgård-Geisler-Krøigaard. 16
DT Decision Tree. 26, 43

FHE Fully Homomorphic Encryption. 95
FMR False Match Rate. 102, 111, 112, 119, 134
FNMR False Non-Match Rate. 102, 111, 112, 119, 134

GAN Generative Adversarial Network. 12, 15
GC Garbled Circuit. 16, 17, 93, 118, 120
GDPR General Data Protection Regulation. 4, 127
GMM Gaussian Mixture Model. 62, 77, 78
GNB Gaussian Naive Bayes. 27, 43–45

HD Hamming distance. 17, 18, 97, 100, 105–111, 117–119
HE Homomorphic Encryption. 16–18, 93–99, 102–104, 106, 112, 114, 115, 119–123, 127
HIST Greyscale histogram. 24, 34, 36, 43–45
HOG Histogram of Oriented Gradients. 25, 35, 36, 43–45, 91
HW Hamming weight. 97, 110, 118

KNN K-nearest neighbours. 9, 26, 43

LBP Local Binary Patterns. xix, 9, 11, 12, 24, 30, 31, 35, 36, 42–45, 91
LDA Linear Discriminant Analysis. 27, 43
LG Log-Gabor. 105, 115
LOO leave-one-out. 9, 12, 15, 66, 67, 80, 81, 86–88, 90, 91

LRCN Long-term Recurrent Convolutional Network. 54, 55, 59, 61, 73, 74, 79, 81, 82, 84, 86–88, 90, 91, 123, 126

LSCI Laser Speckle Contrast Imaging. 15, 22, 34–36, 42, 43, 45–48, 50, 123, 134

LSSC Linearly Separable Subcode. 100

LSTM Long Short-term Memory. 15, 54, 55, 60, 61, 66, 67, 72, 73, 81, 82, 90, 91, 123, 126

LWE Learning With Errors. xx, 17, 96

MC Maximum Curvature. 30, 36, 43

MSE mean squared error. xx, 57, 58, 62, 75

NIR Near Infrared. 15, 19, 20, 29, 30, 48, 111, 124

NN Neural Network. xviii, xx, 9, 11, 27, 54, 57

NTRU N-th degree truncated polynomial ring. 17, 96, 97, 99–101, 103–105, 109, 111, 113–115, 119–121

OC one-class. 56, 61, 62, 67, 77, 78, 91

OCT Optical Coherence Tomography. 14

ORB Oriented FAST and Rotated BRIEF. 91

PA Presentation Attack. 4, 5, 7, 19, 77

PAD Presentation Attack Detection. 4–9, 11, 12, 14, 15, 19, 21–24, 26, 27, 29–31, 33–36, 38–43, 46–51, 54, 56, 58–63, 65–68, 70, 72–75, 77–82, 84, 86–88, 90, 91, 123–128, 134

PAI Presentation Attack Instrument. 4, 7, 8, 14, 19–21, 29–31, 38, 45, 47, 48, 65, 66, 71, 73, 77, 81, 82, 86–88, 90, 123, 134

pAUC partial AUC. 74, 75, 77, 78

PHE Partially Homomorphic Encryption. 94, 95

PLBP Gaussian Pyramids and LBP. 30, 31, 40, 42

PRNG Pseudo Random Number Generator. 98

QDA Quadratic Discriminant Analysis. 27, 43

QSW Quadratic Spline Wavelet. 105

RE Reconstruction Error. 56, 58, 59, 62, 75

ReLU Rectified Linear Unit. 29

| | |
|---|---|
| RF | Random Forest. 26, 43–45 |
| RLWE | ring-LWE. 17, 96, 97 |
| RNN | Recurrent NN. 54 |
| RoI | Region of Interest. 21, 48, 125 |
| SGD | Stochastic Gradient Descent. 27, 43, 44 |
| SHE | Somewhat Homomorphic Encryption. 94, 95 |
| SIFT | Scale Invariant Feature Transform. 91 |
| STPC | Secure Two-Party Computation. 16, 18, 94, 97–99, 115, 118–123, 127 |
| SURF | Speed-Up Robust Features. 91 |
| SVM | Support Vector Machine. 9, 11, 26, 27, 30–32, 36, 40, 41, 43–45, 62, 77, 78, 126 |
| SWIR | Short Wave Infrared. 15, 21, 31–33, 36, 42–49, 59, 60, 62, 63, 66–68, 70, 71, 73–75, 77–79, 81, 82, 84, 87, 88, 90, 91, 123–126 |
| wMSE | weighted MSE. 58, 62, 75 |
| XOR | Exclusive-OR. 17, 97, 100, 110, 111, 116–118, 121 |

# 1 INTRODUCTION

User authentication methods can be classified into three categories that are based on *i)* knowledge, *ii)* possession, or *iii)* biometrics. Remembering passwords and PINs are examples for the first category, while classical keys, smartcards, and tokens belong to the second one. In context of the last category, biometric recognition is defined as the "automated recognition of individuals based on their behavioural and biological characteristics" [146]. Modalities as face, fingerprint, and iris are considered biological, and gait, signature, and keystrokes are behavioural characteristics. On the other hand, the human voice combines both as it has a biological origin but also a behavioural proportion since the ability to talk needs to be learnt. The main advantage of biometrics with respect to the other two categories is that biometric characteristics cannot be forgotten or shared among users. On the contrary, the nature of biometrics makes it impossible to renew or exchange biometric characteristics. Within the last decade, the news reported multiple data breaches comprising biometric data of several million individuals[1][2][3][4][5][6]. *Therefore, security and privacy are essential elements of biometric systems.*

Nowadays, biometric recognition is present in a wide range of applications in our daily life. Typical scenarios include smartphone unlocking, access control for specific rooms or buildings, and identity confirmation during border control. Hence, biometric systems can be used for convenience, such that data subjects can authenticate with their biometric characteristics, or in high security scenarios, where e. g. law enforcement requires to identify a specific person or verify an ID claim.

In general, biometric verification is the process where one probe is compared to one reference (1:1 comparison) in order to verify a claimed ID. Biometric identification, on the other hand, compares one probe to all references (1:n comparison) in order to find the

1 Washington Post - Hacks of OPM databases compromised 22.1 million people, federal authorities say (2015/07/09)
2 CNN - Hackers stole 5.6 million government fingerprints - more than estimated (2015/09/23)
3 Washington Post - U.S. Customs and Border Protection says photos of travelers were taken in a data breach (2019/06/10)
4 The Guardian - Major breach found in biometrics system used by banks, UK police and defence firms (2019/08/14)
5 vpnMentor - Report: Data Breach in Biometric Security Platform Affecting Millions of Users (2019/08/14)
6 Website Planet - Report: Retail-focused Used Electronics Business Leaks Customers' IDs & Fingerprints in Data Breach (2020/11/17)

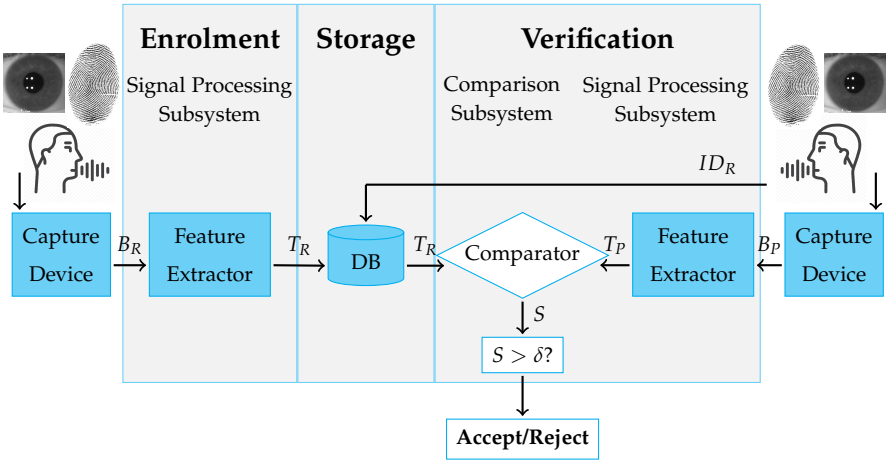



Figure 1.1: Enrolment and verification steps in a biometric system.

biometric reference identifier of the probe. The reference can either be stored in a database (e. g. company access control) or at a device of the data subject itself (e. g. biometric travel documents). Figure 1.1 illustrates the procedure of a biometric system for the enrolment and biometric verification use case. During enrolment, the biometric characteristic is presented to the capture device, which generates a captured biometric sample $B_R$ to be used as reference. After the subsequent feature extraction, the reference template $T_R$ is stored in the database (DB). For biometric verifications, the first steps are identical. The capture device acquires the biometric probe $B_P$ from the presented characteristic and the feature extraction creates the probe template $T_P$. Additionally, the data subject specifies a reference $ID_R$, that is used for comparison in order to verify the identity of the data subject. Thus, the comparator receives $T_R$ and $T_P$ and derives from them a comparison score $S$, which is compared to a decision threshold $\delta$ to finally accept or reject the verification attempt. For a biometric identification, no ID claim is send to the DB but $T_P$ is compared to all references in the DB. In the end, the resulting list is sorted in order to find the most similar candidate(s). While the computational cost are considered trivial for biometric verifications, the complexity for biometric identifications increases with the number of enrolled subjects.

## 1.1 ATTACKS ON BIOMETRIC SYSTEMS

As most information systems, biometric systems attract different attacks [95, 247]. This thesis focusses on those threatening the *security or privacy* of the system and its data subjects.

As one of the earliest attacks, hill-climbing methods received attention since the biometric system was attacked regardless of the used modality. Research has shown efficient hill-climbing approaches for

face [99], fingerprint [201], and iris [237] recognition systems. These attacks utilise a weakness in case the comparison score is disclosed to the data subject, who is interacting with the system. A straightforward way is to submit a randomly generated probe sample and save the comparison score. Now, the attacker can change one part of this sample at a time and see whether the score gets better or worse. Changes that decrease the similarity to the claimed reference template are reverted and changes that increase the similarity are kept. Therefore, the attacker can iteratively create a new probe sample that is eventually accepted for verification. In case the attacker aims to get access, the attack succeeded. In other cases, the attacker continues until the comparison score is optimised in order to reconstruct the enrolled reference template. Hence, the attacker can repeatedly target all references in order to reconstruct the whole database. As a countermeasure, biometrics systems can conceal the comparison score and only forward the final decision.

In case of stolen templates or database leakages, it is also possible to reconstruct biometric samples. Since the template representations contain features of the particular characteristic that enable biometric systems to verify or identify a specific data subject, it is also possible to reconstruct the original sample (e. g. the captured image). Successful sample reconstruction methods have been published for fingerprint minutiae [44, 45, 94], iris codes [100], traditional face templates [2] as well as deep face templates [195], and hand-shape recognition [108]. It is important to note that these reconstructed images may contain some distortions or artefacts and thus might not fool a human. However, they are good enough to fool the biometric system as similar features are extracted resulting in a comparison score that gets accepted. A successful defence mechanism against sample reconstruction requires an irreversible transformation as defined in ISO/IEC 24745 on biometric information protection [144].

ISO/IEC 30107-1 [145] defines multiple attack points on biometric systems in order to bypass the authentication as depicted in Figure 1.2. While steps 2 to 9 require knowledge and access to the inner functions of a biometric system, the capture device is accessible for interaction with data subjects and possible attackers. Hence, there are no barriers to perform a Presentation Attack (PA) at the sensor since no expertise is required to e. g. print a photo and present it to the face recognition system. In general, PAs are a threat to all biometric modalities and for each modality various materials [161, 165] can be utilised to create a Presentation Attack Instrument (PAI) (e. g. silicone face mask, contact lens, or wax fingerprint overlay) that is presented to the capture device.

The intention of the attacker can be twofold as PAs can be classified into impersonation attacks and concealing attacks [145]. The biometric impostor tries to get recognised as another legit data subject, while the identity concealer aims to hide its own identity e. g. due to black-listing.

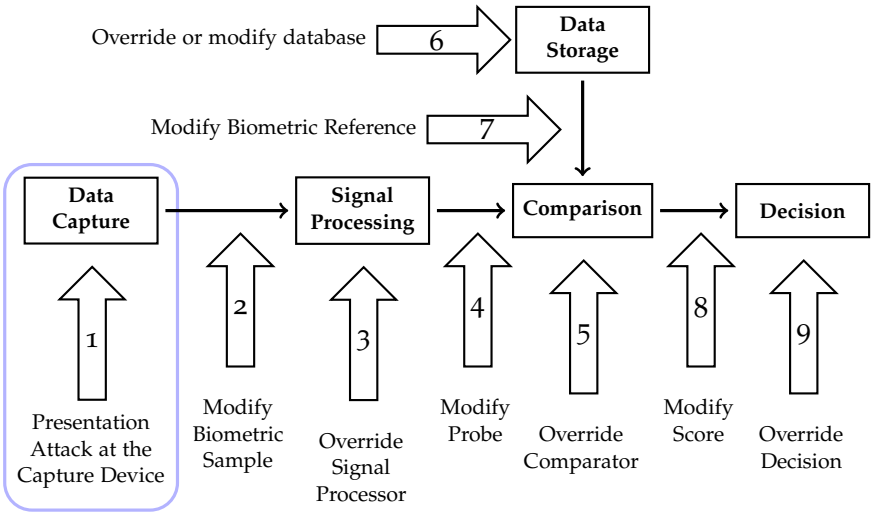



Figure 1.2: Attack points on biometric systems according to [145].

As a consequence, Presentation Attack Detection (PAD) methods [199] are essential for biometric systems in order to distinguish between bona fide presentations and attack presentations.

## 1.2 MOTIVATION

The possibility of attacks against biometric systems are reason enough to motivate research on security and privacy. Nobody is interested in using or operating a system that is known to be vulnerable.

Furthermore, biometric data are considered sensitive data as declared in the General Data Protection Regulation (GDPR) [91]. This regulation requires that biometric data is always protected to preserve the privacy of the data subjects [171]. In this context, ISO/IEC 24745 [144] defines three requirements on Biometric Information Protection (BIP) for biometric systems: *i) irreversibility*, retrieving original samples from a given protected sample is not possible, *ii) unlinkability*, it is impossible to link two protected templates to the same subject, *iii) renewability*, old templates can be revoked and new ones created without needing the subject to re-enrol. Additionally, the biometric recognition performance should not decrease for protected systems in comparison to the unprotected performance. Since those requirements imply additional computational costs, this Thesis investigates efficient solutions with long-term security. The goal is to utilise generic approaches that can be applied to various biometric modalities and execute in real time.

In addition to privacy protection, system operators have a high interest on PAD to detect attacks on the capture device. In this regard, especially applications in the German public sector need to fulfil the requirements defined in the technical guideline of the Federal Office for Information Security to prevent attempts to deceive e. g. during

border control [39] or passport enrolment [40]. Hence, companies manufacturing biometric capture devices have a strong interest to include PAD in their products in order to raise sales. Otherwise companies are completely excluded from selection in public sector applications, since PAD is a key element in public calls for tenders. Since the PAD methods highly depend on the target biometric modality and in some cases additionally on the capture device, this Thesis focusses on PAD for fingerprint recognition only.

All in all, this Thesis presents approaches enhancing the security for fingerprint recognition applications in terms of PAs and further methods for privacy preservation in terms of BIP. The overall goal is to strengthen the trust of data subjects as well as system operators for their usage of biometric systems.

## 1.3 RESEARCH QUESTIONS

Derived from the motivation, the following research questions are defined for this Thesis to investigate solutions for fingerprint PAD in particular and BIP in general.

*Security Enhancement*

RQ1: Which type of data needs to be captured for reliable fingerprint presentation attack detection?

Since samples acquired from legacy fingerprint capture devices might not include enough details to distinguish between bona fide presentation and attack presentation, new capture devices could be developed. In this case, further questions arise:

- What type of sensors are included in the capture device?
- Does the captured data require particular pre-processing?
- Is this system still compatible with legacy fingerprint sensors?

The PAD performance is not as relevant if the captured fingerprints cannot be compared to legacy database entries. In this context, different PAD methods are evaluated based on data captured by new developed fingerprint capture devices.

RQ2: Which machine learning classifiers aid the detection of attack presentations while keeping the false alarm rate low?

Given the vast amount of different machine learning tools, it is of interest to benchmark several algorithms instead of using only one. Due to multiple possibilities, the following point needs to be considered:

- Does the combination of classifier and PAD data require further pre-processing of the data?

Depending on the data type, additional steps might be required to extract relevant features for further processing. In general, RQ1 and RQ2 are strongly linked as capturing PAD data on its own is useless without feeding them to a classifier. On the other hand, the choice of the classifier highly depends on the acquired PAD data.

*Privacy Protection*

RQ3: Which concepts are suited for the protection of biometric systems while allowing real time efficiency?

Since privacy-preserving transformations and comparisons imply computational overhead, protected systems still need to operate in real time. The challenge is to investigate mechanism to speed up the transaction time without losing biometric accuracy, while definitely preserving the cryptographic security.

## 1.4 THESIS OUTLINE

This Section provides an overview of the content covered in this Thesis and how it is organised in the following Chapters:

- Chapter 1 introduces biometric systems and possible attacks on those to the reader. As a result, motivation and research questions are defined in order to investigate on countermeasures within this Thesis.
- Chapter 2 summarises related work and the state of the art for fingerprint PAD on the one hand and BIP on the other hand. This Chapter focusses on research without own contributions to give a broad overview on the topic.
- Chapter 3 presents own contributions in the area of fingerprint PAD. This includes the utilised concepts of machine learning as well as the experimental evaluations.
- Chapter 4 contains the work on BIP across multiple biometric modalities. Long-term security of the proposed methods is assured by relying on post-quantum cryptography.
- Chapter 5 concludes the contributions of this Thesis by answering the research questions and finally outlines perspectives for future work.

# 2 RELATED WORK

This Chapter summarises the state of the art without own contributions to provide an overview in the areas of PAD and BIP. Although security and privacy are somehow connected, since biometric systems require both, the modules are completely independent. Hence, the first Section reviews related work on fingerprint PAD while the second Section looks at methods for BIP.

## 2.1 FINGERPRINT PRESENTATION ATTACK DETECTION

As defined in ISO/IEC-30107-3 [147], PAD distinguishes between bona fide presentations (i. e., the sample stems from a real data subject) and attack presentations (i. e., an artefact is presented to the capture device). During the attack, a PAI that represents a fingerprint pattern is presented to the capture device. In order to fabricate a fingerprint PAI, either a cooperative target subject or a latent fingerprint is required [104]. For the first option, the target finger is pressed into some modelling compound, which will harden subsequently. The solid mould is then filled with silicone, gelatin, or a similar material to cast the PAI. In the second scenario, a latent fingerprint can be made more visible by adding for example iron powder as forensic experts do. This analogue representation is then digitised and potentially manually enhanced. Subsequently, the negative image is printed on a transparent film. By using acid-treating, for example on a printed circuit board, the mould is created and can again be filled to create the PAI. As an alternative to latent fingerprints, a digital photo of the fingerprint works equally well. Since fingerprints are unique even for relatives and twins [126], PAs are restricted to single instances rather than attacking all enrolled subjects of a system.

Solutions for fingerprint PAD can be grouped into two classes: *i)* software-based, where legacy fingerprint samples are deeply examined by software algorithms, and *ii)* hardware-based, where additional sensors capture further PAD data which then is analysed by corresponding software. Given the vast number of articles studying fingerprint PAD, the following Sections summarise the most relevant ones for this Thesis and the interested reader is referred to [138, 169, 197, 261, 265] for more comprehensive surveys.

### 2.1.1 *Metrics for Presentation Attack Detection*

In the following, the main definitions of the standards ISO/IEC 30107-1 [145] and ISO/IEC 30107-3 [147] on biometric presentation attack detection are introduced, since these are used throughout the Thesis:

- bona fide presentation: "interaction of the biometric capture subject and the biometric data capture subsystem in the fashion intended by the policy of the biometric system" [147]. A normal or genuine presentation.
- attack presentation: "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system" [147]. An attack presentation to the capture device to either conceal the own identity or impersonate someone else.
- Presentation Attack Instrument (PAI): "biometric characteristic or object used in a presentation attack" [145]. For instance, a printed face photo, a contact lens, or a silicone fingerprint overlay.
- PAI species: "class of presentation attack instruments created using a common production method and based on different biometric characteristics" [147].
- Attack Presentation Classification Error Rate (APCER): "proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario" [147].
- Bona fide Presentation Classification Error Rate (BPCER): "proportion of bona fide presentations incorrectly classified as attack presentations in a specific scenario" [147].
- Detection Equal Error Rate (D-EER): Error rate at the operation point where APCER = BPCER.

In addition to the D-EER, further operation points can be fixed in order to benchmark different systems. In this context, $APCER_{0.2}$ describes a convenient system for a fixed BPCER = 0.2% and $BPCER_{0.2}$ a high security application where APCER = 0.2%.

### 2.1.2 *Software-based Fingerprint PAD*

The main advantage of software-based fingerprint PAD approaches is that those solutions could iteratively be enhanced via software updates. Thus, there is no need to replace existing capture devices if the supplier offers update possibilities. A summary of the reviewed works[1] is

1 Parts of this Section are derived from our publications [176, 177].

presented in Table 2.1. A continuous baseline was established with the biannual LivDet competitions[2] [104, 293] starting in 2009, which are widely used for fingerprint PAD development and benchmarking. In addition to providing public datasets to enable research, the organisers further define the Average Classification Error Rate (ACER) as given in Eq. (2.1), which is often the only reported metric in corresponding articles.

$$ACER = \frac{APCER + BPCER}{2} \tag{2.1}$$

A total of six fingerprint datasets has been released so far: LivDet 2009 [200], LivDet 2011 [292], LivDet 2013 [105], LivDet 2015 [208], LivDet 2017 [209], and LivDet 2019 [221]. The fingerprints are acquired with multiple capture devices, thus allowing evaluations regarding generalisation capabilities. Moreover, some unknown attacks are included in the test set of 2015 and starting in 2017, all PAI species in the test sets are unknown. In addition to the LivDet datasets, further research is based on other acquisitions: ATVS FFP [97], MSU-FPAD and PBSKD (both [55]).

An early trend of fingerprint PAD was the localisation of sweat pores [52, 87, 203] based on high resolution images. Due to their tiny size and since they cannot be extracted from latent fingerprints, this approach seemed promising. However, the algorithms did not achieve a stable performance and hence research shifted the focus to other features. In this context, Nikam and Agarwal [215] extracted textural information from the fingerprint images as well as wavelet energy features based on ridge frequency and orientation. Fusions of both features were then used to train Support Vector Machine (SVM), K-nearest neighbours (KNN), and Neural Network (NN) classifiers. A general PAD approach, that is also applicable for other modalities, was proposed by Galbally et al. [97, 98]. The authors applied a set of complementary image quality metrics and evaluate different combinations of those for their usability of PAD. Through analysing ridge signal, valley noise, and region labelling, Tan et al. [275] found that unknown scenarios such as environmental conditions and new unseen PAI species are challenging for fingerprint PAD algorithms. On the other hand, results were much better, when the PAI species are available included in the training process. In a similar setting, Marasco and Sansone [198] evaluated the impact of unknown materials following a leave-one-out (LOO) protocol. The results showed that a combination of multiple fingerprint PAD algorithms based on static and intensity features, helped to improve the detection of unseen PAI species.

The freely available LivDet datasets quickly established a baseline for fingerprint PAD algorithm benchmarks. Following the direction of texture analysis, Jia et al. [155] combined multi-scale Local Binary Patterns (LBP) feature extraction with SVM classification. The results

2 https://livdet.diee.unica.it/ (previous competitions: https://livdet.org)

| Year | Ref. | Description | Performance * | #PAI | Database |
|---|---|---|---|---|---|
| 2007 | [52] | Pore spacing | † CCR=85.2% | 1 | Own DB |
| 2008 | [215] | LBP + wavelet energy | † CCR=97.4% | 2 | Own DB |
| 2010 | [275] | Valley noise, region labelling | D-EER≤5.9% | 6 | Own DB |
| 2011 | [87] | Closed sweat pore extraction | APCER=21.2%<br>BPCER=8.3% | 4 | Own DB |
| | [203] | Active sweat pore localisation | N/A | 0 | BFBIG-DB1 |
| | [198] | Static + intensity features | D-EER=12.45% | 3 | LivDet 2009 |
| 2012 | [97] | 10 fingerprint quality metrics | ACER=10.4% | 4 | LivDet 2009,<br>ATVS FFP |
| 2014 | [98] | 25 image quality metrics | APCER<13%<br>BPCER≤14% | 3 | LivDet 2009 |
| | [155] | Multiscale LBP | D-EER=7.52% | 7 | LivDet 2011 |
| 2015 | [242] | One-class SVM re-calibration | D-EER=7.0% | 7 | LivDet 2011 |
| 2016 | [74] | One-class SVMs | ACER=14.7% | 7 | LivDet 2011 |
| | [218] | Pre-trained CNNs | ACER=2.90% | 8 | LivDet 2009-13 |
| | [170] | Deep believe network | D-EER=1.16% | 5 | LivDet 2013 |
| 2017 | [154] | Contrast enhancement + CNN | ACER=0.2% | 2 | ATVS FFP |
| | [118] | Bag of Words + SIFT | APCER=5%<br>BPCER=4.3% | 7 | LivDet 2011 |
| | [283] | Patch-based CNN | ACER=3.26% | 8 | LivDet 2011-13 |
| | [282] | Handcrafted feature fusion | ACER=1.6% | 8 | LivDet 2009-13 |
| 2018 | [157] | LBP extracted from<br>Gaussian pyramids (PLBP) | ACER=21.21% | 7 | LivDet 2013 |
| | [55] | Minutiae-centered CNN<br>Fingerprint Spoof Buster | APCER<7.3%<br>BPCER=1% | 12 | LivDet 2011-15,<br>MSU-FPAD, PBSKD |
| 2019 | [57] | Minutiae-centered CNN<br>generalisation | APCER=4.7%<br>BPCER=0.2% | 12 | MSU-FPAD,<br>PBSKD |
| | [119] | Fisher vector encoding | D-EER=1.88% | 13 | LivDet 2011-15 |
| 2020 | [3] | Incremental learning | ACER=2.4% | 13 | LivDet 2011-15 |
| | [156] | DenseNet + genetic algorithm | ACER=1.78% | 13 | LivDet 2009-15 |
| | [56] | Minutiae-centered CNN<br>sample generator | APCER=8.22%<br>BPCER=0.2% | 12 | LivDet 2017,<br>MSU-FPAD, PBSKD |
| | [124] | CNN + ARL<br>generalisation | APCER=7.06%<br>BPCER=0.2% | 11 | LivDet 2015-17<br>MSU-FPAD |
| 2021 | [255] | CycleGAN | ACER=2.17% | 10 | LivDet 2015 |
| | [120] | Local feature encoding | ACER=1.74% | 15 | LivDet 2011-19 |

* Some authors evaluated multiple experiments and only one result is included in the table.
† CCR = Correct Classification Rate

Table 2.1: Summary of reviewed software-based fingerprint PAD methods.

prove their assumption that a multi-scale approach is superior for fingerprint PAD. Later, Jiang et al. [157] computed three Gaussian pyramids before extracting LBP histograms. This multi-resolution analysis achieves more robustness against outliers. An extensive study of handcrafted fingerprint PAD methods was done by Toosi et al. [282]. Their benchmark of different feature fusions and classifiers revealed that these solutions are not robust against cross sensor scenarios. In order to counter the problem of unknown PAI species, one-class approaches consider all attack presentations as unknown and are only trained on bona fide presentations. In this context, Rattani et al. [242] utilised one-class SVMs, which were further fine-tuned on specific PAI species. The resulting binary classifiers were then tested on unknown attacks and additionally re-calibrated on those in order to evaluate the impact of the particular material. Likewise, Ding and Ross [74] trained one-class SVMs on twelve feature sets, before refining the hypersphere on a small number of attack presentations. Their final score fusion allows for more generalisability as it counters the instability of single instances.

In addition to unknown attacks, further generalisation experiments include cross sensor as well as cross database evaluations. Starting with an unknown attack scenario, González-Soler et al. [118] showed that the Bag of Words [262] feature encoding of handcrafted features achieves better fingerprint PAD performance than other existing feature descriptors without encoding. In a follow-up study [119], the error rates could be further reduced by utilising the Fisher vector [252] feature encoding. These results are further confirmed in [120] through an extensive evaluation on unknown attacks, cross sensor, and cross database scenarios. Across all experiments, the Fisher vector encoding accomplished the best fingerprint PAD performance.

In contrast to the aforementioned handcrafted approaches, latest research focusses mostly on deep learning methods, which combine feature extraction and classification in one model. For instance, Toosi et al. [283] fine-tuned AlexNet [180] for fingerprint PAD. As most deep learning algorithms require large training sets, the authors apply data augmentation of image patches to artificially increase the number of training samples. In a similar way, Jang et al. [154] trained a Convolutional NN (CNN) inspired by VGG [260] on non-overlapping patches. Moreover, they showed that prior contrast enhancement is beneficial for fingerprint PAD. Using a deep believe network, Kim et al. [170] also rely on data augmentation and patch-based input. However, the authors additionally report stable PAD performance across different capture devices. In another approach, Chugh et al. [55] centered their patches on the located minutiae points. Hence, only relevant patches were used to train their *Fingerprint Spoof Buster* and consequently this network outperformed several other state of the art approaches. More recently, Agarwal et al. [3] proposed an incremental

learning model, which offers a high level of robustness by clustering the training data. The advantage is an efficient way to retrain the model on new data without forgetting previous clusters. In another approach, Jian et al. [156] proposed DenseNet in combination with a genetic algorithm optimisation. Hence, the model architecture is adjusted during the training procedure in order to optimal fit to the particular use case.

While this might improve the performance for specific scenarios, it is most unlikely to generalise well on unknown test scenarios. A general lack of generalisability was already discussed by Nogueira et al. [218], who tested three different CNNs on multiple LivDet datasets. Their deep learning approach clearly outperforms a handcrafted LBP-based fingerprint PAD algorithm. However, when evaluating unknown attacks, cross sensor, and cross database scenarios, significantly higher errors rates were reported. In order to address the poor generalisability, Sandouka et al. [255] proposed the usage of Generative Adversarial Networks (GANs). In particular, the authors utilise a vision transformer in combination with CycleGAN in order to evaluate cross sensor as well as cross material scenarios. The results show that data augmentation is very important in order to train more generalisable networks. In this context also Chugh and Jain [57] evaluated their *Fingerprint Spoof Buster* towards unknown attacks following a LOO protocol. As a result, they defined a generalisation training set of selected PAI species, which allows to detect the left-out materials due to similar appearances. The authors further extended this work by including an universal material generator, which was trained to generate additional synthetic samples [56]. This generator was then used by Grosz et al. [124] in combination with adversarial representation learning (ARL) to train a generalisable CNN. Since ARL is independent of the target domain, it is especially strong in cross sensor and unknown attack scenarios.

### 2.1.3 *Hardware-based Fingerprint PAD*

As any other pattern recognition tasks, PAD can also benefit from further data captured by additional sensors incorporated into the capture device, which includes e. g. different illumination techniques or pulse measurements. Table 2.2 summarises reviewed hardware-based approaches[3]. Since these works usually require an own database collection, the number of samples are listed instead of a database name.

Manivanan et al. [196] followed the idea of detecting active sweat pores with help of a high resolution camera. However, the authors demonstrated the concept only on one bona fide presentation and miss the opportunity to prove the soundness of this technique. As one

3 Parts of this Section are derived from our publications [177, 178].

| Year | Ref. | Description | Performance | #PAI species | #PA samples | #BF samples |
|---|---|---|---|---|---|---|
| 2008 | [249] | Multi-spectral<br>wavelet transform | APCER=0.9%<br>BPCER=0.5% | 49 | 27,486 | 17,454 |
| 2010 | [196] | Active sweat pores | N/A | 0 | 0 | 1 |
| 2011 | [129] | Multi-spectral<br>blanching effect, pulse | APCER=0%<br>BPCER=0% | 4 | 7-15 | 11-28 |
| 2013 | [76] | Optical methods<br>pulse, pressure, reflections | APCER=10%<br>BPCER<2% | N/A | N/A | N/A |
| 2016 | [68] | OCT, double bright peaks<br>+ autocorrelation | APCER=0%<br>BPCER=0% | 3 | 28 | 540 |
| 2018 | [137] | SWIR, LSCI<br>+ patch-based CNN | APCER=0%<br>BPCER=0% | 17 | 227 | 551 |
| 2019 | [189] | OCT peak analysis | APCER=0%<br>BPCER=0% | 4 | 60 | 30 |
| | [205] | LSCI CNN + LSTM<br>patch-based | APCER≤0.14%<br>BPCER≤0.11% | 6 | 218 | 3,743 |
| | [58] | OCT patch-CNN<br>(no fingerprints) | APCER=0.27%<br>BPCER=0.2% | 8 | 357 | 3,413 |
| | [84] | One-class GANs<br>RaspiReader | APCER=50.2%<br>BPCER=0.2% | 12 | 5,531 | 11,880 |
| | [231] | Colour time series | APCER=3.55%<br>BPCER=0.2% | 16 | 21,700 | 14,892 |
| 2020 | [139] | Video-based<br>dynamic features | APCER=5%<br>BPCER=18.1% | 7 | 1,386 | 396 |
| | [268] | Multi-modal<br>3-fold and LOO analysis | depends on<br>experiment | 45 | 4,507 | 21,998 |
| 2021 | [140] | Spatial + temporal<br>dynamic features | APCER=5%<br>BPCER=1.11% | 7 | 1,386 | 396 |
| | [190] | OCT<br>Autoencoder | APCER=5%<br>BPCER=3.41% | 101 | 121 | 233 |

Table 2.2: Summary of reviewed hardware-based fingerprint PAD methods.

of the most reliable techniques, Optical Coherence Tomography (OCT) sensors [34, 266] are especially suited for fingerprint PAD. By scanning up to two millimeter into the skin, a 3D model of the fingertip is constructed and fingerprint PAIs can be detected. Furthermore, a scan of bona fide presentations also includes the inner fingerprint, which enables fingerprint recognition of subjects with worn-out instances. In the context of fingerprint PAD, Darlow et al. [68] focussed on the occurring double bright peaks in the OCT scans to distinguish bona fide presentations from attack presentations of three different PAI species. Likewise, Liu et al. [189] analysed that exactly two peaks appear for bona fide presentations. Moreover, the maximum peak is always on second position, which can be successfully exploited for fingerprint PAD with a threshold comparison. However, also deep learning can be used as demonstrated by Chugh and Jain [58]. The authors extract overlapping patches from one OCT scan to train a CNN. Based on a larger dataset, the OCT PAD performance drops slightly below 100%. However, since only one B-scan is captured, the actual fingerprint information is not available. In another approach Liu et al. [190] developed a one-class convolutional autoencoder based on ResNet [128]. During training only bona fide presentations are used and their test set contains 101 PAI species, which are not further described. One fingerprint sample comprises 400 depth scans, which are separately used in their model. The final PAD score is a score fusion based on the decoded and the latent representations. Additional fingerprint PAD methods based on OCT data were reviewed by Moolla et al. [207]. Despite the fact, that OCT sensors became commercially available in the last years (e. g., ThorLabs[4]), their high costs remain a disadvantage in contrast to different PAD techniques.

Hence, other approaches utilise multiple illumination techniques to develop fingerprint PAD methods. The research of Rowe et al. [249] led to the first multi-spectral capture device by Lumidigm. Fingerprint images are acquired in four differently coloured illuminations in order to improve the recognition accuracy as well as to detect attack presentations. The evaluation is carried out with 49 PAI species and over 44,000 samples in total. Hengfoss et al. [129] followed a more general approach and analysed all wavelengths between 400 nm and 1,650 nm. In particular, the authors observed the reflections during blanching (i. e., while a finger is pressed against a surface, the fingertip loses colour since the blood is squeezed out). This liveness indicator can also be used to detect cadaver fingers. In addition to this multi-spectral analysis, this study also measures the pulse. However, the authors conclude it is less applicable for efficient fingerprint PAD since the capture process itself lasts longer. In a similar way, Drahansky et al. [76] proposed optical methods for pulse, pressure, and skin reflections. They show that reflections from multi-spectral illuminations

4 https://www.thorlabs.com/navigation.cfm?guide_id=2039

are more reliable in terms of fingerprint PAD performance than both other methods. However, no information about the collected dataset are published.

Based on images captured with the *RaspiReader* [83], Engelsma and Jain [84] evaluated one-class GANs regarding their ability to detect unknown attack presentations. The utilised models are based on the DCGAN architecture [234]. The results show room for improvement since especially transparent PAI species are wrongly classified. In another work, Plesh et al. [231] analyse static features of traditional greyscale images as well as dynamic features of colour time series. Their final feature-level fusion combines both methods for fingerprint PAD.

Husseis et al. [139] used thermal and optical sensors and captured videos in order to analyse dynamic fingerprint statistics. Their PAD method describes visual features and the results show that statistical differences between bona fide presentations and attack presentations exist. The authors further improve their work in [140] by taking into account spatial and temporal features within the captured image sequence. The evaluation reveals that the performance depends on the combination of feature extraction and sensing technique. All in all, these methods achieve significantly better results than their first approach. Within a short video sequence (e. g., 1 second), the Laser Speckle Contrast Imaging (LSCI) technique [258] visualises movement within blood tissues beneath the skin. In particular, the laser illumination has a specific penetration depth for non-solid surfaces. If pointed at living skin, the scattered reflections change over the capture time due to the movement within the tissues [288]. In the area of fingerprint PAD, Mirzaalian et al. [205] process these LSCI sequences with multiple deep learning networks. While classical CNNs do not take temporal information into account, the Long Short-term Memory (LSTM) is especially designed for this use case. As a result, the LSTM model achieves better PAD performance than the tested CNNs.

Utilising multi-spectral illuminations in the Short Wave Infrared (SWIR) domain has proven suitable for face PAD [270] since different skin types [96] reflect these wavelengths in a similar way. In the context of fingerprint PAD, Hussein et al. [137] fused LSCI and SWIR PAD methods. This combination turned out to benefit from both strengths and improve the PAD performance. More recently, Spinoulas et al. [268] additionally include Near Infrared (NIR) and finger vein samples in their extensive fingerprint PAD algorithm benchmark. The evaluation consists of 3-fold partitions as well as LOO protocols in order to analyse the performance towards unknown attacks. The results show more robustness for fused systems in contrast to standalone algorithms.

## 2.2 BIOMETRIC INFORMATION PROTECTION

A considerable amount of research deals with privacy-preserving methods for BIP, whereof the most relevant ones are summarised in this Section[5]. Generally, three categories of BIP approaches can be defined, namely: *i*) cancelable biometrics [183, 226], where the sample or its template are irreversibly transformed; *ii*) cryptobiometric systems [43, 46, 152], which either extract a key from the biometric data or bind one to it; and *iii*) biometrics in the encrypted domain [7, 287], where cryptographic methods such as Homomorphic Encryption (HE) [1] or Secure Two-Party Computation (STPC) [71, 230] are applied for BIP. Except for Bloom filters [78, 114–116], cancelable and cryptobiometric BIP systems usually suffer from an accuracy degradation due to the applied schemes [238]. On the other hand, comparing biometric templates in the encrypted domain maintains the performance of the unprotected system, since identical distance computations can be applied. Additionally, provable security can be granted as the BIP methods build upon state-of-the-art cryptographic problems. Furthermore, in contrast to the more specific solutions for cancelable and cryptobiometric BIP, comparisons in the encrypted domain can be done independently of the biometric modality used and thus are more general applicable. Because of these properties and recent advances in HE [4, 9], the focus in this Thesis lies on BIP in the encrypted domain. While the most relevant approaches are introduced in the following, the interested reader is referred to [210, 211, 238, 253] for more extensive surveys, which also include the first two categories. In addition, an overview of BIP schemes in the encrypted domain is presented in [13, 37].

Anonymous biometric access control is possible at the cost of a biometric identification, in case the protocol only evaluates whether a subject is enrolled in the system or not. In this context, Ye et al. [299] as well as Luo et al. [192] presented approaches for iris recognition based on HE. In 2009, Erkin et al. [86] proposed BIP for face recognition based on Eigenfaces [285]. The distances are computed using the Paillier HE scheme [224] based on quantised integer templates and the final threshold comparison is achieved using Damgård-Geisler-Krøigaard (DGK) HE [64, 65]. In a subsequent work, Sadeghi et al. [251] improved the efficiency by introducing Garbled Circuits (GCs) [296] for the secure distance computation. This has the advantage that complex calculations can be shifted to the offline phase, thus accelerating the online comparison.

Gomez-Barrero et al. [107] implemented BIP for fixed-length signature templates and evaluate different distance measures. Real time biometric verifications are achieved by using the Paillier HE scheme [224]. However, this scheme has the disadvantage that the client computes

5 Parts of this Section are derived from our publications [16, 173, 174].

the distance between reference and probe templates. Hence, security is only theoretically given in the honest but curious model [127], since the client (who wants to get authenticated) can simply encrypt a distance score that gets accepted in order to bypass the authentication attempt. A further improvement in [109] allows template sizes of variable length, while also increasing the biometric accuracy of the system. A subsequent study [113] combines the signature approach with fixed-length fingercodes [150] in order to evaluate three different fusion schemes for a multi-biometric system. Since the general framework remains identical, malicious clients can always get authenticated.

The same problem exists for other approaches based on Paillier HE [224] due to the limited homomorphic properties. Additions are fully supported but for multiplications in the encrypted domain, one part needs to be available in plaintext. In this context, Rane et al. [236] let the client compute the distance of quantised fingerprint templates. On the other hand, the system by Osadchy et al. [222] is secure against malicious clients at the cost of an unprotected database at the server. Likewise, Barni et al. [12] protect only the probe fingercodes [150] while working with a plaintext database. By reducing the size of the fingercode before encryption, Bianchi et al. [20] are able to gain computational efficiency by trading some biometric accuracy. Combining HE with GCs, Blanton and Gasti [23] proposed an efficient solution for fingercodes as well as iris-codes. However, as their database also is not encrypted, no protection against database leakages is given.

Yang et al. [294] proposed a fingerprint authentication method based on modified minutiae pair representations [289, 295]. They further quantise the features to retrieve binary templates which they encrypt with Paillier [224] for database storage. However, probes are send in plain to the authentication server, which computes the bitwise Exclusive-OR (XOR) of probe and encrypted reference. Despite their two server architecture, the client holds the secret key to decrypt the Hamming weight and authenticate itself. Moreover, the client can send a zero-vector to retrieve a database record.

Using the HE scheme by Gentry and Halevi [103], Yasuda et al. [297] gain efficiency with their newly developed packing method. Hence, computing the Hamming distance (HD) requires less homomorphic operations in the encrypted domain. The authors further improve the packing in a subsequent study [298] based on a particular variant of the ring-LWE (RLWE) assumption [193], which applies for multiple HE schemes [32, 33, 93]. This system computes the HD of 2,048 bit vectors in the encrypted domain within 5.3 ms. Furthermore, Patsakis et al. [227] split iris-codes into blocks before encrypting these with the N-th degree truncated polynomial ring (NTRU) HE scheme [133] and computing the distance in the encrypted domain. Additionally, they benchmark their system in terms of transaction time and communication costs with the RSA-based protocol of [24].

Drozdowski et al. [77] focus on the challenges for HE in the context of biometric identifications. The authors describe an example architecture for face recognition by evaluating two HE schemes. While Cheon-Kim-Kim-Song (CKKS) can directly operate on float templates, Brakerski/Fan-Vercauteren (BFV) first requires an integer quantisation. The results show that computing the Euclidean distance in the encrypted domain is too slow for efficient biometric identification systems. Boddeti [26] found a way to speed up face verifications in the BFV HE scheme. First, the float vectors are rounded to two decimal digits before they are converted to integers. Moreover, the author applies batching in order to execute multiplications on multiple templates within one homomorphic operation. Additionally, principal component analysis [291] can be used to reduce the feature vector to 64 dimensions, thus needing less operations for the distance computation. However, this implies a noticeable performance degradation. Also the secret key is stored at client side, which creates a two-factor authentication system for the user. This is also the case for the system of Kulkarni and Namboodiri [181], who utilised the Boneh-Goh-Nissim (BGN) HE scheme [27] to compare iris-codes and palmprints in the encrypted domain. However, as BGN supports only one multiplication next to additions of ciphertexts, the system needs binary inputs, which can be compared with the HD.

Finally, THRIVE [163] was proposed to protect binary templates in a malicious environment. However, their proof is based on the semi-honest model and hence the security for malicious servers is not given. This problem was recently addressed by Bassit et al. [15] to additionally cover malicious adversaries. Their security is based on the Elgamal HE scheme [82] in combination with zero-knowledge proofs and works for all biometric modalities with fixed-length feature vectors. In order to enhance the efficiency of the biometric recognition, the authors implement lookup tables, where they can store precomputed results for fast access. Hence, the real time execution is maintained while securing the biometric data against malicious adversaries. In another approach, Barni et al. [14] utilise the SPDZ tool by Damgård et al. [66, 67] to protect face and iris authentications against malicious parties. Based on secret sharing, their STPC setup consists of one client and one server. Hence, the references are partly stored at the client and furthermore the authors assume that the client completed the required pre-processing steps (offline phase) before an authentication attempt. The timed online phase takes a fracture of a second and only the final decision is disclosed in plaintext. Besides, malicious attempts can be detected through the usage of message authentication codes [223].

# 3

# FINGERPRINT PRESENTATION ATTACK DETECTION

Research has proven that PAs are a severe threat and that biometric recognition systems accept attack presentations for comparison [21, 125]. Regarding fingerprints, it must be assumed that a skilled attacker is able to cast a good quality fingerprint PAI [106]. However, due to mass fabrication of PAIs in order to collect databases for research, the quality of the fingerprint pattern varies. Hence, the PAD approaches in this Chapter focus only on the detection of attack presentations and do not evaluate whether attacks would successfully match to mated bona fide references. Examples of usable materials and some fingerprint PAIs are included in the appendix (Figure a.1). The following Sections present and discuss own contributions for fingerprint PAD.

## 3.1 FINGERPRINT PAD USING MULTIPLE SENSING TECHNIQUES

The evaluated fingerprint PAD approach follows a hardware-based attempt. Therefore, in a project context related to the scope of this Thesis, a new design for a capture device was developed in cooperation with the project partners[1]. Hence, this capture device is introduced in the following and how the acquired data is expected to assist the detection of attack presentations. Furthermore, the particular methods and algorithms for fingerprint PAD are presented and finally the discussion of the experimental results concludes this Section.

### 3.1.1 *Capture Device and Fingerprint Data*

The idea for the capture device design was to combine the strengths of multiple sensing techniques in order to detect various different PAI species. As already addressed in research question RQ1, additional sensors can provide reliable PAD data, but on the other hand, compatibility with legacy fingerprints is required. The distinct parts of the capture device[2] are illustrated in Figure 3.1. The box contains two cameras and multiple illumination units. As soon as the finger is placed in the open slot, all ambient light is blocked and only the desired illuminations within the box are captured. The first camera (Basler acA1300-60gm) acquires images in the visible (VIS) and NIR domain. Similar to Lin and Kumar [187], a finger photo (260 × 840

1 https://www.isi.edu/projects/batl/overview

2 A more detailed description and illustration is given in [137] as well as in the appendix (Figure a.2).

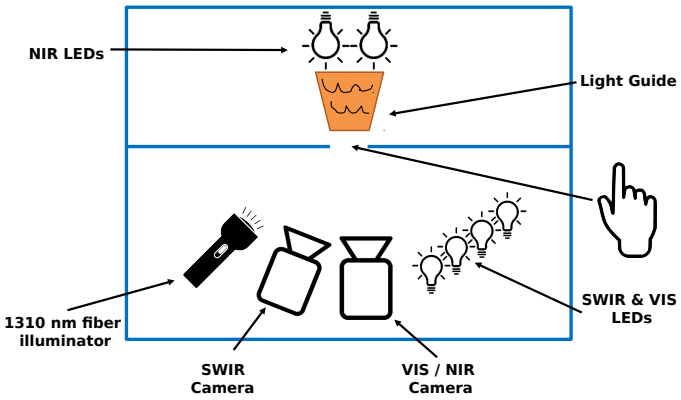



Figure 3.1: The capture device is a closed box with only one free slot for the finger. Two cameras in combination with multiple illuminations are able to capture the fingerprint and additional PAD data.

(a) (b) (c)

Figure 3.2: Fingerprint recognition pipeline for legacy compatibility: (a) original finger photo, (b) segmented fingerprint, and (c) detected minutiae points.

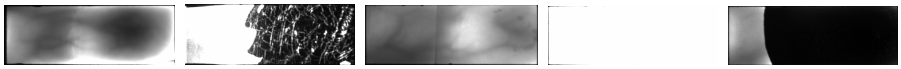
pixels) is captured in the visible domain in order to grant compatibility with legacy fingerprint sensors. The processing pipeline to extract the fingerprint of the first phalanx is shown in Figure 3.2, where Neurotechnology Verifinger SDK[3] was used to segment the fingerprint (b) and locate the minutiae points for comparison (c). The advantages of touchless fingerprint acquisition are no skin deformations, nor latent prints on the screen, and a reduction of hygienic concerns [185]. Samples of different captured PAIs in comparison to a bona fide sample are shown in Figure 3.3.

Inspired by the multi-modal capture device in [235], which acquires fingerprint and finger vein images with the same camera, the capture device also includes a *back-illumination* part. As depicted in the top of Figure 3.1, NIR LEDs (940 nm) are placed above the finger slot and an additional *light guide* concentrates the illumination on the finger in order to prevent diffusion. A wavelength of 940 nm is absorbed by oxygenated veins and a wavelength of 660 nm can reveal deoxygenated ones [243]. In both cases, the light *shines* through the finger and only the veins appear as dark lines. Since classical fingerprint PAIs do not include a vein pattern, these can be easily detected. It should be noted that the subjects are enrolled with their fingerprints only and the vein

3 https://www.neurotechnology.com/verifinger.html

(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.3: Samples in the visible domain.



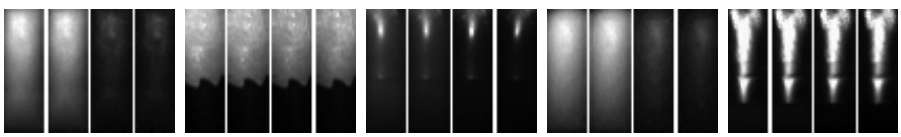
(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.4: Back-illumination samples in the NIR domain.

patterns are not used for comparison but for PAD only. In this context, it is only of interest to check that a vein pattern exists in the image. However, as visible in Figure 3.4, the bona fide vein pattern is still recognisable for thin and transparent overlay PAIs as dragonskin.

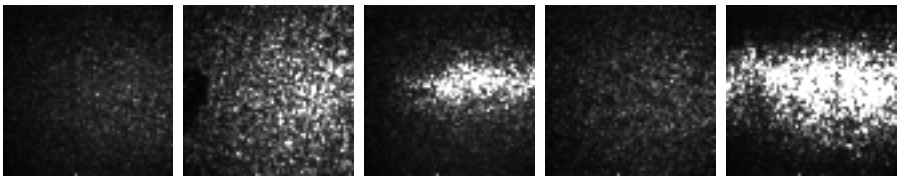
In order to deal with these issues, additional sensors acquire complementary information. In particular, a second camera (Hamamatsu InGaAs, $64 \times 64$ pixels at 1025 fps) is used for the SWIR wavelengths between 1200 nm and 1700 nm. This invisible domain is especially suited for PAD because all skin types in the Fritzpatrick scale [96] reflect in the same way as shown by Steiner et al. [270] for face PAD, but on the contrary PAI species reflect this light quite different from skin [271]. In this context, four SWIR wavelengths are captured with this device: 1200 nm, 1300 nm, 1450 nm, and 1550 nm. Samples across all four wavelengths (left to right) are depicted in Figure 3.5. The Region of Interest (RoI) of the finger slot comprises $18 \times 58$ pixels.

Finally, the Hamamatsu camera additionally captures a laser speckle sequence comprising 1,000 frames with 1310 nm fiber illumination. For this, the lens of this camera is automatically switched in order to zoom into the finger slot (RoI of $64 \times 64$ pixels). Furthermore, a fast steering mirror is utilised to focus three consecutive patches to capture a total of $1 \times 3$ cm of the finger. This technique comes from biomedical applications and is used to visualise and monitor microvascular blood movement of biological tissues. Applications include among others investigations in skin, retina, and neuroscience [258, 288]. For the use case of fingerprint PAD the laser penetrates the skin, producing a random interference effect [35] resulting in a granular pattern of



(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.5: SWIR samples in the four wavelengths 1200, 1300, 1450, 1550 nm.



(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.6: Samples of one laser speckle frame.

dark and bright spots [258]. Figure 3.6 includes one frame of the captured image sequence to show examples of the speckle effect. Depending on the structure and roughness of the surface, variations of the interference pattern are visible. Especially moving scatterers (e. g., red blood cells) make the speckle pattern change over time [25, 258, 288], which can be utilised for fingerprint PAD.

### 3.1.2 *Underlying Concepts for Fingerprint PAD*

This Section introduces the concepts that are used for fingerprint PAD based on the aforementioned capture device and PAD data. These include specific pre-processing, utilised feature extraction methods, and classifiers.

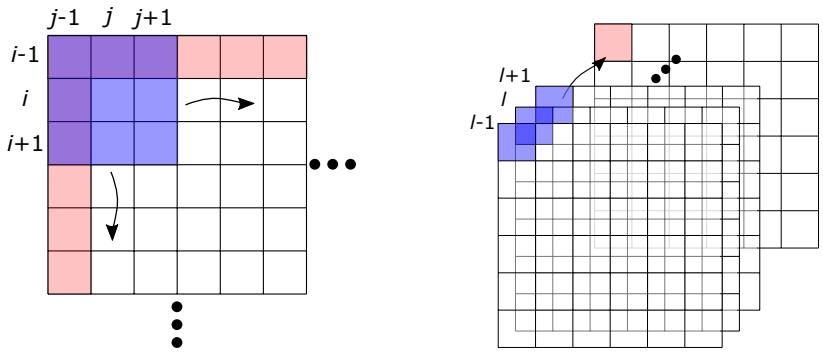
#### 3.1.2.1 *Laser Speckle Contrast Imaging*

The speckle pattern[4] has different intensity values based on the occurring interference [259]. Furthermore, depending on the velocity of flow and the exposure time (1 ms for this capture device), moving scatters appear blurred [35]. The speckle contrast, as the main characteristic of Laser Speckle Contrast Imaging (LSCI), describes the degree of blurring and thus quantifies the pictured movement. In this context, Goodman [123] demonstrates that, given perfect conditions, standard deviation and mean intensity of the speckle pattern are equal. Hence, the speckle contrast $C$ can be defined as the ratio between standard deviation $\sigma$ and mean intensity $\langle I \rangle$:

$$C = \frac{\sigma}{\langle I \rangle} \tag{3.1}$$

The contrast of the captured laser speckle sequence can be analysed either temporally ($C_t$) or spatially ($C_s$) in order to quantify the degree of motion causing the blurring [81]. The calculation for both cases is illustrated in Figure 3.7. Generally, a larger contrasting neighbourhood results in a more accurate contrast estimation [288]. Given the small resolution of $64 \times 64$ pixels and a sequence of 1,000 frames, the temporal contrast $C_t$ is computed for fingerprint PAD approaches. As depicted in Figure 3.7b, this approach quantifies how a single pixel

4 The descriptions within this Section are derived from our publications [167, 175].




(a) spatial contrast (b) temporal contrast

Figure 3.7: Laser speckle contrast calculation for (a) spatial and (b) temporal domain as described in [288].

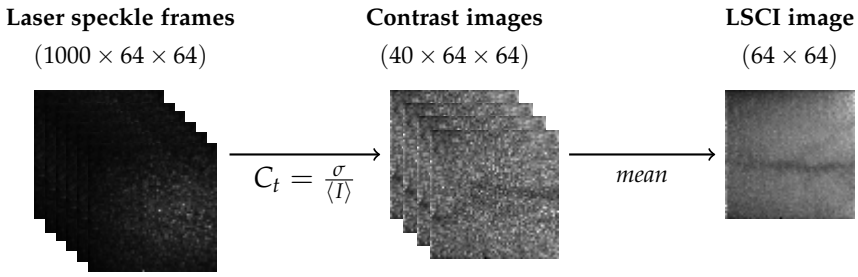
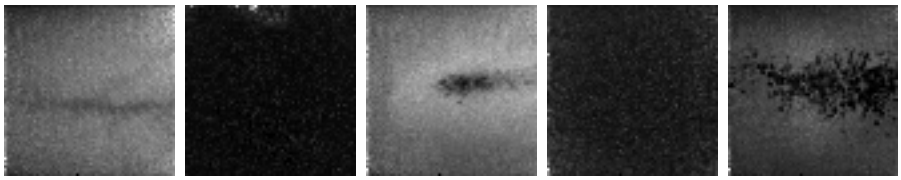



Figure 3.8: LSCI pre-processing pipeline computing the temporal contrast of the captured laser speckle sequence.

changes over the capture time. A pixel without motion has steady intensity values over the time, which results in a low standard deviation and thus a low contrast $C_t$. On the other hand, motion causes varying intensities within the sequence that increase the standard deviation as well as the contrast $C_t$. Based on the findings in [25, 258, 259, 288], the temporal neighbourhood is set to 25 frames. Hence, the temporal contrast $C_t$ is computed for each 25 consecutive laser speckle frames of the captured sequence, thus resulting in 40 contrast images as depicted in Figure 3.8. Subsequently, these 40 intermediate contrast images are averaged in order to combine their information in a single image. For the remaining of this Thesis, this averaged contrast image is simply referred to as LSCI image. Figure 3.9 shows samples of LSCI images that can then be further processed by the PAD algorithms.



(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.9: Samples of LSCI images.

(a) original (b) HIST (c) LBP (d) BSIF (e) HOG

Figure 3.10: Overview of utilised image descriptors.

#### 3.1.2.2 *Handcrafted Feature Extraction*

This Section describes feature extraction methods that are utilised in the following PAD approaches. This is of particular interest for the sub-question of RQ2, regarding pre-processing steps in the combination of PAD data and classifier. As one of the major image characteristics, texture can be analysed by different descriptors [191]. Given the captured 2-dimensional greyscale images (width $\times$ height), the following well-known image descriptors [92] are taken into account for fingerprint PAD. Figure 3.10 provides an overview of the utilised image descriptors based on the bona fide fingerprint captured in the visual domain.

Greyscale histogram (HIST). The most simple image descriptor is the histogram as it only counts the number of occurrences per specified bin. Hence, the greyscale histogram for 8-bit images uses 256 $(2^8)$ bins (i. e., one for every possible value). This analysis can be applied on its own or as the last step of other feature extraction methods in order to obtain a 1-dimensional feature vector for classification.

Local Binary Patterns (LBP) [219]. LBP convinces with computational simplicity as an efficient and greyscale invariant texture descriptor. The process to compute the LBP value for one pixel is illustrated in Figure 3.11. The greyscale values of neighbouring pixels are compared to the central pixel (orange) resulting in a binary value for each position. Those binary values are combined and transformed to a decimal value, which finally replaces the original central pixel in the image. This block (e. g. $3 \times 3$) shifts pixelwise through the whole image and replaces all original pixels (except for the most outer ones). Besides, the binary representation can be flipped and the starting point and reading direction can be changed as LBP operates independently of these factors. However, these settings are required to be identical throughout one system in order to keep the features comparable. Finally, the histogram of the LBP image creates a compressed feature vector for further processing.

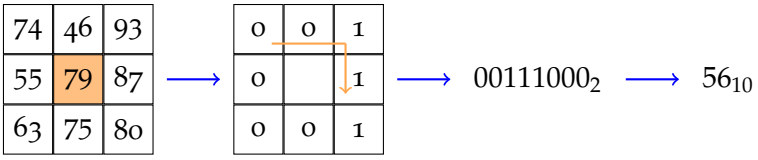
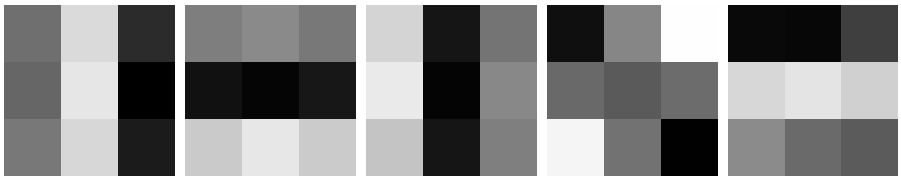
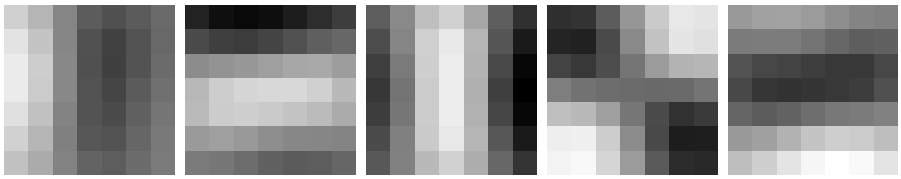



Figure 3.11: LBP computation steps for a $3 \times 3$ window.



(a) Filter 1 (b) Filter 2 (c) Filter 3 (d) Filter 4 (e) Filter 5

Figure 3.12: 5-bit $3 \times 3$ BSIF filters.



(a) Filter 1 (b) Filter 2 (c) Filter 3 (d) Filter 4 (e) Filter 5

Figure 3.13: 5-bit $7 \times 7$ BSIF filters.

Binarized Statistical Image Features (BSIF) [162]. In contrast to LBP, the BSIF filter utilises statistics of natural images instead of heuristic computations. In particular, the pixelwise responses of multiple linear filters applied to the image results in a binary code. Example filters for two sizes are shown in Figure 3.12 ($3 \times 3$) and Figure 3.13 ($7 \times 7$). Each filter of the selected series is subsequently applied to the original image and their responses are combined to retrieve the output image. Finally, a histogram of the binarised result is computed with the number of bins directly corresponding to the number of BSIF filters (e. g., $2^5 = 32$). Kannala and Rahtu [162] additionally provide details on how to create new BSIF filters using independent component analysis [141]. However, training own filters involves the risk of reducing the generalisability compared to the ones proposed by the authors. Thus, in most cases, as in this Thesis, their pre-trained filters [162] are used.

Histogram of Oriented Gradients (HOG) [63]. Extracting features from changes or frequency of information is another non texture-based approach. Using image gradients, HOG characterises spatial structures of the input image by detecting regular patterns within the structures. In particular, direction and magnitude of each pixel are related with previously computed gradients to estimate whether the structure continues or changes. In the next step, the image is divided into a grid of even cells (e. g. $16 \times 16$ pixels) and a HOG is computed for each cell separately. Then, overlapping blocks of multiple cells are created to normalise the HOGs and improve global accuracy. The final results are flattened to a 1-dimensional feature vector. One advantage of HOG towards the previously introduced texture descriptors is its contrast invariance.

All presented feature extraction methods are applied on greyscale images and return a 1-dimensional feature vector. Hence, efficient classification is supported by these compact features.

#### 3.1.2.3 *Classifiers*

According to research question RQ2, different machine learning classifiers [228] are benchmarked against each other. These classifiers for handcrafted features are briefly introduced in this Section[5].

Support Vector Machine (SVM) [60]. As a binary classifier, the SVM is especially suited for PAD tasks. SVMs map the input data to a high-dimensional feature space and define a hyperplane during training that allows separation of both classes. Hence, predictions are efficiently achieved by mapping the test sample into the feature space and comparing it to the hyperplane.

K-nearest neighbours (KNN) [61]. During training, the features are grouped into clusters (two in the case of PAD). Unknown test samples are mapped into the same feature space and the classifier checks its $k$ nearest neighbours to predict the corresponding class. In order to apply simple majority voting, uneven values are used for $k$.

Decision Tree (DT) [233]. In a DT, class labels are stored in the leafs and the branches contain the conjunctions to reach the conclusion. Given the binary PAD case, all leafs contain either labels for bona fide presentations or attack presentations. DTs are still efficient for huge datasets since they need scarce data preparation. However, as little alterations in the training set cause largely different structures, DTs do not generalise well on unseen data.

Random Forest (RF) [131]. Building a forest of multitude DTs, RFs aim to amend the missing robustness by averaging several DTs. In order to prevent over-fitting, all DTs are trained on deviating, in some parts overlapping, sets of the training data. By design, this composition generalises much better than a single DT, which is a desired property for PAD approaches as unknown PAI species must always be anticipated.

AdaBoost Classifier (ADA) [300]. As the RF, this classifier also trains multiple instances, but uses the full training set in all cases. The first model is evaluated before fitting additional copies with adjusted weights for misclassified samples. Hence, the subsequent instances learn to correctly distinguish more difficult cases and improve the general performance.

Gaussian Naive Bayes (GNB) [47]. The GNB classifier requires only small training sets for parameter adjustment and convinces with an efficient execution time. The prediction takes into account previous observations using a conditional decision chain and the possibility to update model parameters after deployment is the main advantage of the GNB classifier.

5 The descriptions within this Section are derived from our publication [175].

Stochastic Gradient Descent (SGD) [29]. This method is a combination of linear classifiers and SGD training, during which the gradients are evaluated by shuffling the samples for each epoch. Additionally, decreasing the learning rate prevents over-fitting while processing single training samples at a time. Several different loss functions can be selected to fit the model.

Linear Discriminant Analysis (LDA) & Quadratic Discriminant Analysis (QDA) [184]. These classifiers either come with linear or quadratic decision boundaries and are created by applying Bayes' theorem [158]. Both approaches fit a Gaussian density to each class and for classification of new test data, these class conditional densities are employed.

#### 3.1.2.4 *Convolutional Neural Networks*

The rise of deep learning [122] within the last years correlates with increasing processing power auf GPUs as well as the amount of available data. These facts allow the training of deep architectures that are able to outperform traditional methods. Thus, deep learning techniques are deployed, among other areas, for biometrics in the fields of fingerprint [277], face [73, 256], iris [214], ocular [241], signature [281], and speaker [59] recognition. In fact, even use cases with scarce data can benefit from deep learning. In this case, pre-trained models are fine-tuned on a different task (i. e. fingerprint PAD), an approach called transfer learning [276][6].

Convolutional NNs (CNNs) are one of the most successful architectures in the area of deep learning when it comes to image classification. Its convolutional layers extract different patterns as e. g. horizontal and vertical edges, while pooling layers are added to achieve invariance towards little alterations of the input data to aid the pattern extraction. In the area of fingerprint PAD, CNNs have a twofold use as they can be trained to extract features from input images or additionally train a prediction layer for classification (end-to-end). In this work, only the latter case is used to directly predict the class of the given sample. A benchmark between both scenarios is done in [279], where the extracted features are classified with SVMs and result in a worse performance than the end-to-end CNNs.

In this context, two CNNs are evaluated for fingerprint PAD within this Thesis since they combine both parts of research question RQ2. VGG19 [260] was pre-trained for the *ImageNet* challenge [72, 250] and is fine-tuned on the fingerprint data. Therefore, the top layers (i. e., prediction block) are replaced to support a binary decision between bona fide presentations and attack presentations. Furthermore, the weights of the last convolutional block are re-trained together with the newly added prediction block. In fact, more general features are extracted by the first layers while the last ones extract more abstract

6 This Section is based on the descriptions within our publications [112, 280].

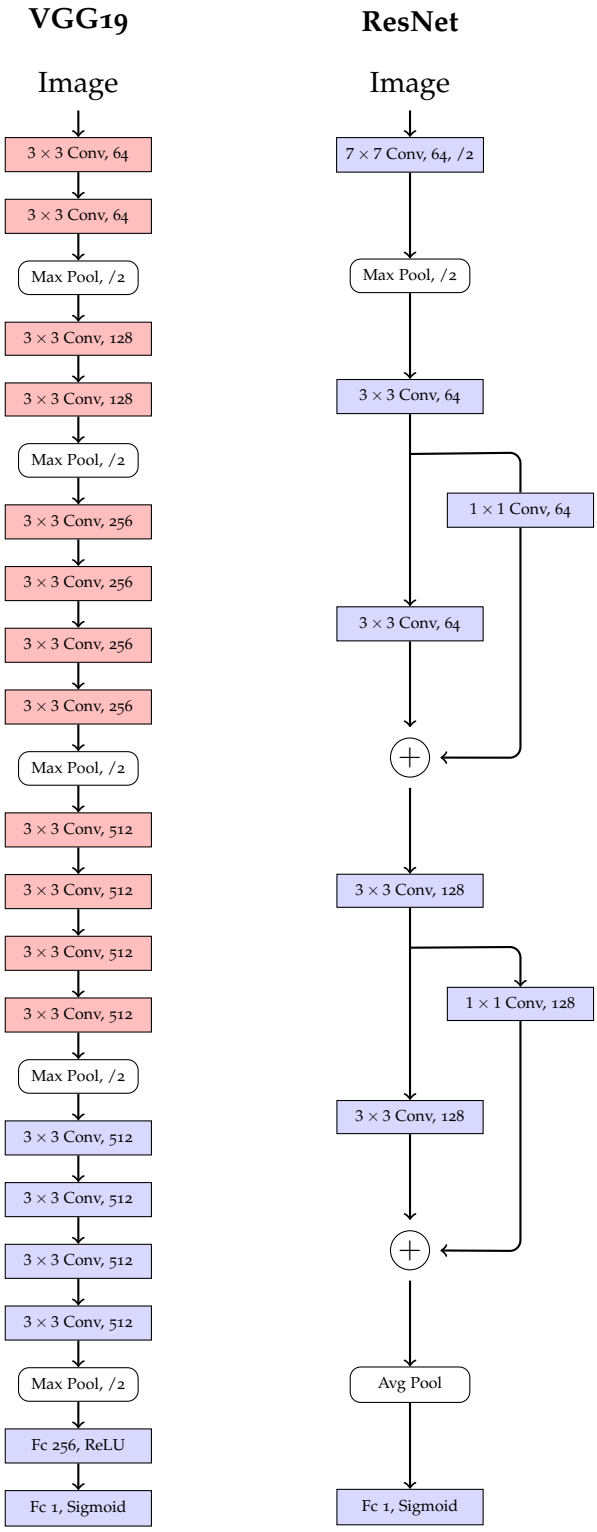



Figure 3.14: CNN architectures of the fine-tuned VGG19 and the developed residual CNN ResNet. Trainable layers are coloured in blue and red marks pre-trained layers that keep original weights.

features that are already specific for the particular task. In addition, re-training the whole CNN with such a small database (Section 3.1.7) would most likely result in over-fitting, which is not beneficial for PAD tasks. The second CNN is a small self-developed network with residual connections [273], which will be referred to as ResNet. The small size of six layers allows training from scratch without the risk of over-fitting. Additionally, in contrast to plain networks, residual connections combine the output of one layer with the output of the subsequent layer. These shortcuts make deeper architectures possible and reduce the training time significantly [128]. Following the approach of [142], each convolutional layer is followed by batch normalisation and Rectified Linear Unit (ReLU) activation and for the final decision a fully connected layer including Sigmoid activation is added. For both CNNs Adam optimizer [172] with a learning rate of 0.0001 and binary cross-entropy loss function is chosen. The architectures of both CNNs are shown in Figure 3.14, highlighting the layers that are trained on the fingerprint data.

### 3.1.3 *Vein-based Fingerprint PAD Methods*

The vein samples in Figure 3.4 indicate that some PAI materials completely block the NIR light producing a black image (e. g. silicone), while others appear transparent for 940 nm illumination, resulting in a blinded camera and a white capture (e. g. playdoh)[7]. Hence, the most simple vein-based PAD method defines two thresholds $\left(\delta^{min}, \delta^{max}\right)$ to classify the average greyscale value of the back-illumination sample. Captures that are either too bright or too dark to stem from bona fide presentations can thus be detected with this *Luminosity* PAD method. In order to allow compliance with other PAD scores that rely on a single decision threshold, the obtained average greyscale value $g \in [0, 255]$ (given 8-bit images) is normalised to a PAD score $s_{lum} \in [0, 100]$ as follows:

$$s_{lum} = \frac{100}{255} \cdot \left(g - \delta^{min} \mod 256\right) \tag{3.2}$$

And a single decision threshold $\delta_{lum} \in [0, 100]$ can be computed as:

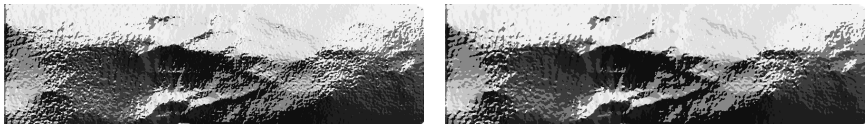
$$\delta_{lum} = \frac{\delta^{max} - \delta^{min}}{255} \cdot 100 \tag{3.3}$$

However, PAI species that only partially block the NIR illumination are not detected as the average luminosity is within the thresholds. Hence, a more advanced PAD method analyses whether finger vein patterns can be extracted with the Maximum Curvature (MC) [206] algorithm that is commonly used for finger vein recognition [286]. The implementation within the freely available Bob toolbox[8] [10, 11] is

7 This Section is based on our publications [111, 177].
8 https://www.idiap.ch/software/bob/

(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.15: Vein patterns that are extracted by MC.



(a) 1st Pyramid level (b) 6th Pyramid level

(c) 11th Pyramid level (d) 16th Pyramid level

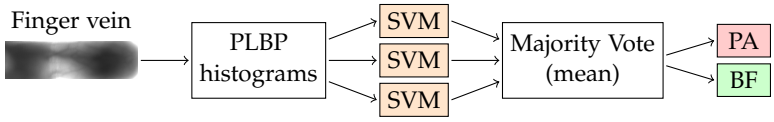
Figure 3.16: Increasing blurriness of PLBP images for higher pyramid levels.

used to obtain binary images showing only the extracted vein pattern. The examples in Figure 3.15 indicate that MC completely fails to extract a pattern for particular PAI species (e. g. playdoh) and creates random patterns for other materials (e. g. 3D printed and silicone). On the other hand, bona fide samples and dragonskin overlays result in a sound vein skeleton.

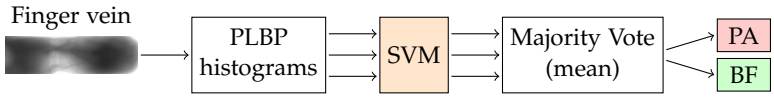
Identically to the luminosity, the mean value of these binary image can be compared to two thresholds to derive a fast decision and filter obvious attack presentations. The only difference towards the computations in Eq. (3.2) and Eq. (3.3) is that MC creates 1-bit images instead of 8-bit images. This vein-based PAD method is referred to as *MC mean*. In addition, a SVM is trained on the column-wise histograms of the MC image. Since finger veins are located along the direction of the finger (horizontal in the captured images), vertical lines indicate the presence of a PAI. Hence, the *MC hist* PAD method adds up the pixels of each vertical column and the histogram of all sums is used as input vector for the SVM.

Furthermore, the texture of the captured vein samples is analysed using a combination of Gaussian Pyramids and LBP (PLBP) [232], which was successfully applied for fingerprint PAD on the LivDet 2013 dataset [105] by Jiang and Liu [157] using three pyramid levels. PLBP allows to extract texture information (LBP features) from multiple hierarchical spatial pyramids considering different resolution levels, which tends to increase its robustness. In this regard, the PLBP method is benchmarked against the classical versions of LBP and BSIF.

Gaussian pyramids are widely used for multi-resolution image analyses [220]. By repeatedly down-sampling the input image through applying a Gaussian blur lowpass filter, a series of consecutive smaller




(a) Separate SVM approach for three pyramid levels.




(b) Single SVM approach for three pyramid levels.


(c) Basic LBP pipeline.


(d) Basic BSIF pipeline.

Figure 3.17: Tested texture-based PAD methods. PLBP example case of three pyramid levels for separate (a) and single (b) SVM approaches in contrast to basic LBP (c) and BSIF (d) feature extractions.

images is obtained. This process compresses image information as a fixed-size area is replaced by one pixel in its smaller successor. When aligning all images from small (top) to large (bottom) it resembles a pyramid. For this PAD method up to 16 pyramid levels are evaluated in order to find the best-suited setting. Moreover, the resulting images are up-sampled again to the original image size to extract feature vectors of identical lengths for classification. Thus, higher pyramid levels seem blurrier due to the information loss as depicted in Figure 3.16.

In accordance with the normal LBP process, the greyscale histograms from each PLBP image are computed and serve as input for the classifier. Furthermore, two SVM setups are benchmarked to either train separate instances for different pyramid levels or to train one common SVM on the input from all pyramid levels. Both use cases combine their multiple outputs through majority voting as illustrated in Figure 3.17 for three pyramid levels. Additionally, the pipelines for basic LBP and BSIF processing are shown for completeness. All SVMs produce binary results and are trained using cross-validation following the guide in [136] to automatically search for the best parameter setting. Besides, creating the Gaussian pyramid as well as feature extraction are done with the publicly available Bob toolkit [10, 11].

### 3.1.4 *SWIR Fingerprint PAD Methods*

Based on the acquired SWIR data, this Section[9] presents handcrafted as well as deep learning algorithms for fingerprint PAD.

As proposed in [110], the spectral signature of fingerprint images captured in the SWIR domain can be used for fingerprint PAD. This method exploits the fact that all skin colours reflect SWIR illumination in a very similar way while most PAI materials look differently, which was shown in [271] (here: Figure a.3). As visible in Figure 3.5, for bona fide samples the greyscale intensity decreases for increasing SWIR wavelengths. On the other hand, the reflections remain stable for particular PAI species (e. g., 3D printed, dragonskin, or silicone). This fact was first utilised by Steiner et al. [270] for wavelengths between 935 nm and 1550 nm to discriminate each pixel of facial images whether it shows skin or no skin (e. g., hair, make-up, or PAI). Due to the pixel-wise classification, even partly covered areas can reliable be detected. Furthermore, since each pixel results in a feature vector, the used SVM can be successfully trained on small datasets. Given the four captured wavelengths ($\lambda_1 = 1200$ nm, $\lambda_2 = 1300$ nm, $\lambda_3 = 1450$ nm, $\lambda_4 = 1550$ nm), the spectral signature $ss$ of a pixel at location $(x, y)$ is computed from the intensity values of the corresponding wavelengths:

$$ss\,(x,y) = (i_{\lambda_1}, i_{\lambda_2}, i_{\lambda_3}, i_{\lambda_4}) \tag{3.4}$$

The final spectral signature $ss$ comprises six differences $d$ from all possible wavelengths combinations as given in Eq. (3.5):

$$d\,[i_{\lambda_1}, i_{\lambda_2}]\,, d\,[i_{\lambda_1}, i_{\lambda_3}]\,, d\,[i_{\lambda_1}, i_{\lambda_4}]\,, d\,[i_{\lambda_2}, i_{\lambda_3}]\,, d\,[i_{\lambda_2}, i_{\lambda_4}]\,, d\,[i_{\lambda_3}, i_{\lambda_4}] \tag{3.5}$$

whereas a single normalised distance $d$ is defined as:

$$d\,[i_a, i_b] = \frac{i_a - i_b}{i_a + i_b}. \tag{3.6}$$

Computing the normalised signature is desired in order to counter possible changes in the illumination and to focus on the trend across different wavelengths instead of absolute brightness values. To support this, only the central area ($18 \times 20$ pixels) of the SWIR images is used as highlighted in Figure 3.18 since the top and bottom of the finger slot are not equally illuminated. Finally, for each pixel ($18 \times 20 = 360$) with coordinates $(x, y)$ six normalised differences $d$ across all wavelengths are calculated to obtain 360 spectral signature feature vectors. The number of classified non-skin pixels proportional to all pixels result in the final PAD score of that specific sample. Following the definition in ISO/IEC 30107-2 [148], PAD scores close to zero correspond to bona fide presentations and higher scores close to 100 indicate attack presentations.

9 This Section is based on our publications [110, 112, 280].

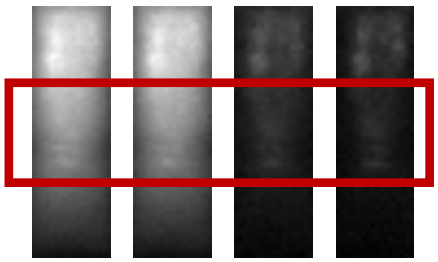
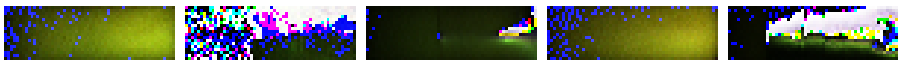


Figure 3.18: Used RoI of the SWIR frames to compute the spectral signature.



(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.19: RGB images created from all four SWIR wavelengths.

In addition to the handcrafted PAD method, CNNs combine feature extraction and classification in one instance. Through deep learning they can be trained to extract those features that are relevant for classification. For this problem of fingerprint PAD, two approaches[10] are evaluated: *i)* fine-tuning the pre-trained VGG19 [260] by applying transfer learning, and *ii)* training the small ResNet from scratch. Both CNNs are applied on the SWIR data, which requires further pre-processing in order to combine the four SWIR wavelengths to one 3-dimensional RGB image.

Those four wavelengths were selected because all skin types [96] reflect SWIR illumination in the same way [271], thus providing low intra-class variations for bona fide samples. It is of interest to maintain this property, which is valuable for the PAD task, while creating the RGB image. Furthermore, the inter-class variance between attack presentations and bona fide presentations should be as large as possible to aid correct classification. After an empirical evaluation of different combinations, the RGB image $I$ can be computed as follows:

$$I\,(R,G,B) = (|\lambda_4 - \lambda_1|, |\lambda_4 - \lambda_2|, |\lambda_4 - \lambda_3|) \tag{3.7}$$

Figure 3.19 shows examples of those RGB images. In contrast to the spectral signature approach, the CNNs work on the full image. As orange playdoh resembles bona fide skin in the SWIR domain (Figure 3.5), also the crafted RGB images look similar. On the other hand, the other attack presentations are easily separable from bona fide presentations. Those RGB images can then be used as input for the CNNs as illustrated in Figure 3.20. In contrast to the handcrafted classifiers that produce binary decisions, the CNNs yield a floating point PAD score in the range $[0,1]$. As a result, an additional PAD threshold is required for final classification.

10 This Section is based on our publications [112, 280].

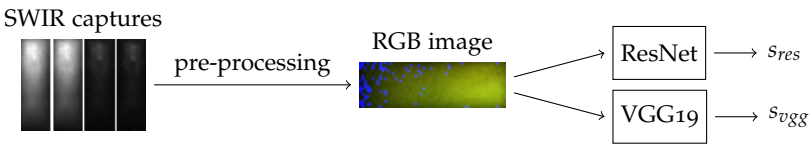



Figure 3.20: PAD pipeline for SWIR CNNs.

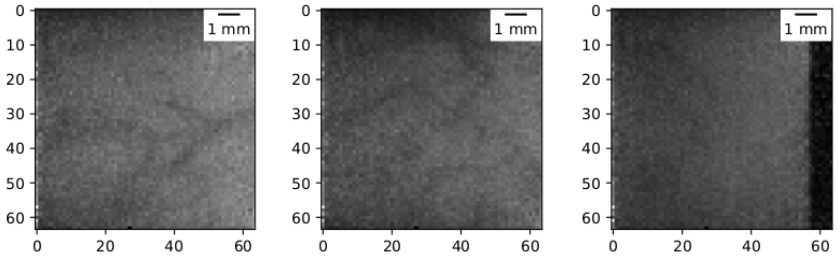
### 3.1.5 *LSCI Fingerprint PAD Methods*

The LSCI PAD methods[11] consist of three steps: pre-processing, feature extraction, and classification. The pre-processing that computes the temporal contrast from the captured laser speckle sequence is described in Section 3.1.2.1 and the feature extraction methods directly work on the resulting LSCI image. Analysing the LSCI images from all three captured areas (Figure 3.21) revealed the presence of noise in the first area since 1 mm $\left(\frac{1}{10}\right)$ of the frames are covered by the edge of the finger slot. Hence, this part shows no blood movement even for bona fide samples which troubled the classification in [167], such that the following experiments focus on the second and third area and ignore the noisy capture.

Through computing the temporal contrast for the LSCI image, high intensity values are a tangible indication for motion. Hence, the Greyscale histogram (HIST) extracts features based on the brightness alone, which yields heavier right tails for bona fide presentations and heavier left tails for PAI species without motion. This can directly be used for an additional pre-selection that compares the position of the histogram peak to a threshold to easily filter the most obvious attack presentations. In this regard, a very conservative threshold (50) is chosen in order to maintain a convenient BPCER. Example histograms with peaks (green) and thresholds (red) are depicted in Figure 3.22.

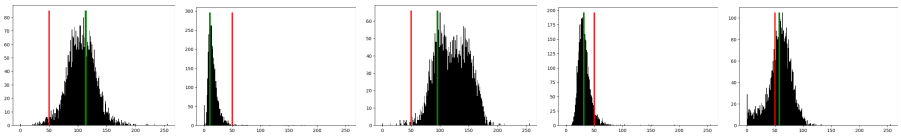
Furthermore, the LSCI images from multiple PAI species appear more rough compared to the smooth texture of bona fide presentations. LBP and BSIF are used to extract these more robust texture-based features as both showed good performance on other fingerprint PAD

11 This Section is based on our publications [167, 175].




(a) area 3 (b) area 2 (c) area 1

Figure 3.21: LSCI images from all three captured areas.



(a) bona fide (b) 3D printed (c) dragonskin (d) playdoh (e) silicone

Figure 3.22: Histograms of LSCI images. The pre-selection threshold is marked in red and the peak bin in green.

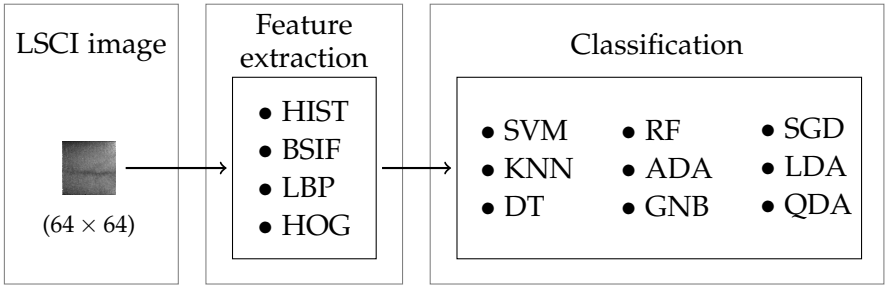



Figure 3.23: LSCI PAD pipeline. After the pre-processing, each feature extraction method is combined with each classifier.

tasks [104, 293]. Finally, in some bona fide LSCI images a horizontal vein pattern is visible (Figure 3.9), which can be recognised by HOG. Figure 3.23 summarises this PAD pipeline: four feature extractors are combined with nine classifiers, resulting in a benchmark of 36 PAD algorithms (+1 optional peak pre-selection). In this regard, separate entities of the classifiers are trained for each different feature set to produce binary decisions and cross-validation is used to find the best parameters for each classifier.

However, this benchmark creates too many individual results for a fusion with the other sensing techniques. Hence, the best scores are first internally fused to retrieve a final LSCI PAD score which is then combined with further PAD scores. In this context, multiple fusion schemes have been evaluated [111, 112, 167, 175], which can be grouped into two categories as illustrated in Figure 3.24: cascade decision and majority voting. The cascade provides more security as attack presentations need to pass each algorithm individually but leads to higher BPCERs as errors sum up. On the contrary, majority voting maintains a low BPCER for a slightly higher APCER. In order to fuse complementary information and prevent over-fitting on a specific data partition, each extracted feature is only used once but particular classifiers have no limit. Furthermore, both designs can easily be adjusted to e. g. additionally include the peak pre-selection or remove a feature set in case for decreasing PAD performance.

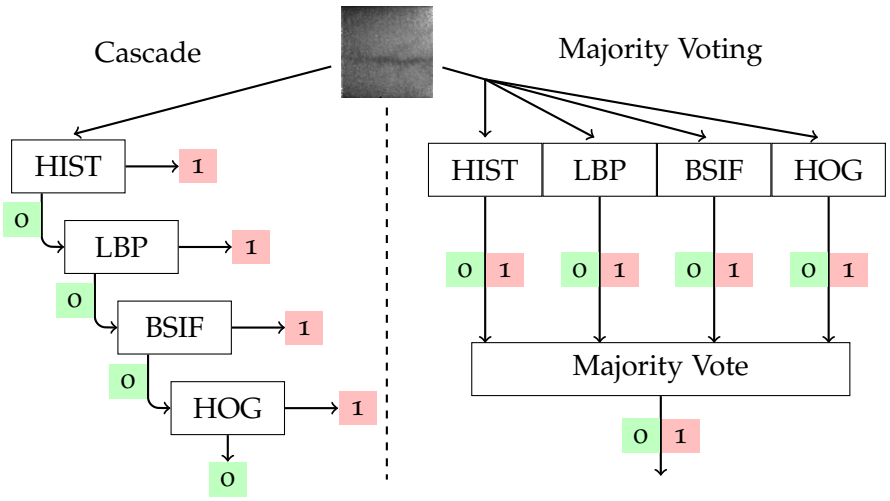



Figure 3.24: LSCI fusion schemes. Either cascading or majority decisions yield the final PAD score.

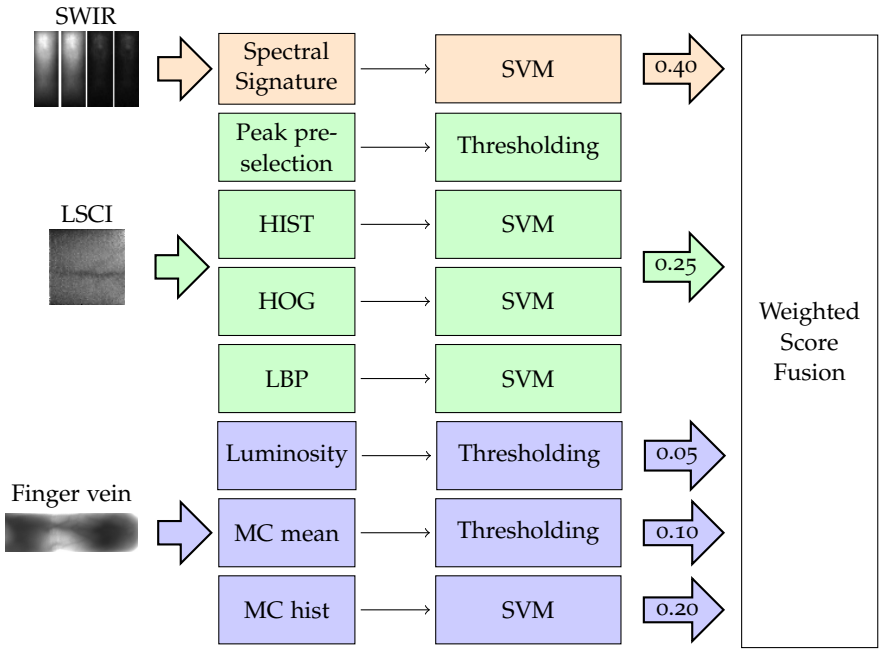
### 3.1.6 *Fused PAD Methods*

In addition to the stand-alone PAD methods, fusion schemes combine the strengths of the different complementary sensing technologies. In accordance with the LSCI fusion analysis, a cascading structure is not desired since it adds up all errors and leads to higher BPCERs. Moreover, a simple majority voting is only beneficial for equal shares of PAD algorithms with identical strengths. As a result, the PAD scores are fused using a weighted score fusion, thus taking into account the individual performances. It is important to note that all fusion weights are optimised on the development / validation dataset and the reported results are obtained from the unseen test set.

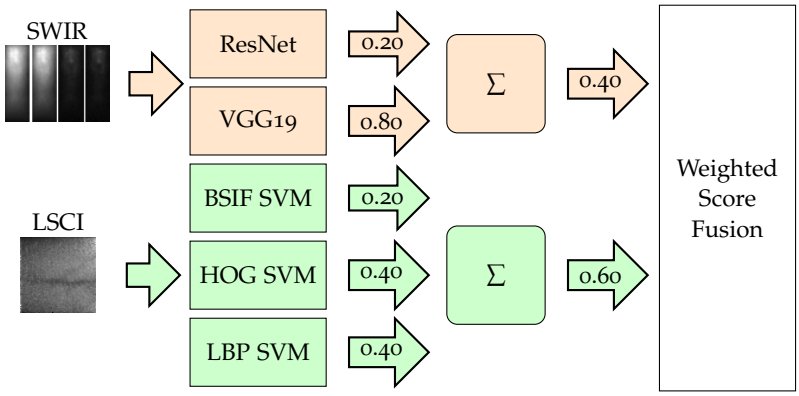
Since the PAD algorithms were partly trained on different partitions, not all possible combinations can be evaluated for the fusion. Hence, the presented fusions align with previous published results as the PAD methods as well. Beginning with handcrafted PAD algorithms only, *Fusion I* [111] combines SWIR spectral signatures, LSCI peak pre-selection, HIST, HOG, and LBP SVMs, and vein luminosity, MC mean, and MC hist. *Fusion II* [112] connects handcrafted and deep learning architectures using LSCI BSIF, HOG, and LBP SVMs and the SWIR CNNs ResNet and VGG19. A summary of all fusions is shown in Figure 3.25.

### 3.1.7 *Database*

The previously described capture device was used for three data collections that are summarised in Table 3.1. In addition to bona fide presentations, attack presentations include full fake finger as well as more challenging fingerprint overlays made from various




(a) Fusion I [111].




(b) Fusion II [112].

Figure 3.25: Summary of fused PAD algorithms. The original fusion weights from [111, 112] are kept.

| DB | # subjects | # PAI species | # BF samples | # PA samples |
|---|---|---|---|---|
| A0 | 4 | 14 | 63 | 58 |
| A1 | 163 | 32 | 547 | 226 |
| A2 | ≈ 340 | 8 | 3743 | 216 |

Table 3.1: Summary of the collected datasets showing the number of subjects, PAI species, bona fide (BF), and PA samples.

materials. Examples of both PAI types are shown in the appendix (Figure a.1). DB A0 was mainly used for a first evaluation and final adjustments of the capture device. Moreover, it was the basis for PAD development as this was the first capture device that combines this specific combination of sensing techniques for data collection. Hence, the intention of the next acquisition (DB A1) was to collect a broad range of PAI species in order to analyse and improve the PAD performance. Finally, DB A2 focusses on the most challenging PAI species in addition to collecting a quantity of bona fide presentations that is required to evaluate whether the PAD algorithms are suited to operate in a convenient scenario (i. e. BPCER=0.2%). Achieving a low APCER is useless when on the other hand most bona fide presentations would be rejected. DBs A1 and A2 are compatible, resulting in 35 different PAI species as five are identical in both datasets.

A detailed listing of all captured PAI species with their number of samples and variations is given in Table 3.2. Additionally, the specific PAI species are divided into *Fakefinger* and *Overlay* groups.

For privacy reasons, the datasets include no information about the originating fingerprints that are used to facilitate the PAIs. Hence, this Thesis provides no experimental evaluation whether the attack presentations would successfully match to the enrolled reference[12]. Furthermore, it should be noted that the consent form for DB A1 collection does not allow sharing of the data. Therefore, the PAD algorithms could only be remotely evaluated at the site of the data controller during the project runtime. Since a large part of the data is unavailable, it was not possible to benchmark the developed PAD algorithms on a unified partition for this Thesis. Instead, the presented results are the ones reported in our own publications.

### 3.1.8 *Experimental Protocol*

All experiments use disjoint training, validation, and test sets to grant a fair evaluation on unseen data. Furthermore, a similar number of bona fide presentations and attack presentations is used during training with the aim to avoid creation of biased systems.

12 Only the project organisers had access to these information and evaluated this case.

| PAI Group | PAI | # samples (# variations) | | |
|---|---|---|---|---|
| | | **A0** | **A1** | **A2** |
| Fakefinger | 3D printed | 4 (2) | 33 (2) | – |
| | dragonskin | – | 33 (4) | – |
| | ecoflex | 6 (1) | 35 (4) | – |
| | gelatine | 6 (1) | – | – |
| | latex | – | 8 (1) | – |
| | playdoh | 3 (1) | 28 (9) | – |
| | silly putty | 6 (2) | 15 (3) | – |
| | wax | 4 (2) | 6 (1) | – |
| Overlay | dragonskin | 1 (1) | 11 (1) | 27 (2) |
| | ecoflex | 2 (1) | 13 (2) | 49 (1) |
| | glue | – | 6 (1) | – |
| | latex | – | 10 (1) | – |
| | printout | 24 (2) | 18 (2) | 37 (2) |
| | silicone | 2 (1) | 6 (1) | 103 (3) |
| | wax | – | 4 (1) | – |

Table 3.2: Summary of the PAIs in the datasets including the number of samples and number of variations. Variations include e. g. different colours and conductive augmentations.

Although some publications [110, 280] used the small dataset DB A0, those results are excluded from this Thesis as both PAD algorithms were also evaluated on other datasets. Thus, these more significant results are reported instead. As already mentioned before, the PAD algorithms could not be trained on a unified partition, hence Table 3.3 lists all PAD algorithms that are trained on a particular partition. In addition, a detailed description with the specific number of samples per partition is given in Table 3.4. As the vein partitions are limited to DB A1, the total number of samples is lower than for the hand-crafted or deep learning partition. Furthermore, the vein algorithms are evaluated on two scenarios, first with identical numbers of attack presentations and bona fide presentations in the training set and second with slightly more bona fide samples. The idea behind the second scenario is to reduce the BPCER in order to achieve a convenient operation point. Training only on 69 samples (scenario 1) might not generalise well due to the variations within bona fide vein samples, which can occur for different captured fingers (e. g., thumb vs. little finger). Another property of all partitions is that samples acquired from one subject are forbidden to appear in different sets, hence all samples of one subject are in the same set (i. e., training, validation,

| DB A1 [177] (vein partition) | DBs A1 + A2 [111] (handcrafted partition) | DBs A1 + A2 [112, 175] (deep learning partition) |
|---|---|---|
| vein Luminosity | SWIR spectral signature | SWIR ResNet |
| vein MC mean | LSCI peak | SWIR VGG19 |
| vein MC hist | LSCI HIST SVM | LSCI full benchmark* |
| vein PLBP | LSCI HOG SVM | Fusion II |
| vein LBP | LSCI LBP SVM | |
| vein BSIF | vein Luminosity | |
| | vein MC mean | |
| | vein MC hist | |
| | Fusion I | |

* Includes all four feature extraction methods in combination with all nine classifiers.

Table 3.3: List of PAD algorithms that are evaluated on particular partitions.

or test). As a result the number of attack presentations and bona fide presentations are not identical for the handcrafted training partition. The biggest training and validation sets are used for the deep learning partition, which generally requires more samples than handcrafted algorithms.

### 3.1.9 *Experimental Results*

This Section[13] presents the results of the different experiments and aligns the structure with the described data partitions.

#### 3.1.9.1 *Vein Partition*

Based on the two scenarios of the vein partition, the following six fingerprint PAD methods are evaluated: Luminosity, MC mean, MC hist, PLBP, LBP, and BSIF. In contrast to the other algorithms, multiple settings are tested for the PLBP algorithm including two SVM approaches and different numbers of computed pyramid levels. The results of these tests are plotted in Figure 3.26. Due to the binary output of the SVM, each run results in a fixed pair of APCER and BPCER. The plots show how the error rates (%) change for different pyramid levels. The single SVM approach for scenario 1 (Figure 3.26a) achieves a minimum for both error rates at pyramid level 6 with BPCER = 3.38% and APCER = 5.81%. However, both have inverse peaks at this level in a general changing appearance with many peaks and valleys. Despite the fact, that the separate SVMs (Figure 3.26b) reach minimum error rates at different pyramid levels (BPCER = 2.54% at level 5 and APCER = 6.45% at level 4), the curves stabilise much more from the

13 This Section is based on our publications [111, 112, 175, 177].

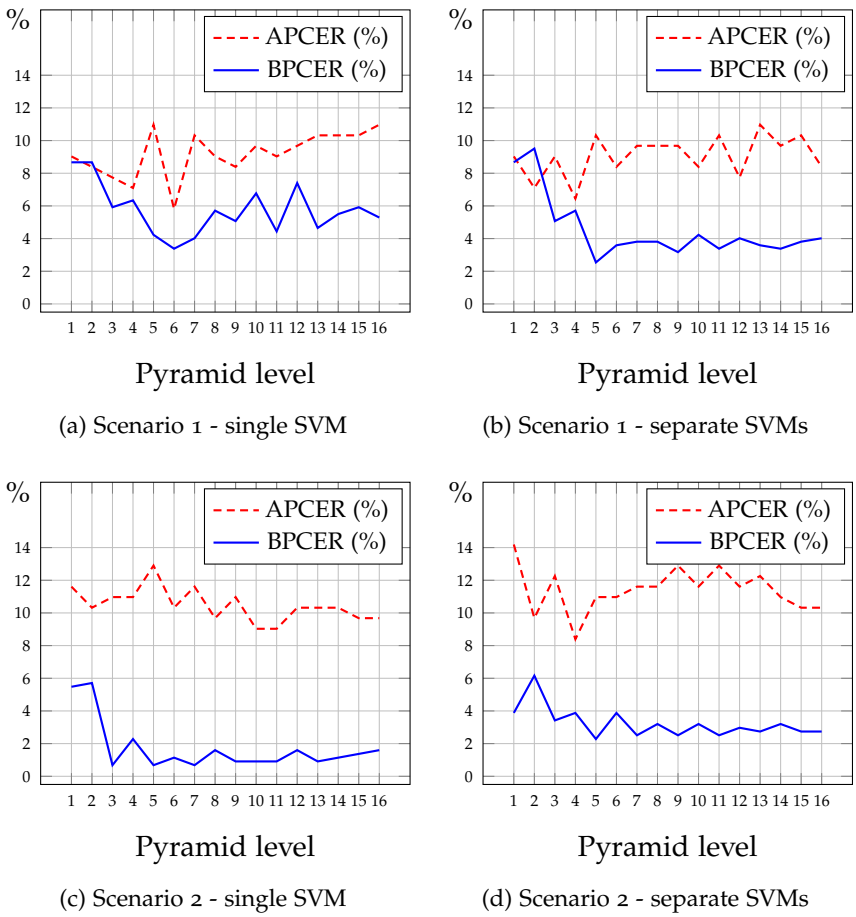
| Name | Set | # PA samples | # BF samples |
|---|---|---|---|
| vein - scenario 1 | Training set | 69 | 69 |
| | Test set | 155 | 473 |
| vein - scenario 2 | Training set | 69 | 104 |
| | Test set | 155 | 438 |
| handcrafted | Training set | 70 | 66 |
| | Validation set | 32 | 32 |
| | Test set | 341 | 4190 |
| deep learning | Training set | 130 | 130 |
| | Validation set | 90 | 90 |
| | Test set | 222 | 4071 |

Table 3.4: Specifications of the used dataset partitions.

| Algorithm | Scenario 1 | | Scenario 2 | |
|---|---|---|---|---|
| | APCER | BPCER | APCER | BPCER |
| Luminosity | 68.39 | 0.00 | 68.93 | 0.00 |
| MC mean | 43.87 | 0.21 | 43.87 | 0.23 |
| MC hist | 13.55 | 9.51 | 12.90 | 8,22 |
| BSIF | 28.39 | 5.71 | 26.45 | 4.57 |
| LBP | 10.32 | 1.90 | 11.61 | 1.14 |
| PLBP (lvl 7) | 10.32 | 4.02 | 11.61 | 0.68 |

Table 3.5: PAD results on the vein partition.

$6^{th}$ level upwards in contrast to the first test. Using more bona fide samples for training in scenario 2 reduces the BPCERs for both SVM setups significantly while APCERs remain similar, thus confirming the assumption for the second scenario. The minimum values for the BPCER are 0.68% for the single SVM (Figure 3.26c) and 2.28% for separate SVMs (Figure 3.26d). Furthermore, the first four pyramid levels show bigger peaks and higher error rates compared to the following levels. Hence, for PAD the latter ones are more interesting with an average BPCER of 1.12% (single) and 2.87% (separate), respectively. In addition to a lower BPCER, also the average APCER is lower for the single SVM setup: 10.32% to 11.50% for the separate SVMs. The single SVM generalises better than separate ones since it analyses complementary information from all used pyramid levels, while these samples are split in the other case.




(a) Scenario 1 - single SVM

(b) Scenario 1 - separate SVMs

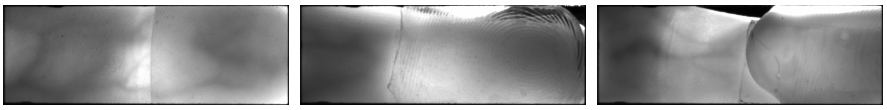
(c) Scenario 2 - single SVM

(d) Scenario 2 - separate SVMs

Figure 3.26: PLBP PAD results depending on the number of pyramid levels for both SVM approaches and both scenarios.

Depending on the use case (i. e., is a low APCER or a low BPCER required), different numbers of pyramid levels can be chosen. Focussing on convenience, level seven is selected for the benchmark with the other vein-based fingerprint PAD methods. The results from all vein algorithms for both scenarios are summarised on Table 3.5. The lowest BPCER values are achieved by the threshold-based methods luminosity and MC mean, which have by design a very convenient operation point. On the other hand, with APCERs above 40%, both only detect the most obvious attack presentations. Except for BSIF (APCERs above 25%), the other algorithms detect much more attack presentations yielding APCERs between 10% and 14%. However, the BPCER of MC hist is nearly at 10% as well, thus unsuited for convenient requirements. Hence, only two algorithms, LBP and PLBP achieve usable performance with identical APCERs, whereas LBP has the lowest BPCER in the first scenario (1.90%) but is then outperformed by PLBP in the second scenario (0.68%).

Finally, Figure 3.27 shows examples of undetected PAI species. Similar to a bona fide capture, the vein patterns are clearly visible



(a) dragonskin (b) ecoflex (c) glue

Figure 3.27: Example vein images of undetected PAI species.

| PAD algorithm | APCER (BPCER = 0.1%) | APCER (BPCER = 0.2%) | Fusion weight |
|---|---|---|---|
| SWIR spectral signature | 15.43% | 12.00% | 0.40 |
| LSCI fusion | 21.00% | 21.00% | 0.25 |
| vein Luminosity | 84.29% | 84.00% | 0.05 |
| vein MC mean | 58.86% | 54.14% | 0.10 |
| vein MC hist | 58.43% | 58.43% | 0.20 |
| **Fusion I** | **9.71%** | **6.57%** | |

Table 3.6: PAD results on the handcrafted partition for fixed BPCERs.

through the thin and transparent fingerprint overlays, showing the limitations of fingerprint PAD based on finger vein images. However, a wide range of different PAI species can be successfully detected.

#### 3.1.9.2 *Handcrafted Partition*

Fingerprint PAD algorithms for all data types are evaluated on the handcrafted partition, namely: SWIR spectral signature, a LSCI fusion of HIST, HOG, and LBP SVMs with peak pre-selection, vein luminosity, vein MC mean, and vein MC hist. Due to the differences of the single PAD algorithms, the results in Table 3.6 are given for fixed operation points of BPCER = 0.1% and BPCER = 0.2% in order to ease a direct comparison of the PAD performance. The vein-based algorithms yield APCERs above 50% up to 84%, thus achieving much worse performance than the SWIR (APCER between 12% and 16%) and LSCI (APCER = 21%) algorithms. The fusion weights were optimised on the validation set, thus not fully aligning with the individual PAD performance. However, the detection rates are significantly improved by 38% (BPCER = 0.1%) and 45% (BPCER = 0.2%) compared to the best-performing individual algorithm (SWIR spectral signature).

In addition, the Detection Error Trade-off (DET) curve of this fusion is plotted in Figure 3.28 showing a D-EER of 2.57%. Moreover, in a high-security application, the PAD threshold could be adjusted resulting in a BPCER = 14.30% for an APCER = 1%.

#### 3.1.9.3 *Deep Learning Partition*

Despite the name, also handcrafted PAD algorithms are evaluated on this partition in order to allow a complementary fusion since deep learning algorithms are only applied on SWIR data. In particular, this includes the full LSCI benchmark with separate combinations of the four feature extractors HIST, LBP, BSIF, and HOG with the nine classifiers SVM, KNN, DT, RF, ADA, GNB, SGD, LDA, and QDA. Additionally, a fusion of the best-performing combinations is presented as well as the results for both SWIR CNNs ResNet and VGG19.

Since the handcrafted classifiers are trained using cross-validation, only the best results from multiple parameter ranges are taken for each classifier. As Table 3.7 shows, no classifier achieves the best performance for all four features. Instead, the results vary for different combinations. In general, nearly all algorithms, except for some outliers, report a low BPCER < 1%, thus supporting convenient applications. The lowest APCERs (around 16%) for the HIST features are achieved with RF and ADA classifiers, while all others are above 20% or even 40% APCER. On the other hand, for BSIF features all APCERs are below 19% with the best classifiers SVM and ADA (both around 12%). Similar to HIST, again RF and ADA present the lowest APCERs (around 15%) for LBP features. However, this time GNB follows closely with 17% APCER. The highest error rates (up to 73%) are obtained with HOG features but the lowest APCER of nearly 15.77% (SGD) is close to the other performances. Although e. g. ADA achieves always among the lowest APCERs, its average BPCER of 1.07% is higher than others. Hence, it is not part of the fusion, where the BPCER was minimised. Also HOG features are completely excluded from the fusion as they increased the error rates in all settings. Instead, HIST RF, BSIF

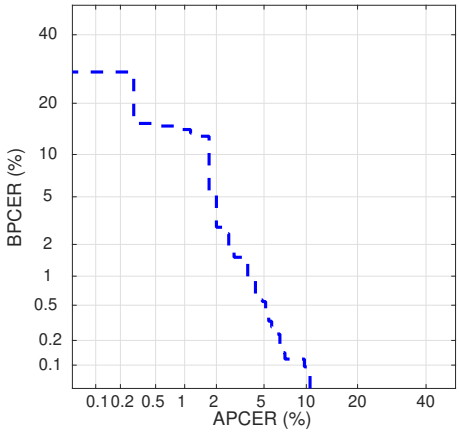



Figure 3.28: DET curve of Fusion I on the handcrafted partition.

| | HIST | | BSIF | | LBP | | HOG | |
|---|---|---|---|---|---|---|---|---|
| | APCER | BPCER | APCER | BPCER | APCER | BPCER | APCER | BPCER |
| SVM | 40.99 | 0.05 | **12.61** | **0.12** | 18.92 | 0.17 | 18.02 | 0.25 |
| KNN | 23.42 | 0.17 | 18.92 | 0.07 | 24.77 | 0.05 | 63.51 | 0.07 |
| DT | 26.58 | 1.42 | 18.92 | 0.79 | 30.18 | 3.54 | 43.69 | 0.96 |
| RF | **15.32** | **0.71** | 14.86 | 0.49 | **15.32** | **0.27** | 19.37 | 0.20 |
| ADA | 16.67 | 0.98 | 12.16 | 1.50 | 15.77 | 0.88 | 18.02 | 0.93 |
| GNB | 41.44 | 0.59 | 15.77 | 0.12 | 17.12 | 0.07 | 19.37 | 0.10 |
| SGD | 37.84 | 5.48 | 18.02 | 0.02 | 22.52 | 0.00 | **15.77** | **1.28** |
| LDA | 43.24 | 8.25 | 17.12 | 0.00 | 28.38 | 2.90 | 50.45 | 7.76 |
| QDA | 47.75 | 0.59 | 13.51 | 1.82 | 27.48 | 47.41 | 72.97 | 25.10 |
| **Fusion: APCER = 9.01, BPCER = 0.05** | | | | | | | | |

Table 3.7: PAD results in percentage on the deep learning partition for all LSCI combinations. The fusion is a majority vote of HIST RF, BSIF SVM, and LBP GNB.

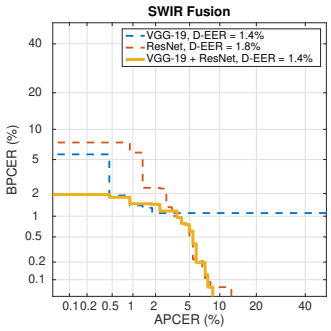
| type | error rate | correct |
|---|---|---|
| bona fide | 0.05 | 4069/4071 |
| overlay dragonskin | 58.33 | 5/12 |
| overlay ecoflex | 4.76 | 40/42 |
| overlay glue | 100.00 | 0/2 |
| overlay latex | 50.00 | 1/2 |
| overlay printout | 26.32 | 14/19 |
| overlay silicone | 4.00 | 72/75 |

Table 3.8: Classification errors of the LSCI fusion.

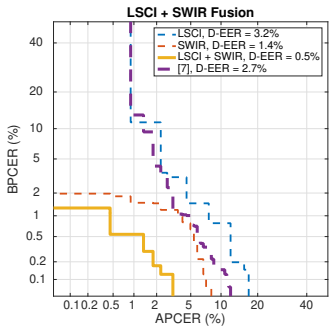
SVM, and LBP GNB are fused, resulting in an improved APCER = 9.01% for a BPCER = 0.05%.

The classification errors of this LSCI fusion are mostly on account of thin and transparent fingerprint overlay PAIs. Due to the nature of the laser illumination, both PAI and skin are penetrated by the light, such that blood movement is still detectable. On the contrary, all attack presentations by means of full fake fingers are correctly classified as visible from Table 3.8. It should be noted that previous publications [111, 112] use only the LSCI SVMs from [167] within the specified fusions I and II. However, this benchmark proves that other classifiers are better suited in case of three out of four extracted features.

Based on the same deep learning partition, also both SWIR CNNs ResNet and VGG19 are evaluated. The corresponding DET curves including their fusion are plotted in Figure 3.29a. VGG19 (dashed




(a) SWIR CNN fusion




(b) Fusion II including its two parts and in comparison to Fusion I (here: [7])

Figure 3.29: PAD results on the deep learning partition as from [112].

blue) achieves a D-EER = 1.4% but constantly misclassifies some bona fide presentations, which results in a permanent BPCER > 1%. On the other hand, ResNet (dashed red) has a slightly higher D-EER = 1.8% but is able to reduce the BPCER < 1% for a trade-off in increasing APCERs > 4%. However, their fusion (solid yellow) benefits from both models when it comes to low BPCERs or low APCERs, but stays at a D-EER = 1.4%.

Figure 3.29b includes the DETs for fusion II as well as its LSCI and SWIR parts. Additionally, the results from fusion I (denoted as [7]) are plotted for direct comparison. Although the SWIR curve (dashed red, D-EER = 1.4%) is continuously below the LSCI curve (dashed blue, D-EER = 3.2%), both algorithms have only five identical misclassifications from all 222 attack presentations in the test set. Hence, fusion II (solid yellow) profits from these complementary information and decreases the D-EER to 0.5%. For a BPCER = 0.12%, fusion II achieves an APCER = 3.15%, which relates to seven undetected attack presentations. These all belong to the overlay group and are made from dragonskin, ecoflex, and printout. For high-security applications, the PAD threshold can be further adjusted, resulting in an APCER = 0% while maintaining a BPCER < 2.5%, thus detecting all attack presentations. Furthermore, fusion II clearly outperforms fusion I (dashed purple, D-EER = 2.7%).

### 3.1.10 *Summary*

The results have shown that it is in fact helpful for fingerprint PAD to utilise multiple sensing techniques and capture complementary data. The wide variety of different PAI species that are available, and those which are still unknown, pose a severe threat to biometric recognition systems. In this context, a compound dataset with 442 attack presentation samples from 35 different PAI species and 4,290 bona fide presentation samples was collected with a newly designed

multi-sensing capture device with the goal to develop fingerprint PAD methods that are robust to various PAI species.

The evaluation of vein-based PAD algorithms revealed that those are strong against the fake finger class but vulnerable to thin and transparent overlays as the bona fide veins are still visible. Furthermore, there are significant differences within the bona fide samples due to a fixed capture process that does not differentiate between e. g. thumbs and little fingers. As a consequence, the finger veins are varyingly strong recognisable, which leads to higher BPCERs. In addition, the LSCI technology also analyses the inside of the finger by observing blood movement within the tissues. Hence, these PAD algorithms come with the same weaknesses as the vein PAD methods, namely thin and transparent overlays. On the other hand, nearly no errors occur for bona fide presentations since those tissues are directly beneath the skin, thus providing better generalisability than vein-based PAD. On the contrary, the surface of the finger is assessed in the SWIR domain. While the spectral signature PAD method is very sensitive regarding illumination changes as it computes the distances between the distinct frames, the CNNs are a more performant alternative for fingerprint PAD. Finally, the fusion of LSCI and SWIR algorithms combines the strengths from both techniques to significantly improve the overall PAD performance.

Summarising the findings, due to the many different ways and materials that can be used to fabricate fingerprint PAIs, PAD concepts can benefit from complementary sensing technologies. However, for finger vein and LSCI advantages and disadvantages overlap in large parts as both focus on the inside of the finger. As a consequence of the less superior results, the finger vein sensor could be omitted in order to reduce the hardware costs.

## 3.2 FINGERPRINT PAD USING SELECTED SENSING TECHNIQUES

Based on the insights from Section 3.1 the fingerprint PAD strategy was adjusted. This includes hardware changes in the capture device, and PAD algorithm development now focuses on data from the SWIR domain, which includes the laser wavelength. Thus vein data are excluded since the results showed similar weaknesses of vein and LSCI PAD, with LSCI being more robust. Following the structure of the previous Section, the next set of own contributions for fingerprint PAD algorithms is presented.

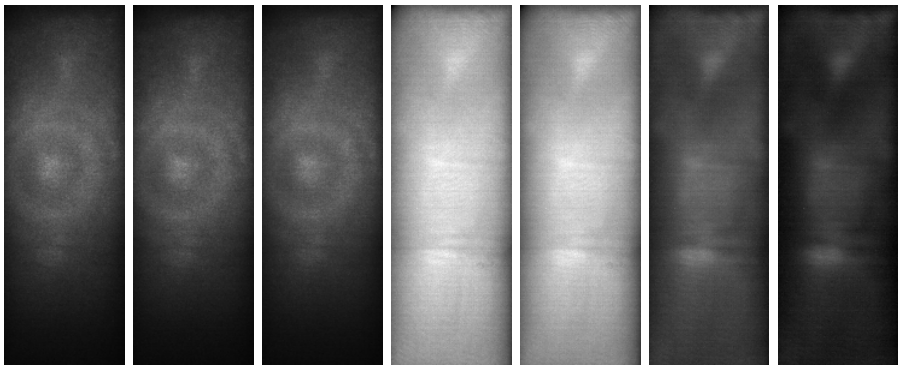
### 3.2.1 *Capture Device and Data*

The general design of the camera-based capture device (Figure 3.1) including multiple sensing techniques was approved by the previous results[14]. In fact, only one camera was replaced to capture images with a higher resolution[15]. The camera to acquire samples in the visible and NIR domain is the same as before. Thus, the legacy compatibility in terms of extracting the fingerprint did not change. However, for the SWIR domain the Hamamatsu was replaced by a 100 fps Xenics Bobcat 320, increasing the resolution from 64×64 pixels to 320×256 pixels. In addition, one lens is used to capture both laser and SWIR data and because of the higher resolution, the fast steering mirror could also be removed. On the other hand, illuminations did not change, within the SWIR domain still a 1310 nm fiber laser and 1200 nm, 1300 nm, 1450 nm, and 1550 nm LEDs are utilised.

Due to the camera change, the captured laser sequence is reduced to 100 frames (before 1025 fps), but still observes the movements within one second, and the new RoI of the finger slot comprises 100×300 pixels. Moreover, the total capture time could be reduced to two to four seconds per finger in contrast to 20 to 30 seconds of the previous prototype. The execution times of the lens flipper and fast steering mirror are saved and additionally distinct wavelengths (i. e., visible and SWIR) are now simultaneously captured. Example frames of a bona fide presentation acquired with the new camera are depicted in Figure 3.30. The first three frames are taken from the laser sequence (1310 nm) and the last four are taken with the specified SWIR wavelengths (1200, 1300, 1450, 1550 nm). Since the subtle temporal changes between the laser frames are not visible, only the middle one is displayed for the samples of attack presentations in Figure 3.31. In general, a circle is recognisable where the laser hits the finger, whereas the SWIR LEDs provide a more consistent illumination. Depending on the type and material of the PAI, the acquired images appear more or

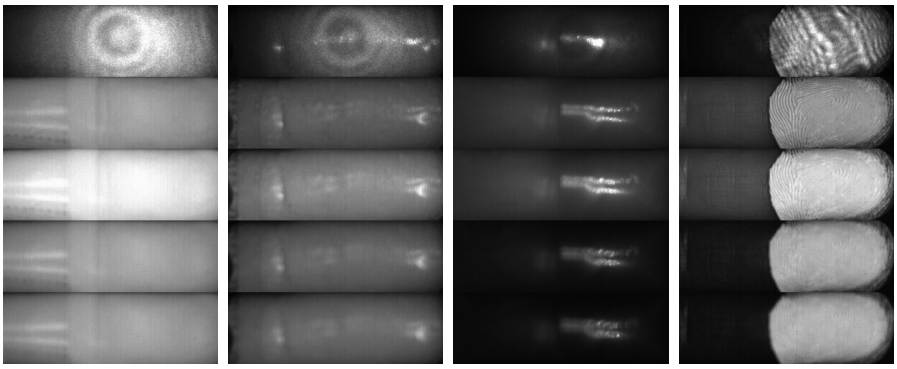
14 Parts of this Section are derived from our publications [176, 178].

15 An in-depth description of the revised capture device is given in [267, 268] as well as in the appendix (Figure a.4).



(a) 1310 nm (first) (b) 1310 nm (middle) (c) 1310 nm (last) (d) 1200 nm (e) 1300 nm (f) 1450 nm (g) 1550 nm

Figure 3.30: Bona fide samples acquired at five different wavelengths.



(a) ecoflex (b) gelatin (c) glue (d) silicone

Figure 3.31: Samples of four different PAIs accross all five wavelengths. Top to bottom: 1310 nm laser, 1200 nm, 1300 nm, 1450 nm, and 1550 nm SWIR.

less similar to a bona fide presentation. This is especially visible for the silicone sample, where the reflections within the same presentations differ very much between the overlay and the skin parts.

### 3.2.2 *Underlying Concepts for Fingerprint PAD*

The novel fingerprint PAD algorithms for the new laser and SWIR data are partly based on different concepts[16], which are introduced here according to research question RQ2. Given that the new camera and lens combination captures shorter laser sequences with a higher resolution, the initial procedure to compute the contrast from the laser speckle sequence returned mostly black images. Hence, additional neighbourhood size were tested to compute the temporal contrast, but for all these settings the major share of computed LSCI images was just black. Furthermore, multiple parameter settings to compute the spatial

16 The concepts in this Section are derived from our publications [117, 176, 178, 179].

contrast failed as well. Since neither temporal nor spatial contrast images are computed, the term Laser Speckle Contrast Imaging (LSCI) is no longer used and instead the data is simply referred to as laser sequences or laser frames. Finally, due to the superior performance of the CNNs in the previous evaluations, the new fingerprint PAD methods focus on deep learning concepts.

#### 3.2.2.1 *Convolutional Neural Networks*

Based on the CNNs introduced in Section 3.1.2.4 (pre-trained VGG19 and ResNet trained from scratch), additional pre-trained CNNs from the literature are fine-tuned for this fingerprint PAD task. These CNNs are commonly trained on the *ImageNet* challenge [72, 250] consisting of millions of images from 1,000 categories. Hence, the last fully connected layers are replaced to support binary predictions between bona fide presentations and attack presentations. Furthermore, only the weights from the last CNN block(s) are retrained on the captured fingerprint data. This should prevent over-fitting as the available training data for fingerprint PAD is not comparable to the millions of images from *ImageNet*, for which the CNNs were designed for. The utilised networks[17] are introduced in the following:

**InceptionV3** [274]. The InceptionV3 was designed to reduced computational costs compared to other very deep architectures while maintaining the classification performance. Additionally, the authors suggest to avoid extreme compressions but gently decrease the input. Furthermore, the network reduces the dimension before applying spatial aggregations, thus saving parameters and execution time. Subsequent to some convolutional layers, three different inception modules are repeatedly deployed, whereby the last two blocks are retrained.

**MobileNet** [135]. The main characteristic of MobileNet is that it uses depthwise convolutional layers, which treat each channel of the input image separately. Subsequently, the information is combined in a pointwise convolutional layer ($1 \times 1$ Conv). Through separating the spatial analysis of the image channels, less parameters are required which allows a faster processing. Due to the large size of 13 blocks, the average pooling and fully connected layers for fingerprint PAD classification are already placed after the $8^{\text{th}}$ block.

**MobileNetV2** [254]. In addition to the depthwise convolutional layers, MobileNetV2 further applies residual connections and inverted bottlenecks. The bottleneck block consists of a convolutional layer, a depthwise convolution, and a final striding convolutional layer as shown in Figure 3.32. The residual connections combine input and output of those bottlenecks for blocks with identical parameters $t, c, s$ as the previous one. Similar to MobileNet, only twelve out of 16 blocks are used and the last two are retrained.

17 The descriptions are partly derived from our publication [117] as additional networks are included in the Thesis.

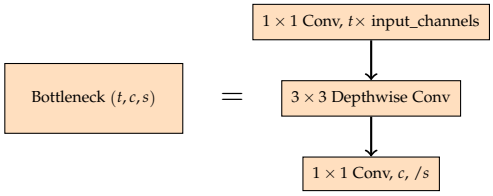



Figure 3.32: The MobileNetV2 bottleneck block comprises three layers, with expansion factor $t$, number of filters $c$, and stride $s$.

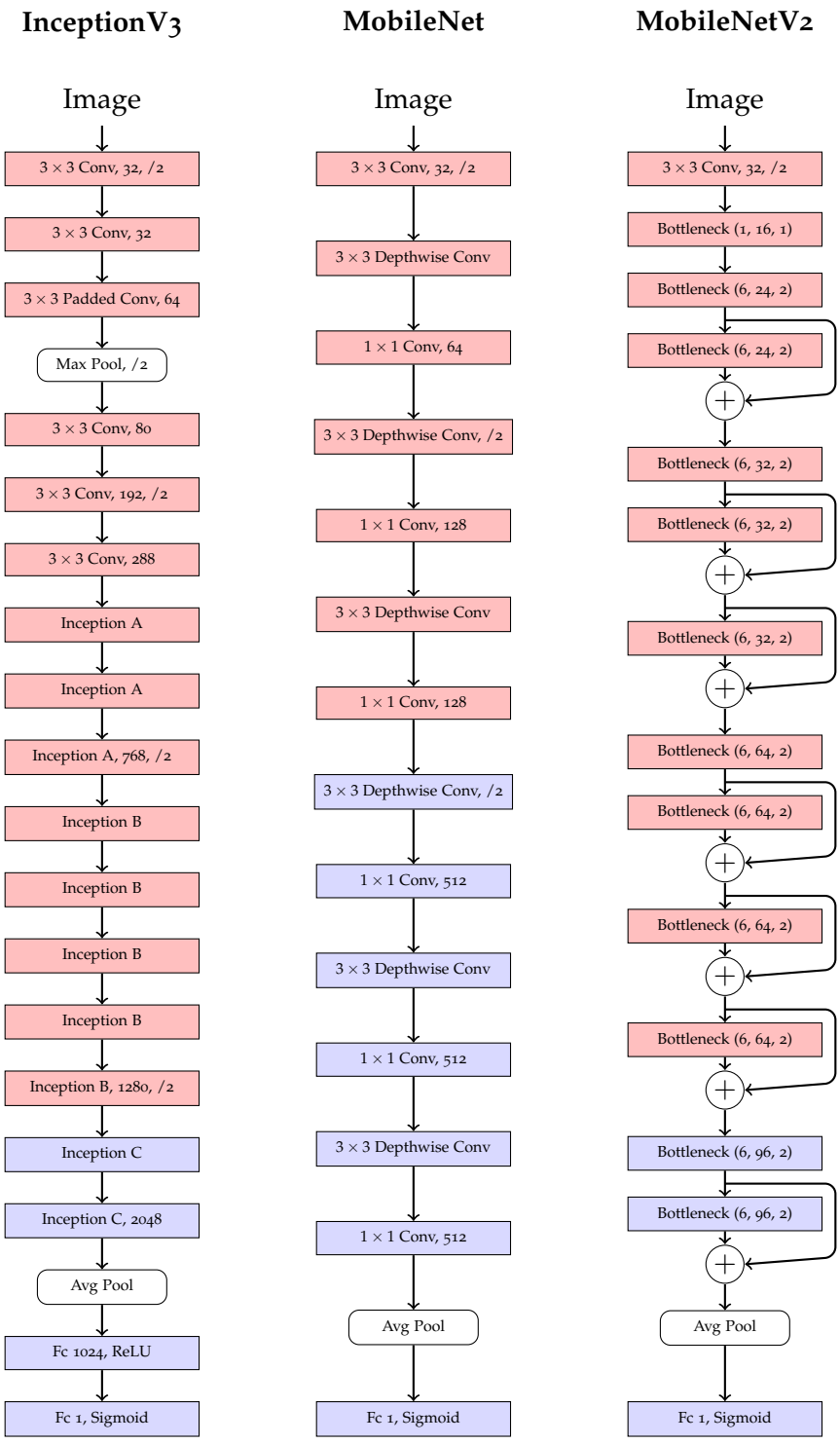
**ResNet**. The self developed ResNet is a small residual CNN that is trained from scratch on the fingerprint PAD data. The network consists of five convolutional layers and two residual connections, thus adding a previous representations into the downstream flow. This design intends to mitigate possible information loss during the process. Furthermore, residual CNNs have a reduced training time, which allows the usage of deeper architectures [128].

**VGG16 and VGG19** [260]. Using 16 and 19 convolutional layers only, the VGG architecture is much simpler than the other pre-trained networks. Although they are older as well, their performance is still among the best in several competitions. Both networks consist of five blocks with two to four convolutional layers. The blocks are separated by max pooling operations that additionally reduce the dimensionality by a stride of two. In both cases, the last three layers are fine-tuned on the fingerprint PAD data.

**VGGFace** [225]. Based on the previously described VGG16, VGG-Face was completely re-trained on a large-scale face dataset with 2.6 million images [225] to perform face recognition. Hence, the weights are optimised by evaluating facial images that show a large proportion of skin, which could be beneficial for the fingerprint PAD task at hand. As for all other approaches, the last fully connected layers are exchanged to support binary classifications.

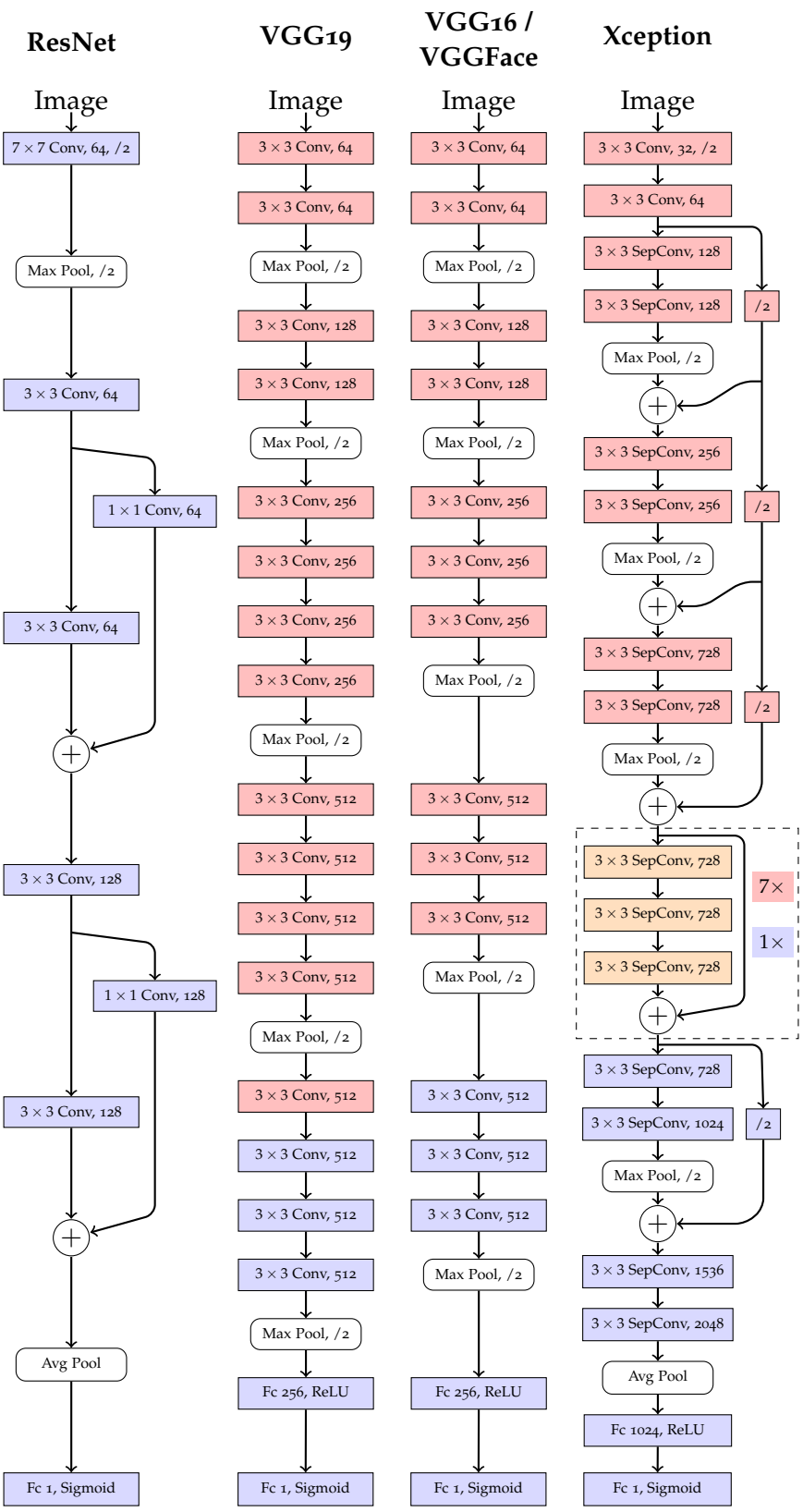
**Xception** [53]. The Xception CNN replaces the inception modules from InceptionV3 with depthwise separable convolutions (SepConv), which are proposed to use model parameters more efficiently. Additionally, residual connections are added to combine input and output of the blocks. In general, the linear stacking of depthwise separable convolutions is much easier to implement than the defined inception modules of InceptionV3.

Figure 3.33 provides an overview of the different CNN architectures. However, this comprehensive overview does not include intermediate *batch normalisations* or *activations* and the *Inception* modules can be found in the original description [274]. The orange Xception block is repeated in total eight times, whereas only the last iteration is retrained. For all CNNs holds, that frozen layers are coloured in red and trainable ones are marked in blue. Pooling operations do not include trainable parameters.




(a) Inception V3, MobileNet, and MobileNetV2.

Figure 3.33: CNN architectures (1/2).




(b) ResNet, VGG16, VGG19, VGGFace, and Xception.

Figure 3.33: CNN architectures (2/2).

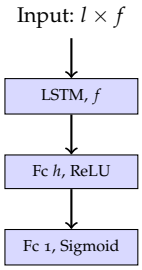



Figure 3.34: LSTM architecture based on sequence length $l$ and extracted feature vector size $f = 4 \times h$. The LSTM layer processes the sequence but preserve the feature length, the first fully connected layer reduces the vector by a factor of four, and the final fully connected layer outputs a PAD score between 0 and 1.

#### 3.2.2.2 *Long Short-Term Memory Networks*

The main property of feed-forward NNs (e. g. CNNs) is their independent processing of given inputs[18]. However, in order to extract information from image sequences, a memory state is required to connect subsequent frames of the sequence. The concept of Recurrent NNs (RNNs) was developed to process information from time sequences in particular and the Long Short-term Memory (LSTM) architecture is optimised to learn long-term dependencies. In this context, Jozefowicz et al. [159] conducted a thorough architecture study to search for better architectures than the LSTM. After an extensive evaluation of more than 10,000 different structures, only one architecture was found to outperform the LSTM on some but not all tested problems. Hence, the LSTM approach is evaluated for fingerprint PAD on the laser sequence data.

However, LSTMs work on 1-dimensional temporal data and are not able to extract this information from images. Therefore, CNN models are used as feature extractors by removing the prediction block at the end. Instead of a PAD score, the CNNs now output a 1-dimensional feature vector, which fulfils the input requirements of the LSTM. These vectors derived from the full image sequence can then be used to train the LSTM, which is depicted in Figure 3.34. The particular input shapes and output sizes depend on the length of the extracted feature vector of the corresponding CNN. In contrast to the complex CNN architectures, the LSTM is rather simple as it does not need to process multi-dimensional data.

The weakness of this approach is that the CNNs are trained to classify the input based on steady features within one frame. Hence, the extracted feature vector might not necessary contain relevant information for the LSTM to connect the temporal relations. As a countermeasure, Donahue et al. [75] propose a Long-term Recurrent Convolutional Network (LRCN) architecture, which includes CNN

18 The descriptions are derived from our publication [179].

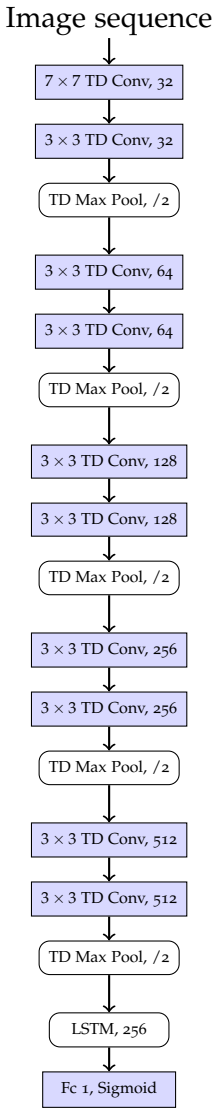



Figure 3.35: LRCN architecture. All layers, except LSTM and fully connected, are time distributed (TD) and all blocks are trained from scratch.

and LSTM layers. Through specifying the CNN layers as *time distributed*, the convolutional operations are applied to all frames of the sequence. The resulting LRCN thus learns convolutional perceptual representations together with temporal dynamics of the sequence. The utilised LRCN implementation [264] was influenced by VGG16 [260], but is slimmed down to two convolutional layers per block to avoid over-fitting on smaller datasets. The block-wise architecture is depicted in Figure 3.35. Each convolutional layer is followed by a *Batch Normalization* and a *ReLU activation*. Since all blocks are jointly trained, the particular weights are adjusted based on the spatial and temporal information gain.

#### 3.2.2.3 *Convolutional Autoencoders*

In contrast to the previous PAD concepts, one-class (OC) classifiers[19] can be trained on bona fide presentations only. Their strength is to focus on the structure of a single class during the training and try to map unseen data into this structure. If the differences are too big, the testing sample is classified as an attack presentation. Hence, finding an optimal decision threshold, that includes varying bona fide presentations and detects sophisticated attack presentations, remains the main challenge. As for all biometric recognition systems, intra-class deviations occur for bona fide presentations due to environmental factors and the interaction with individual data subjects.

As a deep learning based OC classifier, the convolutional Autoencoder (AE) implements multiple CNN layers for its encoding $(h = f(x))$ and decoding $(x' = g(h))$ functions. The encoder converts the input $x$ to a lower dimensional latent representation $h$, while the decoder tries to retrieve the original input from this representation. In order for the AE to focus on the relevant parameters, the training process learns to minimise a loss function:

$$L(x, g(f(x))) \tag{3.8}$$

The AE gets penalised when $x' \neq x$ and can adjust the corresponding weights to reduce the dissimilarities. As a consequence, the performance of the AE highly depends on the selected loss function. In order to reduce the training time, the loss values are computed for a subset, *batch*, which comprises randomly assigned samples and the final loss value is obtained based on all *batches*. Nevertheless, a fundamental design decision is to build an *undercomplete* AE architecture with $h$ having a lesser dimension than the input $x$. This is required for the AE to focus on the most relevant information from the input image and particularly prohibits internalising of the identity function $id(x) = x$ [263].

In the case of fingerprint PAD, the trained model encodes the input $x$ and reconstructs the output $x'$. Subsequently, the Reconstruction Error (RE) between $x$ and $x'$ is computed, which generally is lower for inputs that are similar to the bona fide training data. On the other hand, the reconstruction fails for inputs that differ from the training samples, thus indicating an attack presentation. The resulting RE can be used as a PAD score and classification is achieved through a PAD threshold. Due to their high sensitivity towards unknown inputs, AEs are a popular tool for anomaly detection [143, 216].

In the context of this Thesis, three AE architectures are evaluated that are depicted in Figure 3.36: Conv-AE, Pooling-AE, and Dense-AE. The naming reflects the additional layers in each architecture. Starting with one convolutional layer, that includes a dimension reduction by

19 The descriptions are derived from our publication [178].

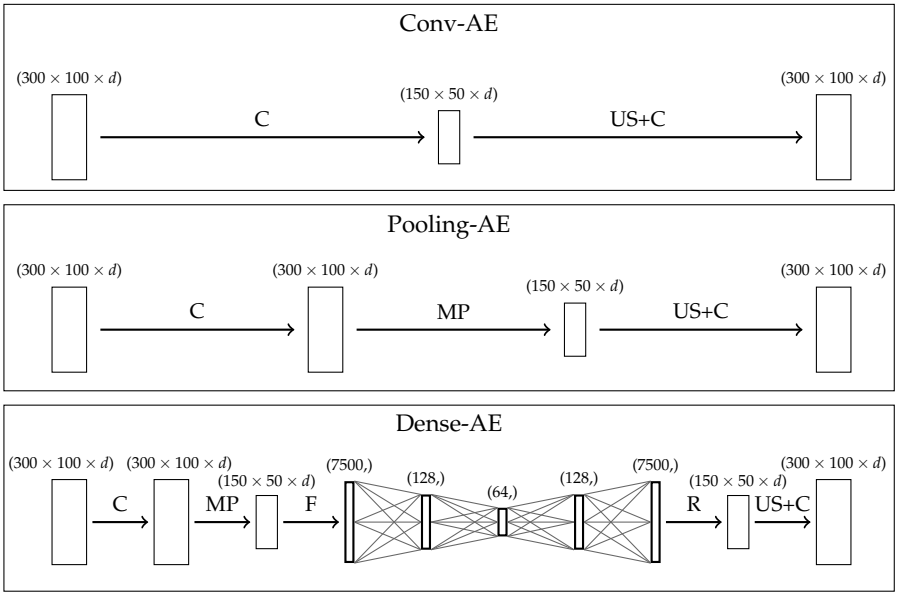



Figure 3.36: Autoencoder architectures for variable image dimensions $d$. The following operations are used: C = convolution, US = up-sampling, MP = max pooling, F = flatten, R = reshape.

a stride of two, the Conv-AE has the most simple structure. For the next AE (Pooling-AE), the stride is moved to an additional pooling layer. On the one hand, Springenberg et al. [269] found that replacing max pooling layers with convolutional layers that include a striding operation leads to no significant performance decrease. On the other hand, Goodfellow et al. [122] report that max pooling layers contribute to achieve translation invariance in small image regions. Hence, both approaches are directly benchmarked against each other. Finally, the Dense-AE adds a fully connected NN to the Pooling-AE, which was suggested by Ke et al. [166]. This additional network reduces the latent representation $h$ to a 64-dimensional vector. All convolutional layers apply twelve filters and *ReLU* activations are used, except for the last decoding layer, where the *Sigmoid* function is applied.

#### 3.2.2.4 *Reconstruction Error*

As mentioned before, multiple functions can be utilised to compute the loss[20] and most commonly the mean squared error (MSE) [18] is used:

$$\begin{aligned} L(x, x') &= \frac{1}{B}\sum_{j=1}^{B}\frac{1}{WHI}\sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}(x^{j}_{whi} - x'^{j}_{whi})^2 \\ &= \frac{1}{B}\sum_{j=1}^{B}\frac{1}{WHI}\sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}e^{j}_{whi}(x, x') \end{aligned} \tag{3.9}$$

20 The descriptions are derived from our publication [178].

with $B$ as number of data samples within one batch iteration, $W$ image width, $H$ image height, and $I$ number of input channels (i. e., three for RGB images). The popularity of MSE arises from its understandable simplicity and that it is often pre-implemented. Nevertheless, MSE is known to be vulnerable to random noise since the squared difference is computed pixel-wise. Hence, the RE is directly effected by outliers, which impacts the PAD performance. This problem of *robust estimation* towards outliers is widespread across all areas of deep learning [248]. As a countermeasure for AEs, Ishii and Takanashi [143] propose a weighted MSE (wMSE) for anomaly detection, which is defined as follows:

$$L_{Ishii}(x, x') = \frac{1}{B} \sum_{j=1}^{B} w^j \cdot mse^j(x, x') \tag{3.10}$$

where

$$mse^j(x, x') = \frac{1}{WHI} \cdot \sum_{w=1}^{W} \sum_{h=1}^{H} \sum_{i=1}^{I} e^j_{whi}(x, x') \tag{3.11}$$

and $w^j$ is defined as

$$w^j = \begin{cases} 1, & mse^j(x, x') \leq C \\ 0, & mse^j(x, x') > C \end{cases} \tag{3.12}$$

$C$ is the $\alpha$-th quantile of $mse = [mse^1, \ldots, mse^B]$, defining the threshold to exclude training samples with exceeding MSEs. As a consequence, outliers are ignored during training and cannot distort the model weights. In general, the phenomenon of unknown outliers is likely for unlabelled training data. However, in this particular fingerprint PAD case, the dataset was manually checked and the training data includes only bona fide presentations. Hence, ignoring full samples means a loss of generalisability since similar bona fide presentations during testing would most likely be misclassified.

Therefore, the loss function of Ishii and Takanashi is further modified to operate pixel-wise such that only outlying pixels are excluded instead of ignoring the complete sample. This leads to an optimised AE that is trained to successfully reconstruct the most significant parts of the image while ignoring occurring noise. For this, the definition of the wMSE changes as follow:

$$L_{Prop}(x, x') = \frac{1}{B} \cdot \sum_{j=1}^{B} \frac{1}{WHI} \sum_{w=1}^{W} \sum_{h=1}^{H} \sum_{i=1}^{I} w^j_{whi} e^j_{whi}(x, x') \tag{3.13}$$

with

$$w^j_{whi} = \begin{cases} 1, & e^j_{whi}(x, x') \leq mse^j(x, x') + C \cdot std^j \\ 0, & e^j_{whi}(x, x') > mse^j(x, x') + C \cdot std^j \end{cases} \tag{3.14}$$

and

$$std^j = \sqrt{\frac{1}{WHI} \cdot \sum_{w=1}^{W} \sum_{h=1}^{H} \sum_{i=1}^{I} \left( e^j_{whi}(x, x') - mse^j(x, x') \right)^2} \tag{3.15}$$

In addition to the mean, also the standard deviation of the squared error is included in the threshold calculation. Finally, this threshold defines which pixels are excluded based on the pixel-wise RE with the assumption that the overall RE is reduced for bona fide presentations. The new challenge lies in estimating an optimal value for the constant $C$. Too high values pose a risk that noisy areas are not excluded from the processing, resulting in a less robust model causing higher error rates. On the other hand, the border between bona fide presentations and attack presentations diminishes for too low $C$ values as relevant information for classification are excluded (i. e., over-generalisation). As a result, the $C$ parameter is not generally fixed but optimised for each use case (i. e., fingerprint PAD).

### 3.2.3 *SWIR Fingerprint PAD Algorithms*

In contrast to other wavelengths, where different skin types vary a lot, all skin types reflect in a very similar way within the SWIR spectrum. This property is exploited[21] to distinguish bona fide presentations from several different attack presentations by utilising the power of CNNs. In addition to the CNNs introduced in Figure 3.33, another small network is trained from scratch. In fact this is the CNN base of the LRCN (Figure 3.35) without the time distributed specifications. Since it is a trimmed-down version of VGG16, it is simply denoted as VGG10. All in all, two CNNs are trained from scratch and additional seven CNNs are fine-tuned on the fingerprint data. Furthermore, given the four captured SWIR wavelengths ($\lambda_1 = 1200$nm, $\lambda_2 = 1300$nm, $\lambda_3 = 1450$nm, and $\lambda_4 = 1550$nm), this benchmark includes two input formats: *i)* the manual RGB conversion from Eq. (3.7), and *ii)* a convolutional input processing block that is jointly trained with the CNNs in order to reduce the 4-dimensional SWIR input to a 3-dimensional image that fulfils the CNN input requirements. A generic overview of the SWIR fingerprint PAD setup is given in Figure 3.37. The proposed pre-processing block receives a 4-dimensional input image and applies three convolutional filters of size $P \times P$. Subsequent batch normalisation and ReLU activation facilitate convergence to the following CNN blocks. Finally, multiple filter sizes $P \in \{3, 5, 7, 9, 10, 11, 13, 20, 30, 40, 50\}$ are tested for each CNN to find the optimal combination.

Whereas the handcrafted RGB conversion is fixed, each CNN can find an own 3-dimensional combination during the training stage. Besides, the RGB conversion affects the whole image area homogeneous,

21 This Section is based on our publication [117] but additional settings are tested.

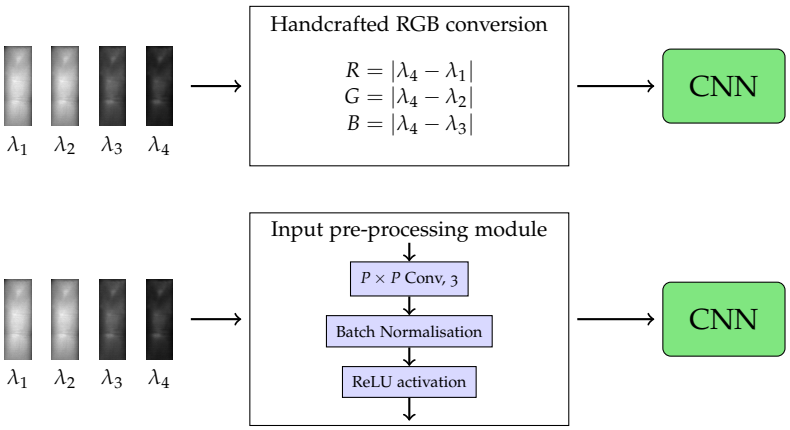



Figure 3.37: General overview of the SWIR PAD methods. The input pre-processing module is jointly trained with the following CNN.

whereas the new input processing module allows the CNNs to execute different linear and non-linear convolutions at different areas of the image. Thus, the most suitable operations are applied due to the joint training with the other blocks. As a result, each CNN might create a unique combination of the four SWIR wavelengths.

### 3.2.4 *Laser Fingerprint PAD Algorithms*

Given the new higher resolution camera in contrast to the old capture device, these fingerprint PAD algorithms[22] can investigate much subtler differences between bona fide presentations and sophisticated attack presentations. In this context, spatial features are evaluated by CNNs and the temporal connections from the laser sequence are extracted with the LSTM approach.

The structural overview of the different fingerprint PAD approaches is illustrated in Figure 3.38. The same CNNs from the SWIR PAD compilation are also trained on the laser data. However, as there are no clearly visual differences between the distinct frames of the laser sequence, only the middle frame is used to train the CNNs in order to prevent over-fitting. Furthermore, the same CNNs are later reused to extract the feature vectors from all 100 frames for the LSTM. Hence, connecting e. g. first, middle, and last frame in the RGB image could affect the generalisability. For the pre-trained CNNs, which require 3-dimensional input, a greyscale to RGB conversion is applied. This simply copies the 1-dimensional greyscale image to all three dimensions of the RGB image. Since these duplicated dimension contain no additional data, the 4-dimensional SWIR input processing layer is not used for this data.

While two CNNs are trained from scratch on this data, the other seven pre-trained models are fine-tuned, thus retraining only the

22 This Section is based on our publication [179].

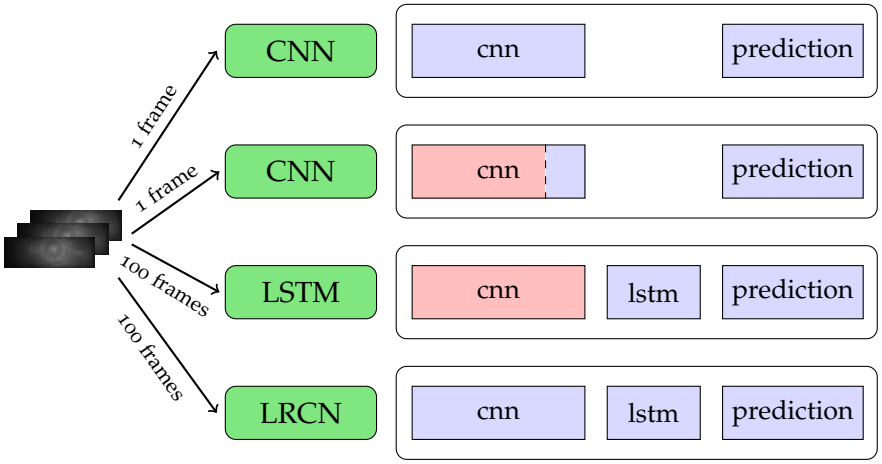



Figure 3.38: Structure of the laser PAD approaches with trainable blocks coloured in blue and fixed ones are marked in red.

last block(s). The LSTMs utilise those trained CNN bases as feature extractors for all 100 frames and feed these to the *lstm* layer that is connected to the prediction block. Hence, all nine CNNs are also used for this setup. In theory it is also possible to extract the feature vector with the original model weights, however those vectors do not contain information that is beneficial to distinguish between bona fide presentations and attack presentations. Finally, the LRCN jointly trains a trimmed CNN with the LSTM on the full image sequence as well. Since this CNN (VGG10) is also used in the LSTM setup, the effect of this joint training can directly be measured.

The contrast of powerful CNNs on the one hand and LSTMs that utilise those CNNs for feature extraction on the other hand, allow a further evaluation whether the temporal information are a key element for fingerprint PAD. Since the LSTM setup requires a trained CNN anyhow, it needs to further improve the PAD performance to be of value. All in all, this benchmark validates nine CNNs trained on the mid-frame next to nine LSTMs and one LRCN that process the full image sequence.

### 3.2.5 *One-Class Fingerprint PAD Algorithms*

In contrast to the previously introduced fingerprint PAD algorithms, which are trained on both classes, one-class (OC) algorithms are only trained on bona fide presentations[23]. Given the diverse possibilities to create new fingerprint PAI species [161, 165], especially unknown attacks [261] threaten established PAD methods. In addition to the effort of collecting extensive datasets comprising new PAI species, two-class classifiers might need to get retrained each time as well. Hence, OC classifiers [278] have the advantage that all attack presentations are

23 This Section is based on our publication [178].

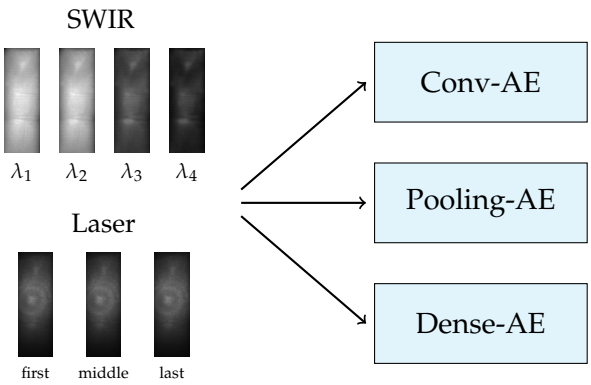



Figure 3.39: Structure of the AE PAD approaches.

treated as unknown attacks and can be detected as anomalies differing from the bona fide presentations seen during training.

In this context, the three introduced AE architectures from Section 3.2.2.3 (Conv-AE, Pooling-AE, Dense-AE) are evaluated on both SWIR and laser data as the structure is flexible towards the input dimension. The four captured SWIR wavelengths are combined to a 4-dimensional image and from the laser sequence a 3-dimensional image is built using the first, middle, and last frame. In contrast to the previously described CNNs, the AE is not used as a feature extractor and thus can combine different frames. Through the concatenation into one single image, the AEs can process all information simultaneously. This fingerprint PAD setup is illustrated in Figure 3.39.

The three AEs are benchmarked against each other using the MSE as a loss function. The impact of the RE is measured in the next step. In particular, only the architecture that shows the best results is further analysed with the modified wMSE for different parameters $C \in \{1.4, 1.6, 1.8, 2.0, 2.2\}$. As the used loss function influences the model's weights in the training process, new instances of the AE need to be trained for each parameter change. In a last step, different weighted fusions are evaluated in order to combine the information from SWIR and laser data.

Finally, the soundness of the approaches needs to be validated by benchmarking the AE to other OC classifiers. For this purpose, OC-SVM [48] and OC-Gaussian Mixture Model (GMM) [245] have been selected as both showed good performance in other PAD tasks [74, 217]. However, both classifiers require 1-dimensional feature vectors and cannot process the constructed input images. Hence, the fine-tuned VGG19 CNN is utilised as a feature extractor and additionally the latent representation of the AE is also used as a feature vector. Therefore, the AE is benchmarked to four combinations of input features and classifier as depicted in Figure 3.40. Additionally, the SWIR and laser results of those OC classifiers are also fused for

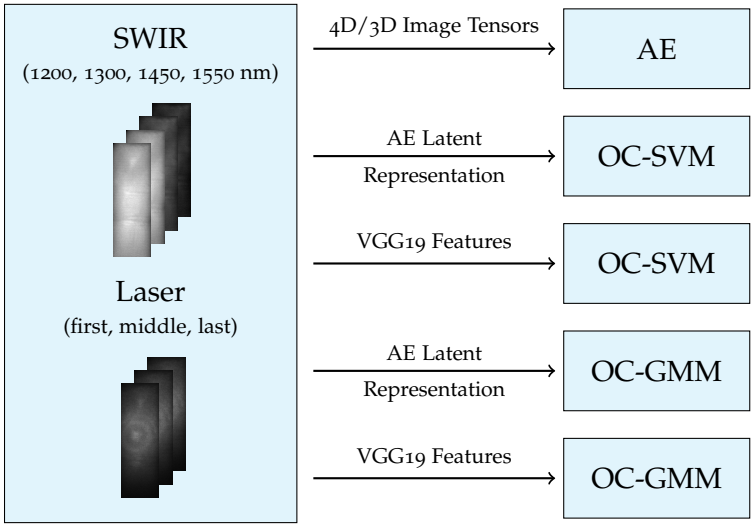



Figure 3.40: Overview of the benchmark between the AE and additional OC classifiers with their corresponding input features.

enhancement of their performance and fairness towards the best AE setup.

### 3.2.6 *Fusion*

The motivation of the fusion is to combine the strengths of both SWIR and laser fingerprint PAD algorithms. However, with all these proposed methods that are evaluated within this Thesis, there are far too many possible combinations for a fusion. Hence, the subset to compute the fusion is reduced to the best performing algorithms in terms of $APCER_{0.2}$. Furthermore, the fusion should contain exactly one SWIR and one laser algorithm. Additionally, to reducing the number of possible combinations for consideration, this limitation is further expected to generalise better as more complex fusion schemes might be over-engineered onto this particular task. On the other hand, the fusion still comprises complementary information from both data types.

### 3.2.7 *Database*

A total of four data acquisitions in two distinct locations contributed to the database[24]. In each capture round, subjects presented six to eight fingers (little fingers were excluded) with different levels of e. g. ink, dirt, and moisture. Besides, subjects were allowed to join the following acquisitions as well, thus contributing multiple times. The full database consists of 24,050 samples apportioned in 17,730 bona fide presentations and additional 4,339 attack presentations stemming

24 This Section is derrived from our publications [176, 178].

| PAI Group | PAI | # variations | # samples |
|---|---|---|---|
| Fakefinger | 3D printed | 2 | 72 |
| | dental material | 1 | 33 |
| | dragonskin | 3 | 477 |
| | ecoflex | 4 | 291 |
| | latex | 2 | 147 |
| | playdoh | 4 | 116 |
| | silly putty | 3 | 55 |
| | wax | 1 | 74 |
| Overlay opaque | bandage plaster | 1 | 14 |
| | dental material | 1 | 51 |
| | dragonskin | 1 | 17 |
| | ecoflex | 2 | 1035 |
| | gelatin | 1 | 194 |
| | printout paper | 1 | 49 |
| | silicone | 5 | 824 |
| Overlay transparent | dragonskin | 1 | 106 |
| | gelatin | 1 | 107 |
| | glue | 2 | 27 |
| | latex | 1 | 34 |
| | printout foil | 1 | 64 |
| | silicone | 1 | 157 |
| | wax | 1 | 18 |
| Overlay semi | dragonskin | 1 | 47 |
| | ecoflex | 1 | 24 |
| | glue | 2 | 146 |
| | silicone | 1 | 160 |

Table 3.9: Summary of PAIs in the database. The total number of samples is given as well as the number of PAI species sharing the same material basis. Modifications include different colours and transparency levels or applied conductive augmentations.

from 45 different PAI species. In addition to the base groups of full fake fingers and fingerprint overlays, the PAI species are further divided regarding their visual properties instead of the main material. Hence, Table 3.9 summarises the number of samples and variations per PAI species in the corresponding groups: *Fakefinger, Overlay opaque, Overlay transparent, Overlay semi*. Printout PAIs were worn as overlays and PAI variations include different colours or conductive augmentations. For example, ecoflex out of the box dries transparent, but it is also possible to add colour during fabrication to create further appearances. In other cases, e. g. electric paint was applied after the casting process. The choice and recipes of PAI species were selected by the project sponsor. On the other hand, the defined PAI groups are mostly relevant for camera-based PAD systems as e. g. capacitive capture devices would more likely focus on the moisture level than colour or transparency. The same holds for defining PAI species, differently coloured playdoh might cause varying reactions for cameras and optical capture devices but not for capacitive ones. It should be noted that the owner of the data knows the importance of reproducibility and benchmarking future results and indicated to make the complete dataset available[25].

### 3.2.8 *Experimental Protocol*

The partitions for all experiments[26] utilise non-overlapping training, validation, and test sets in order to grant fair evaluations. In this regard, all bona fide samples of one subject, also when they participated in multiple capture sessions, are included in the same set. Furthermore, unified partitions are used to test all fingerprint PAD algorithms, thus allowing effective comparisons of the results.

The priority for the baseline partition is to include all PAI species in training, validation, and test sets. Due to the varying numbers of samples per PAI species and to prevent over-fitting towards some of them, training and validation sets got a fixed maximum number of samples per PAI species. Moreover, an identical number of bona fide presentations and attack presentations are used during training and validation to avoid biased classifiers. Thus, some bona fide samples were excluded due to the restrictions of having disjoint sets for samples from the same data subjects. As a result, training and validation sets are rather small but allow unbiased testing on a large set.

The next partition treats all PAI species as unknown, hence training one-class classifiers on bona fide presentations only. As these classifiers are designed to be biased towards the known class, much more samples are utilised for training and validation. In fact, 30% of subjects (with all their samples) are randomly assigned to the training set and

25 https://www.isi.edu/projects/batl/data
26 The partitions are based on our publications [176, 178, 179].

| PAI Group | PAI | # variations | # samples |
|---|---|---|---|
| Mat. group i) | silicone | 7 | 1141 |
| Mat. group ii) | dragonskin | 6 | 647 |
| | ecoflex | 7 | 1350 |
| Mat. group iii) | gelatin | 2 | 301 |
| | glue | 4 | 173 |
| | latex | 3 | 181 |
| | printout | 2 | 113 |
| | wax | 2 | 92 |
| Mat. group iv) | 3D printed | 2 | 72 |
| | dental material | 2 | 84 |
| | playdoh | 4 | 116 |
| | silly putty | 3 | 55 |

Table 3.10: Specifications of PAIs in the evaluated material groups.

additional 20% to the validation set with the remaining bona fide presentations and all attack presentations available in the test set.

Finally, the best-performing fingerprint PAD algorithms are further evaluated towards their generalisation capabilities on unknown attacks using modified LOO protocols. Instead of leaving out a single PAI species at a time, a complete PAI group is left out to produce more relevant results due to the expected similarity within the groups. Hence, one PAI group is completely excluded from training and validation and only available in the test set. The PAD algorithms are thus trained and validated on the remaining PAI samples (85% training and 15% validation). In order to focus on the unknown attacks and the different training partitions, bona fide samples are kept identically across all LOO sets: 50% for training, 15% for validation, and 35% for testing. In addition to the visual PAI groups from Table 3.9, four material-based PAI groups were defined for the LOO experiments. Table 3.10 specifies the material groups that combine similar PAI species given their different numbers of samples as follows: *group i)* silicone; *group ii)* dragonskin and ecoflex; *group iii)* gelatine, glue, latex, printout, and wax; *group iv)* 3D printed, dental material, playdoh, and silly putty. The particular number of samples for each experimental partition is given in Table 3.11.

Based on the specific partition size and the selected PAD algorithm, the system requirements vary since all training and validation samples are loaded into memory for the training process. As a result, the laser CNNs require between 8 and 16 GB RAM, SWIR CNNs around 32 GB, and the laser LSTMs up to 150 GB as they load all 100 frames,

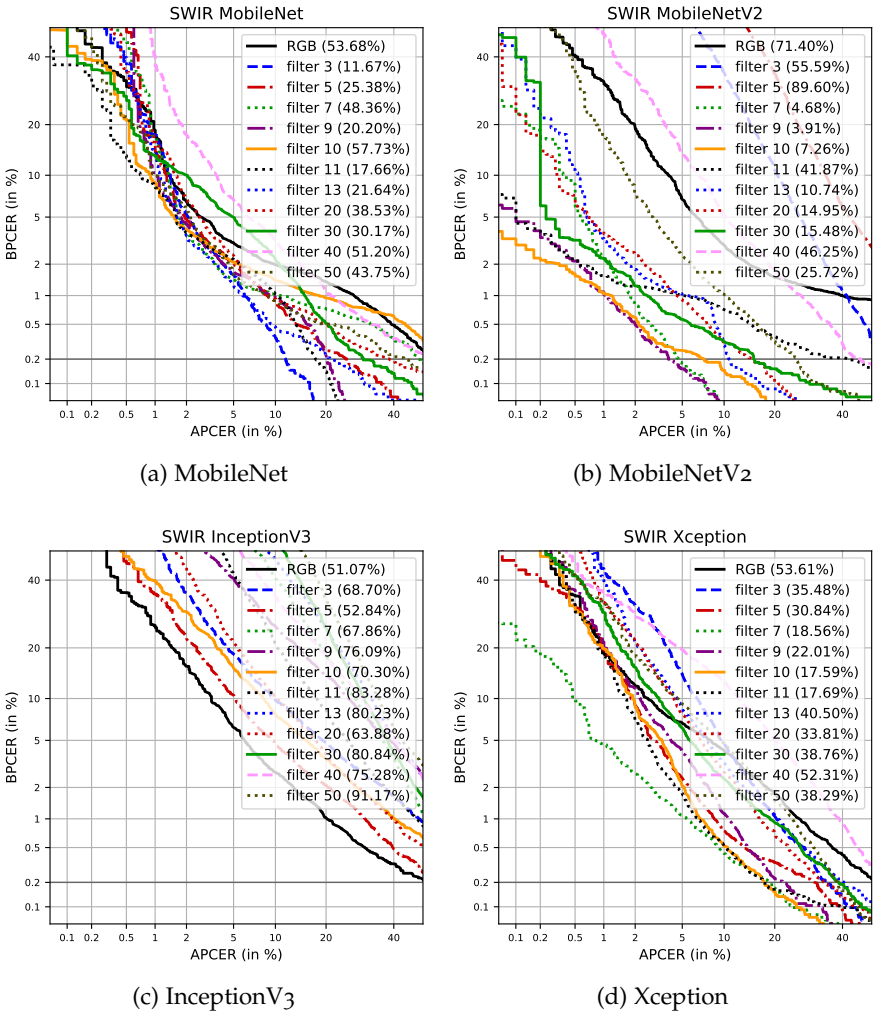
| | Training | Validation | Test |
|---|---|---|---|
| Baseline (BF) | 807 | 542 | 16,381 |
| Baseline (PA) | 807 | 542 | 2,990 |
| Unknown (BF) | 5,717 | 3,553 | 10,441 |
| Unknown (PA) | 0 | 0 | 4,339 |
| LOO (BF) | 9,956 | 3,069 | 6,686 |
| Fakefinger | 2,624 | 450 | 1,265 |
| Overlay | 1,027 | 238 | 3,074 |
| Opaque | 1,801 | 354 | 2,184 |
| Transparent | 3,152 | 674 | 513 |
| Semi | 3,299 | 663 | 377 |
| Mat. group i) | 2,657 | 541 | 1,141 |
| Mat. group ii) | 2,023 | 319 | 1,997 |
| Mat. group iii) | 2,884 | 595 | 860 |
| Mat. group iv) | 3,364 | 648 | 327 |

Table 3.11: Specifications of the used dataset partitions.

respectively. The AEs need slightly less memory than the CNNs as their models are not as deep. As generally for deep learning tasks, training the models is considered expensive due to the amount of data, time, and resources required. On the other hand, the more important predictions for fingerprint PAD are executed in real time with the samples of one capture attempt. The fingerprint PAD algorithms are implemented with Keras [54], which is a deep learning library for Python that includes tools and definitions for several models. For the CNNs and LSTMs the *Adam* [172] optimiser was chosen, while the AEs utilise *RMSprop* [42]. The learning rate is uniformly set to 0.0001.

### 3.2.9 *Experimental Results*

This Section comprises the results of the different experiments for the fingerprint PAD methods. Due to the extensive benchmarks of different settings and architectures, the results are presented in subsections divided into the following topics: SWIR CNNs, laser CNNs and LSTMs, OC-AEs, fusion of SWIR and laser, and finally LOO experiments.




(a) MobileNet (b) MobileNetV2

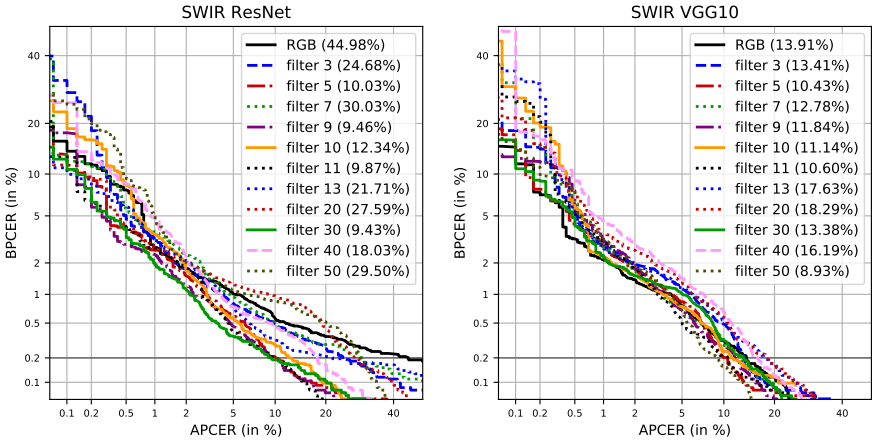
(c) InceptionV3 (d) Xception

Figure 3.41: SWIR benchmark results (1/2). $APCER_{0.2}$ values are given in brackets.

#### 3.2.9.1 *SWIR Fingerprint PAD Results*

The SWIR experiments[27] mainly benchmark the CNNs performance based on handcrafted RGB conversions in contrast to trained convolutional operations. Both approaches reduce the four captured SWIR wavelengths to a 3-dimensional image that is then further processed by identical CNNs. In the case of the convolutional input pre-processing, additionally the impact of different filter sizes is analysed. The DET curves for all SWIR fingerprint PAD algorithms on the baseline partition are shown in Figure 3.41. Starting with the RGB conversion, all tested filter sizes are separately plotted. Additionally, $APCER_{0.2}$ values are included to analyse the results at a particular convenient operation point.

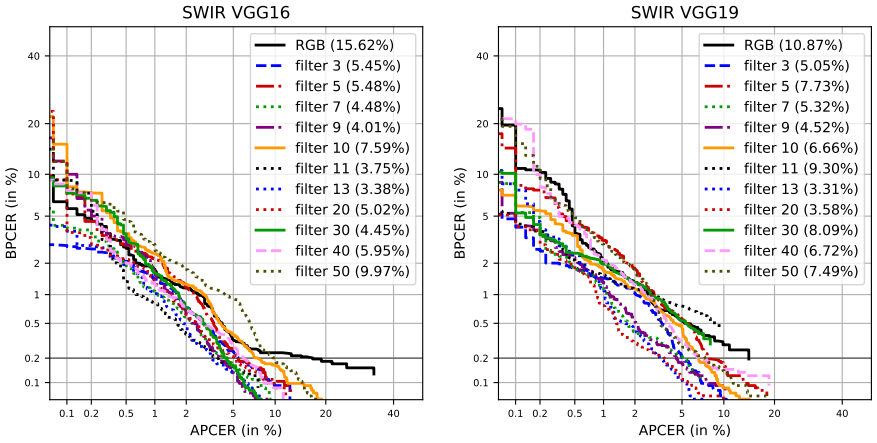
27 This Section is based on our publication [117] but these experiments are run on a different partition and additional CNNs are tested.




(e) ResNet



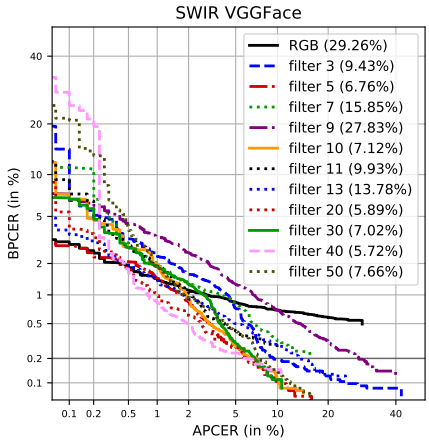

(f) VGG10




(g) VGG16




(h) VGG19




(i) VGGFace

Figure 3.41: SWIR benchmark results (2/2). $APCER_{0.2}$ values are given in brackets.

The DET curves for the MobileNet model (Figure 3.41a) are generally close together but split significantly when comparing the $APCER_{0.2}$ values. The RGB conversion achieves an $APCER_{0.2}$ = 53.68%, while the $APCER_{0.2}$ of the 4-dimensional CNNs range between 58% and 11% for different filter sizes. In contrast to this, the MobileNetV2 results (Figure 3.41b) spread much farther to both ends. The worst performance is an $APCER_{0.2}$ = 89.6% (filter 5), followed by the RGB conversion ($APCER_{0.2}$ = 71.4%). On the other hand, four filter sizes (7, 9, 10, 13) show better performances than the best MobileNet setup, reaching a minimum of $APCER_{0.2}$ = 3.91% for a filter size of 9. InceptionV3 (Figure 3.41c) is the only CNN where the RGB conversion achieves the best detection accuracy compared to all tested filter sizes of the 4-dimensional models. However, with all $APCER_{0.2}$ values above 51% this CNN architecture generally seems unsuited for this particular fingerprint PAD task. On the other hand, the optimisations within Xception (Figure 3.41d) result on average in lower error rates. Again, the RGB version is on the high end ($APCER_{0.2}$ = 53.61%) and the $APCER_{0.2}$ can be reduced to 17.59% for a filter size of 10.

The small CNNs that are trained from scratch, ResNet (Figure 3.41e) and VGG10 (Figure 3.41f), achieve equally good results. The RGB conversion of ResNet can be considered an outlier with an $APCER_{0.2}$ above 44% as all other setups achieve significantly lower error rates. The lowest $APCER_{0.2}$ values of around 9% can be seen for a filter size of 30 (ResNet) and 50 (VGG10), respectively. The best performances in average can be observed for the VGG16 (Figure 3.41g) and VGG19 (Figure 3.41h) CNNs, where all, except for the RGB conversion, $APCER_{0.2}$ values are below 10%. Furthermore, both report the lowest error rates for a filter size of 13 with an $APCER_{0.2}$ = 3.38% (VGG16) and an $APCER_{0.2}$ = 3.31% (VGG19), which is the best result of all SWIR algorithms. In this regard, VGGFace (Figure 3.41i) performs slightly worse with $APCER_{0.2}$ values between 30% (RGB) and 5.72% (filter 40).

Overall, the results show a clear tendency that the proposed 4-dimensional input processing block achieves superior PAD performance in contrast to the handcrafted RGB conversion. However, this requires additional evaluations as the filter sizes need to be optimised in order to utilise the full potential. A summary of the best setups across all analysed CNNs is depicted in Figure 3.42.

Finally, given a fixed operation point of BPCER = 0.2%, the corresponding Attack Presentation Classification Errors (APCEs) of the four best models are reviewed in Table 3.12. From the *fakefinger* group, mostly dragonskin and playdoh samples were misclassified. While especially orange playdoh is known to reflect SWIR illumination in the same way as bona fide skin does [110, 111], dragonskin PAIs yield even more errors. Across all overlays, silicone PAIs are the most challenging ones as they are made in various colours and transparency levels.

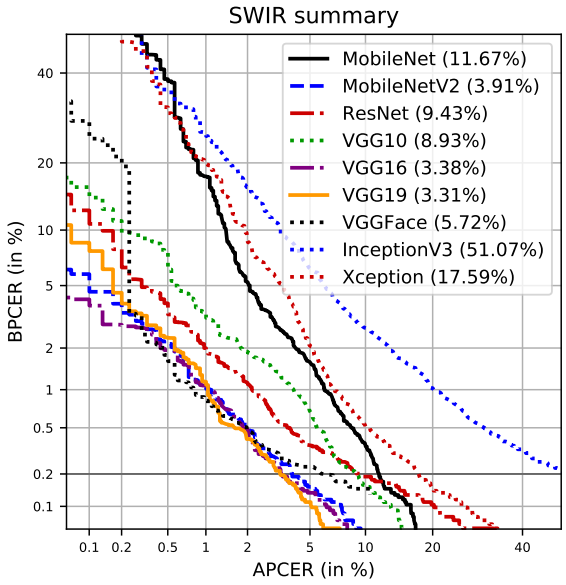



Figure 3.42: Summary of the best SWIR fingerprint PAD results. $APCER_{0.2}$ values are given in brackets.

However, also dragonskin and glue overlays are generally harder to detect than overlays fabricated from ecoflex or wax.

All in all, this benchmark included classical RGB CNNs and a multi-spectral counterpart with a 4-dimensional input layer. In the first case, a handcrafted RGB conversion combines the four captured SWIR wavelengths into one 3-dimensional RGB image. In the second case, a pre-processing module is added to the CNNs and trained to automatically transform 4-dimensional input images into three dimensions to serve as input for the classical CNNs. The results show, that this approach outperforms the handcrafted conversion in all cases but one.

#### 3.2.9.2 *Laser Fingerprint PAD Results*

Given the laser sequence of 100 frames, both CNN and LSTM setups are evaluated[28] to benchmark spatial and temporal classification methods. The same CNNs are used to classify single frames (end-to-end) and to extract feature vectors from the whole sequence. Hence, the focus is whether the additional LSTM module significantly improves the fingerprint PAD performance in contrast to the stand-alone CNNs.

The results in terms of DET curves for all laser CNNs and LSTMs are depicted in Figure 3.43. From the CNN plot it can be seen that four architectures are unsuited for fingerprint PAD based on laser images as their $APCER_{0.2}$ values are above 70%: MobileNet, MobileNetV2, InceptionV3, and Xception. Thus, the focus is on the other CNNs as all report an $APCER_{0.2}$ below 20%. The best results from those CNNs are

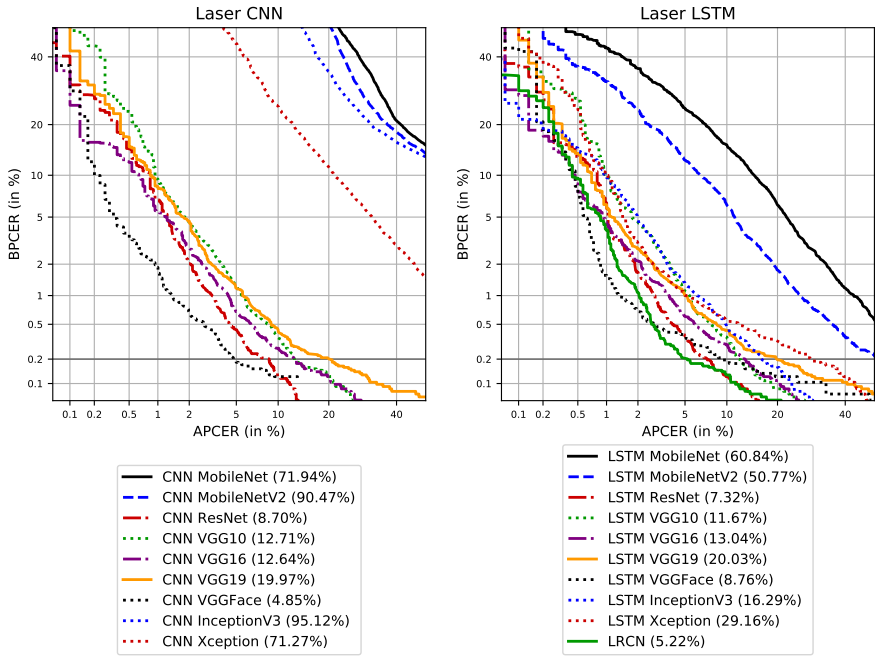
28 This Section is based on our publication [179] but additional LSTMs are tested.

| Group | PAI | MobileNetV2 | VGG16 | VGG19 | VGGFace |
|---|---|---|---|---|---|
| fakefinger | dragonskin | 27 | 33 | 38 | 49 |
| | ecoflex | 3 | 0 | 0 | 10 |
| | latex | 0 | 2 | 2 | 6 |
| | playdoh | 24 | 13 | 23 | 25 |
| | silly putty | 1 | 10 | 7 | 6 |
| ov. opaque | ecoflex | 1 | 2 | 4 | 6 |
| | gelatin | 1 | 1 | 0 | 1 |
| | silicone | 9 | 10 | 4 | 10 |
| ov. transparent | dragonskin | 3 | 15 | 2 | 7 |
| | gelatin | 4 | 0 | 0 | 2 |
| | glue | 3 | 0 | 0 | 0 |
| | silicone | 35 | 10 | 16 | 33 |
| | wax | 0 | 1 | 0 | 1 |
| ov. semi | dragonskin | 0 | 1 | 1 | 2 |
| | glue | 2 | 2 | 2 | 6 |
| | silicone | 2 | 0 | 0 | 4 |
| total | | 115 | 100 | 99 | 168 |

Table 3.12: Number of APCEs at an $APCER_{0.2}$ for the best SWIR algorithms.

achieved by the VGGFace model with an $APCER_{0.2}$ = 4.85%, followed by ResNet ($APCER_{0.2}$ = 8.7%).

When adding the LSTM module and taking into account all 100 frames, the results change as shown in Figure 3.43b. Interestingly, only both MobileNets report $APCER_{0.2}$ values above 50%. Although this is a big improvement towards the pure CNN approach, a more significant drop can be seen for InceptionV3 (16.29%) and Xception (29.16%). Hence, both approaches thrive due to the added LSTM part underlining the importance of the temporal information within the laser sequence. Whereas, the other combinations achieve even better results in terms of a low $APCER_{0.2}$, they do not show significant improvements towards their pure CNN method. However, another important observation can be made by analysing the results of the VGG10 LSTM to the LRCN. Both share basically the same structure with the difference that the first CNN is trained on one frame of the sequence and thus extracts only the feature for the following LSTM module and the LRCN jointly trains convolutional and LSTM layers on the full sequence. As a result, the $APCER_{0.2}$ of the LRCN is half as much as for VGG10 LSTM (5.22% to 11.67%). Hence, this further proves the relevance of the temporal context within the captured image




(a) CNNs (b) LSTMs

Figure 3.43: Laser benchmark results. $APCER_{0.2}$ values are given in brackets.

sequence. However, even the LRCN is outperformed by the VGGFace CNN.

Since these results are still very close (only 0.4% apart) and especially due to the immense improvement for the InceptionV3 ($APCER_{0.2}$ decreased from 95% to 16%), the additional LSTM module can be considered beneficial for this fingerprint PAD task. However, the results also reveal that some CNN architectures are better suited than others. This supports the conclusion that targeted adjustments in the structure could further improve the fingerprint PAD performance for both CNN and LSTM algorithms.

Finally, in accordance with the previous SWIR analysis, the APCEs of the best algorithms are further discussed in Table 3.13. From the *fakefinger* group, the most classification errors occur for dragonskin PAIs, followed by playdoh fingers. Additional misclassified samples can be counted for latex and silly putty materials. However, the LRCN model seems stronger versus the latter ones as its errors are mainly caused by attack presentations using dragonskin. Besides, dragonskin further troubles all classifiers across the different overlay groups. Nevertheless, a higher number of APCEs can be observed for silicone overlays. Especially, the transparent PAIs are not detected to a great extent. Other PAI species made from glue, gelatin, or ecoflex appear as well in the listed errors, but not in a comparable volume as silicone ones. Hence, the most challenging attack presentations are dragonskin fingers and silicone overlays.

| Group | PAI | CNN ResNet | CNN VGGFace | LSTM ResNet | LRCN |
|---|---|---|---|---|---|
| fakefinger | dragonskin | 104 | 33 | 85 | 59 |
| | ecoflex | 1 | 3 | 1 | 2 |
| | latex | 7 | 1 | 7 | 1 |
| | playdoh | 14 | 13 | 9 | 2 |
| | silly putty | 6 | 7 | 5 | 0 |
| ov. opaque | dragonskin | 4 | 1 | 3 | 2 |
| | ecoflex | 13 | 8 | 15 | 0 |
| | gelatin | 9 | 9 | 4 | 10 |
| | silicone | 11 | 5 | 8 | 9 |
| ov. transparent | dragonskin | 9 | 2 | 6 | 6 |
| | gelatin | 2 | 1 | 1 | 3 |
| | glue | 5 | 2 | 4 | 3 |
| | silicone | 51 | 39 | 46 | 48 |
| ov. semi | dragonskin | 4 | 3 | 4 | 3 |
| | glue | 3 | 1 | 1 | 5 |
| | silicone | 12 | 15 | 17 | 0 |
| total | | 255 | 143 | 216 | 153 |

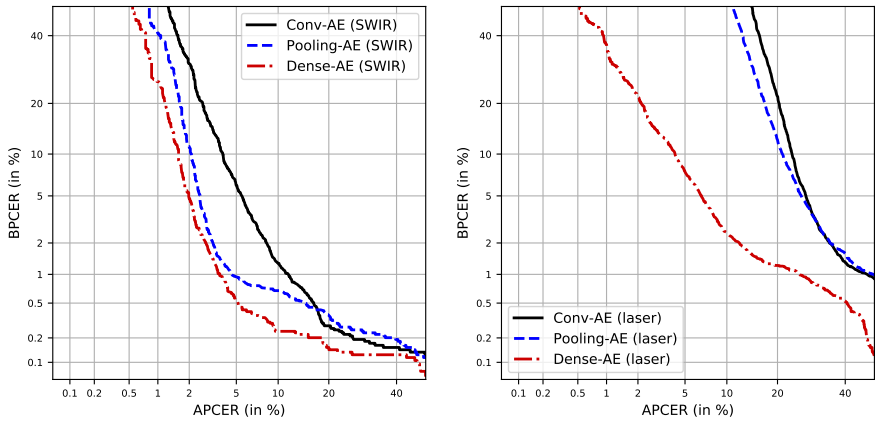
Table 3.13: Number of APCEs at an $APCER_{0.2}$ for the best laser algorithms.

This benchmark revealed that the suitability of network architectures depends on the input data. Whereas the MobileNet CNNs achieved remarkable results on the SWIR data, their performance on the laser data is unusable. Moreover, the InceptionV3 CNN reported the highest $APCER_{0.2}$ on single frames but achieves significantly improvements as LSTM when working on the full laser sequence. Finally, the relevance of temporal information for fingerprint PAD is underlined by the LRCN results that show nearly no errors (except for one PAI species) in the *fakefinger* class.

#### 3.2.9.3 *One-class Fingerprint PAD Results*

The next set of experiments[29] viewed all PAI species as unknown attacks. In contrast to the previous methods, which were analysed at a fixed operation point $APCER_{0.2}$, in this case it is more interesting to observe the general performance. The Area Under Curve (AUC) [30] is a good measure for this task, with small values indicating a more accurate system in the case of DET plots. However, as PAD systems are only usable in a particular range, this Section defines the partial

29 This Section is based on our publication [178].




(a) SWIR (b) Laser

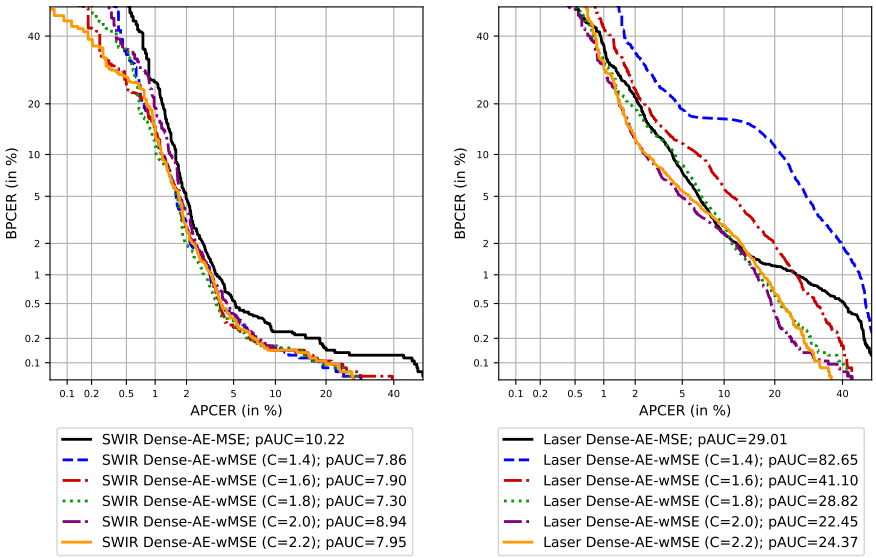
Figure 3.44: DET curves for the three baseline AE architectures: Conv-AE, Pooling-AE, and Dense-AE.

AUC (pAUC) that limits the area at 20% error rate. As a result, the regions above 20% error rate are ignored and not taken into account for benchmarking the systems.

Figure 3.44 allows a side by side analysis of the three baseline AE architectures (Conv-AE, Pooling-AE, and Dense-AE) for SWIR and laser data. The DET curves show that the Dense-AE (red) clearly outperforms the other two architectures for all possible decision thresholds as the curve is constantly below the others. Since the AEs task is to reconstruct the original input from the latent representation, it can be concluded that the Dense-AE extracts the most relevant features for this task. Therefore, the following evaluations focus on this structure and discard the other architectures.

In the following, the MSE loss function from Eq. (3.9) is replaced by wMSE as defined in Eq. (3.13). Therefore, new models of the Dense-AEs are trained for different constant $C$ values and the resulting DET curves are depicted in Figure 3.45. For completeness, the MSE version is again included to visualise the impact of the loss function. It can be observed that the pAUC decreases for growing values of $C$, thus enhancing the detection performance of the model. On the contrary, too low $C$ values imply the exclusion of large parts of the image that equals a loss of information. This aspect changes when reaching values of $C = 2.0$ (SWIR) and $C = 2.2$ (laser), where the threshold gets too high that less of the pixel-wise REs exceed it, thus noisy areas are still included. Therefore, the best configurations are $C = 1.8$ (SWIR) and $C = 2.0$ (laser), resulting in a pAUC of 7.3% and 22.45%, respectively.

Given the previous results that the Dense-AE architecture in combination with the wMSE loss function provides the best results for both SWIR and laser data, the next step is to validate whether both input data complement each other to detect more attack presentations. In




(a) SWIR (b) Laser

Figure 3.45: Performance of the Dense-AEs on SWIR and laser data for MSE and wMSE settings. The pAUC is given in %.

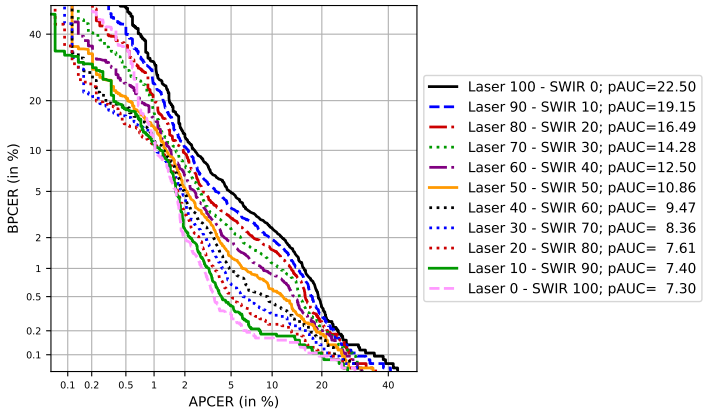


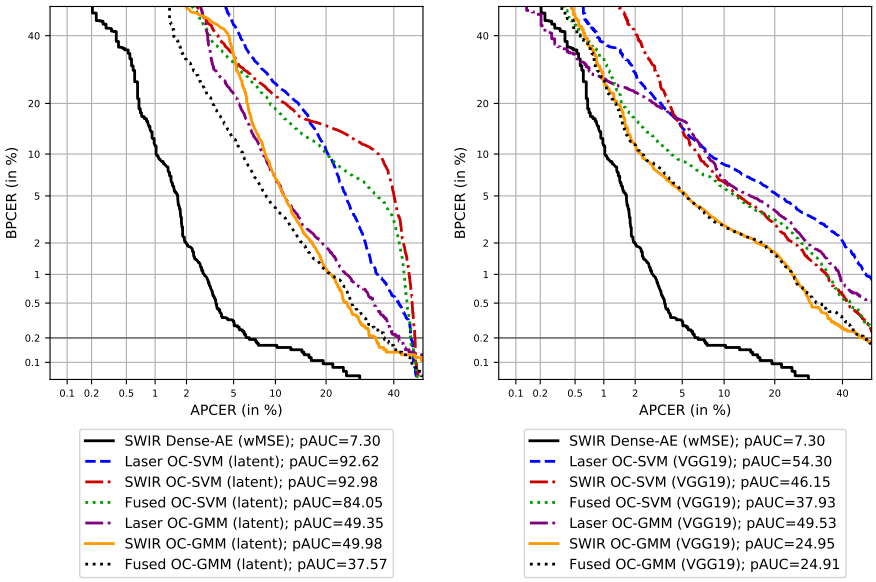

Figure 3.46: Weighted score fusions of the best-performing wMSE Dense-AEs. The pAUC is given in %.

| Group | PAI | SWIR Dense-AE |
|---|---|---|
| fakefinger | dragonskin | 14 |
| | playdoh | 74 |
| ov. opaque | bandage | 2 |
| ov. transparent | dragonskin | 28 |
| | gelatin | 16 |
| | glue | 14 |
| | silicone | 120 |
| ov. semi | dragonskin | 5 |
| | ecoflex | 3 |
| | glue | 13 |
| total | | 290 |

Table 3.14: Number of APCEs at an $APCER_{0.2}$ for the SWIR wMSE Dense-AE.

this regard, multiple weighted score fusions are tested as visible in Figure 3.46. The DET curves show the different performances for fusion weights that are adjusted in steps of 10%. Additionally, the pAUC values constantly decrease for increasing SWIR weights, reaching its minimum of 7.3% using only the SWIR AE. In this case, including the laser AE has a negative impact on the fingerprint PAD performance. Nevertheless, the laser AE adds value to the fusion for high security applications of APCER < 1%. However, as this is only a small part of the plot that additionally exceeds the 20% error limit, this is not visible from the pAUC values.

The occurring APCEs for the $APCER_{0.2}$ operation point reveal that the complete set of SWIR errors is a subset of the laser errors, thus explaining the fusions behaviour for convenient scenarios. The particular number of APCEs per PAI material for the SWIR model are summarised in Table 3.14. Most misclassifications are due to transparent overlays, first of all two part silicone, and the next biggest share results from orange and yellow playdoh fingers. However, when viewing the absolute numbers, it is important to note that no PA samples were used in the training sets, hence all attack presentations are used for testing only. In general, the reconstructions for all APCEs were close enough to bona fide reconstructions during training such that the AE failed to distinguish them. The set of laser APCEs additionally includes more *fakefingers* of ecoflex and latex and higher numbers for the overlay groups. As the laser diode uses only one specific wavelength within the SWIR domain, the laser samples of PAI species generally are more similar to bona fide presentations than those across all four SWIR wavelengths.




(a) SWIR (b) Laser

Figure 3.47: Benchmark of the best AE setup towards additional OC classifiers based on two different features.

Since a performance benchmark with the aforementioned two-class fingerprint PAD methods, that are tested on less PA samples, is not fair, additional OC classifiers are trained and tested on the same data partitions. However, as both OC-SVMs and OC-GMMs cannot be trained on images but require 1-dimensional feature vectors, two pre-trained networks are utilised to extract those. On the one hand, the latent representations of the Dense-AEs are extracted and on the other hand, the fine-tuned VGG19 CNN is modified for this task. The results in terms of DET curves and pAUC (%) are displayed in Figure 3.47. In addition to the SWIR and laser results, both algorithms are also fused in case it improves the fingerprint PAD performance. The analysis shows for all cases that the OC-GMMs performs better than its counterpart (OC-SVMs), which confirms the findings of Nikisins et al. [217]. However, the proposed AE is unbeaten as its DET curve remains significantly below the others. Interestingly, the fusions for the other methods yield better results even for the latent AE representation in contrast to the AEs.

Finally, Table 3.15 lists the performances in terms of $APCER_{0.2}$ and D-EER, allowing to connect these results to other fingerprint PAD methods that do not evaluate the defined pAUC. Furthermore, pAUC measurement is better suited to evaluate the general performance in contrast to particular operation points, which are required to validate whether the proposed PAD algorithm is applicable for the wanted scenario. In addition to the SWIR AE, also the laser AE achieves significantly better results than the other OC algorithms. The next

| Algorithm | $APCER_{0.2}$ | D-EER |
|---|---|---|
| Laser Dense-AE (wMSE) | 24.33% | 4.96% |
| SWIR Dense-AE (wMSE) | **6.59%** | **2.00%** |
| Laser OC-GMM (latent) | 41.25% | 8.80% |
| SWIR OC-GMM (latent | 34.17% | 8.98% |
| Fused OC-GMM (latent) | 36.88% | 7.09% |
| Laser OC-SVM (latent) | 45.95% | 16.21% |
| SWIR OC-SVM (latent) | 47.27% | 16.18% |
| Fused OC-SVM (latent) | 45.86% | 13.93% |
| Laser OC-GMM (VGG19) | 63.98% | 8.64% |
| SWIR OC-GMM (VGG19) | 47.43% | 5.21% |
| Fused OC-GMM (VGG19) | 47.41% | 5.16% |
| Laser OC-SVM (VGG19) | 65.12% | 9.07% |
| SWIR OC-SVM (VGG19) | 52.82% | 7.96% |
| Fused OC-SVM (VGG19) | 55.55% | 7.17% |

Table 3.15: Performance overview of the best AEs in contrast to other OC classifiers.

best results are an $APCER_{0.2}$ = 34.17% (latent SWIR OC-GMM) and a D-EER = 5.16% (VGG19 fused OC-GMM). All in all, the SWIR AE achieves remarkable results with $APCER_{0.2}$ = 6.59% and D-EER = 2%.

#### 3.2.9.4 *Fusion*

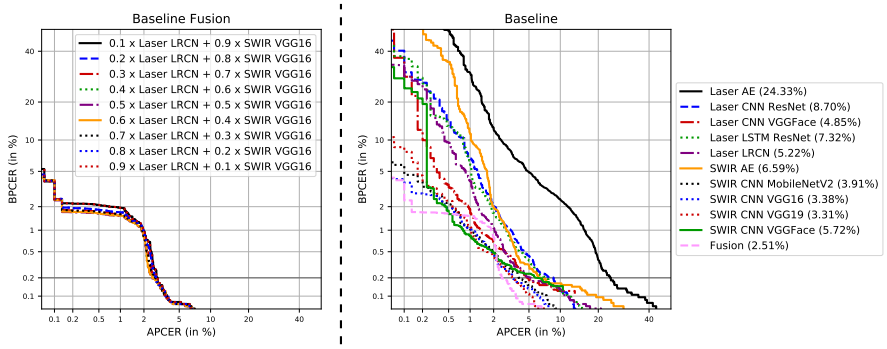
Given all the previous results, the best-performing algorithms[30] are considered for a fusion. As defined in Section 3.2.6, this fusion intends to combine one SWIR algorithm with one laser algorithm in order to limit the number of possibilities and prevent over-engineering on the baseline partition. In this context, the number of identical APCEs are listed in Table 3.16 to validate how the different PAD algorithms complement each other.

The lowest number of identical APCEs (23) is counted for the combination of laser LRCN and SWIR CNN VGG16. Additionally, both algorithms have the lowest average in their row (LRCN) and their column (VGG16), respectively. Hence, this combination is found to be most complementary and suited for the fusion. Since the optimal fusion exploits the strengths from both algorithms, it is generally a good idea to fuse those which have the least APCEs in common.

30 This Section is based on our publication [176] but additional PAD algorithms are considered for the fusion.

| Laser | SWIR AE | SWIR CNN | | | |
|---|---|---|---|---|---|
| | | Mob.NetV2 | VGG16 | VGG19 | VGGFace |
| | (290) | (117) | (101) | (99) | (171) |
| AE (1,070) | 285 | 101 | 61 | 82 | 120 |
| CNN ResNet (260) | 80 | 64 | 45 | 50 | 73 |
| CNN VGGFace (145) | 52 | 49 | 29 | 42 | 59 |
| LSTM ResNet (219) | 69 | 57 | 35 | 43 | 63 |
| LRCN (156) | 66 | 48 | 23 | 28 | 48 |

Table 3.16: Number of identical APCEs for the best-performing algorithms. The total number of APCEs per algorithm is given in brackets.




(a) Fusion trends

(b) Summary of the best results

Figure 3.48: Overview of different weighted fusions (a) and the best PAD algorithms with their corresponding $APCER_{0.2}$ values (b).

The evaluation of different fusion weights in Figure 3.48a approves the complementary fusion of SWIR CNN VGG16 and laser LRCN. Independently of the weights, all DET curves are mostly overlapping and appear as one thick line. Hence, the greatest benefit arises from the fact of fusing alone and is no over-engineered niche solution. Since, the results are so close, equal weights of 50% each are chosen to allow the most general design. This fusion and the DET curves of the best-performing algorithms are plotted in Figure 3.48b. For both extreme cases of high convenience ($APCER_{0.2}$) or high security ($BPCER_{0.2}$), the fusion is an enhancement compared to the single fingerprint PAD algorithms. However, its D-EER is slightly higher than the best CNN approaches. Finally, these fusion weights are fixed for the following experiments as these analyse the generalisability of this subset of algorithms.

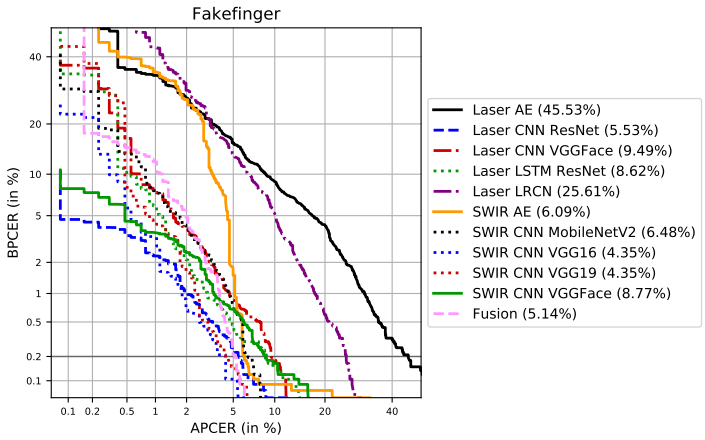



Figure 3.49: DET curves of the *Fakefinger* group and their corresponding $APCER_{0.2}$ values.

#### 3.2.9.5 *PAD Generalisability*

The objective of this Section[31] is to evaluate the generalisation capabilities of the introduced fingerprint PAD methods towards unknown attacks. Therefore, the subset of best-performing algorithms from Figure 3.48b is consecutively trained and tested on the specified LOO partitions and the resulting DET curves are plotted to visualise the differences. Furthermore, the APCEs of the pre-defined fusion and the best two laser and SWIR algorithms are analysed in more detail for the convenient operation point $APCER_{0.2}$. As ResNet appears twice among the best laser algorithms, the CNN ResNet is now abbreviated *C.ResNet* within Tables and the LSTM ResNet *L.ResNet*, respectively. The fusion combines the laser LRCN and the SWIR CNN VGG16 as defined above.

**Visual LOO Groups.** The visual LOO partitions consist of two PAI categories: full fake fingers and fingerprint overlays. Moreover, the overlays are further grouped due to their appearance regarding transparency into opaque, transparent, and semi transparent. Since a camera-based capture device is utilised, the captured samples might be different when the bona fide skin is still visible behind a transparent overlay PAI.

Starting with the fakefinger group, the corresponding DET curves are plotted in Figure 3.49. As the algorithms are trained on overlay PAIs only, the overall PAD performance decreases compared to the baseline partition. However, the best algorithms (SWIR CNNs VGG16 and VGG19) achieve an $APCER_{0.2}$ = 4.35% (+1%), followed by the fusion with 5.14% (+2.6%). On the other hand, laser CNN ResNet improves its $APCER_{0.2}$ by 3%, while e. g. laser LRCN has a 20% higher error rate (25.61%). Despite the fact that the fusion weights were

31 This Section is based on our publication [176] but additional PAD algorithms are evaluated.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | L.ResNet | VGG16 | VGG19 | |
| dragonskin | 30 (6.29%) | 85 (17.82%) | 3 (0.63%) | 6 (1.26%) | 8 (1.68%) |
| latex | 0 | 3 (2.04%) | 1 (0.68%) | 0 | 0 |
| playdoh | 39 (33.62%) | 21 (18.10%) | 48 (41.38%) | 47 (40.52%) | 57 (49.14%) |
| silly putty | 1 (1.82%) | 0 | 1 (1.82%) | 2 (3.64%) | 0 |
| wax | 0 | 0 | 2 (2.70%) | 0 | 0 |
| total | 70 (5.53%) | 109 (8.62%) | 55 (4.35%) | 55 (4.35%) | 65 (5.14%) |

Table 3.17: Summary of APCEs at an $APCER_{0.2}$ on the *Fakefinger* partition.

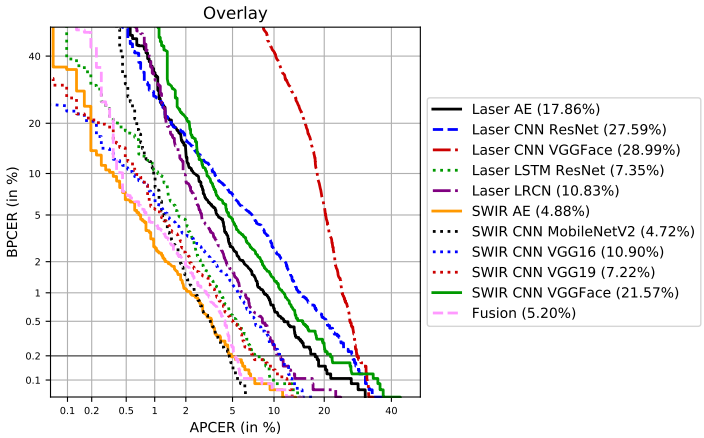



Figure 3.50: DET curves of the *Overlay* group and their corresponding $APCER_{0.2}$ values.

fixed on the baseline partition, the equal fusion of laser LRCN and SWIR CNN is still the third best algorithm. The resulting APCEs are summarised in Table 3.17. The highest number of errors as well as error rates are reported for playdoh PAIs. In fact, especially orange playdoh troubles the PAD algorithms when captured in the SWIR domain. Additionally, some yellow playdoh PAIs are mistaken for bona fide presentations while all other colours are classified correctly. Furthermore, the laser algorithms are challenged by dragonskin PAIs with up to 18% APCER (LSTM ResNet). Similar to the SWIR methods, the fusion is much better at classifying dragonskin presentations (APCER=1.68%) but fails in 50% of the playdoh PAIs (74% for orange and yellow samples only).

On the contrary, the next set of experiments is trained on fake fingers only and treats all overlays as unknown attacks (Figure 3.50). Again the fusion is on the third place ($APCER_{0.2}$ = 5.2%), but this time after SWIR CNN MobileNetV2 with 4.72% and SWIR AE with 4.88%. The laser CNNs show the worst performance ($APCER_{0.2}$ > 27%) and the LRCN improves to 10.8% $APCER_{0.2}$, which is 0.1% lower than its fusion partner SWIR CNN VGG16. Table 3.18 lists all APCEs for the

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | L.ResNet | LRCN | AE | Mob.NetV2 | |
| *opaque* | | | | | |
| bandage | 4 (28.57%) | 5 (35.71%) | 1 (7.14%) | 2 (14.29%) | 5 (35.71%) |
| dragonskin | 2 (11.76%) | 4 (23.53%) | 0 | 0 | 2 (11.76%) |
| ecoflex | 11 (1.06%) | 0 | 0 | 0 | 0 |
| gelatin | 9 (4.64%) | 20 (10.31%) | 0 | 0 | 0 |
| silicone | 19 (2.31%) | 70 (8.5%) | 0 | 2 (0.24%) | 26 (3.16%) |
| *transparent* | | | | | |
| dragonskin | 26 (24.53%) | 39 (36.79%) | 21 (19.81%) | 10 (9.43%) | 21 (19.81%) |
| gelatin | 8 (7.48%) | 33 (30.84%) | 13 (12.15%) | 47 (43.93%) | 15 (14.02%) |
| glue | 20 (74.07%) | 18 (66.67%) | 10 (37.04%) | 3 (11.11%) | 11 (40.74%) |
| latex | 1 (2.94%) | 8 (23.53%) | 0 | 0 | 4 (11.76%) |
| silicone | 87 (55.41%) | 113 (71.97%) | 97 (61.78%) | 74 (47.13%) | 71 (45.22%) |
| *semi* | | | | | |
| dragonskin | 8 (17.02%) | 15 (31.91%) | 0 | 5 (10.64%) | 2 (4.26%) |
| ecoflex | 2 (8.33%) | 5 (20.83%) | 2 (8.33%) | 2 (8.33%) | 3 (12.5%) |
| glue | 4 (2.74%) | 2 (1.37%) | 6 (4.11%) | 0 | 0 |
| silicone | 25 (15.63%) | 1 (0.63%) | 0 | 0 | 0 |
| total | 226 (7.35%) | 333 (10.83%) | 150 (4.88%) | 145 (4.72%) | 160 (5.2%) |

Table 3.18: Summary of APCEs at an $APCER_{0.2}$ on the *Overlay* partition.

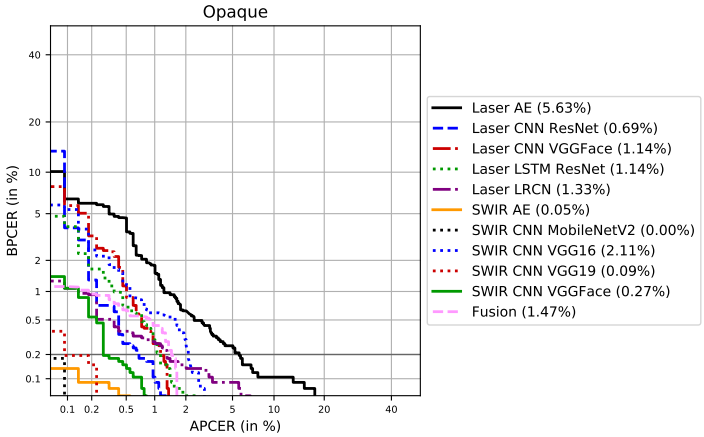



Figure 3.51: DET curves of the *Opaque* group and their corresponding $APCER_{0.2}$ values.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | VGGFace | AE | Mob.NetV2 | |
| bandage | 0 | 0 | 1 (7.14%) | 0 | 1 (7.14%) |
| ecoflex | 7 (0.68%) | 21 (2.03%) | 0 | 0 | 18 (1.74%) |
| gelatin | 5 (2.58%) | 1 (0.52%) | 0 | 0 | 1 (0.52%) |
| silicone | 3 (0.36%) | 3 (0.36%) | 0 | 0 | 12 (1.46%) |
| total | 15 (0.69%) | 25 (1.14%) | 1 (0.05%) | 0 | 32 (1.47%) |

Table 3.19: Summary of APCEs at an $APCER_{0.2}$ on the *Opaque* partition.

three specified overlay groups. From the opaque group, the highest error rates are reported for bandage plasters. Although these presentations are listed as concealing attack presentations, data subjects usually cover only a scratch such that other parts still reveal bona fide skin. Hence, up to 36% are misclassified as bona fide presentations. The error rates of the transparent group are dominated by silicone PAIs ($APCER_{0.2}$ between 45% and 72%), but glue, dragonskin, and gelatin PAIs are also challenging to classify correctly. Finally, the semi transparent group contains the least numbers of APCEs with relatively high error rates due to the smaller number of total samples in this group.

The best overall results can be observed for the opaque overlay group in Figure 3.51, where all DET curves are close to the bottom left corner and an $APCER_{0.2}$ of 5.6% can be seen as an outlier (Laser AE). Four algorithms obtain an $APCER_{0.2}$ below 0.3%, but especially the fused PAD algorithms are among the worst with $APCER_{0.2}$ values above 1.3%. As a result, this fusion is far behind the best results ($APCER_{0.2}$ = 1.5%). As shown in Table 3.19, the reason for this are misclassified samples made from ecoflex and silicone. On the other

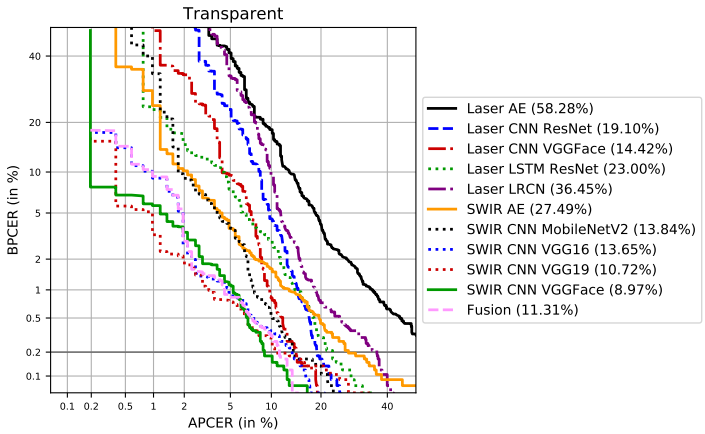



Figure 3.52: DET curves of the *Transparent* group and their corresponding $APCER_{0.2}$ values.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | VGGFace | VGG19 | VGGFace | |
| dragonskin | 15 (14.15%) | 6 (5.66%) | 17 (16.04%) | 2 (1.89%) | 7 (6.6%) |
| gelatin | 2 (1.67%) | 5 (4.67%) | 1 (0.93%) | 0 | 3 (2.80%) |
| glue | 7 (25.93%) | 9 (33.33%) | 1 (3.7%) | 1 (3.7%) | 3 (11.11%) |
| latex | 0 | 0 | 0 | 1 (2.94%) | 0 |
| silicone | 74 (47.13%) | 54 (34.39%) | 36 (22.93%) | 41 (26.11%) | 45 (28.66%) |
| total | 98 (19.1%) | 74 (14.42%) | 55 (10.72%) | 46 (8.97%) | 58 (11.31%) |

Table 3.20: Summary of APCEs at an $APCER_{0.2}$ on the *Transparent* partition.

hand, the detection rate of bandage plasters increases as soon as other partly covering overlays are included in the training set.

With the transparent overlay group next, the PAD performances decrease significantly (Figure 3.52). This time, the best $APCER_{0.2}$ is achieved by SWIR CNN VGGFace (8.97%) followed by VGG19 (10.72%). With an $APCER_{0.2}$ of 11.31%, the fusion is placed third despite the enormous $APCER_{0.2}$ of 36% for the laser LRCN. In accordance to the analysis on the full overlay group, most APCEs are caused by silicone, glue, and dragonskin as presented in Table 3.20. However, due to other overlays with the training set, the error rates are lower than for the transparent subset of the full overlay LOO experiment. However, this remains the most challenging LOO group in general.

The results from the semi transparent group in Figure 3.53 align with the previous observations between the opaque and transparent groups, with a noticeable shift towards the opaque results. Again, the laser AE can be considered an outlier and all other $APCER_{0.2}$ values remain below 6%. In particular, six PAD algorithms (including the fusion) achieve an $APCER_{0.2}$ below 1%. Table 3.21 highlights that the few occurring APCEs result from attack presentations made from

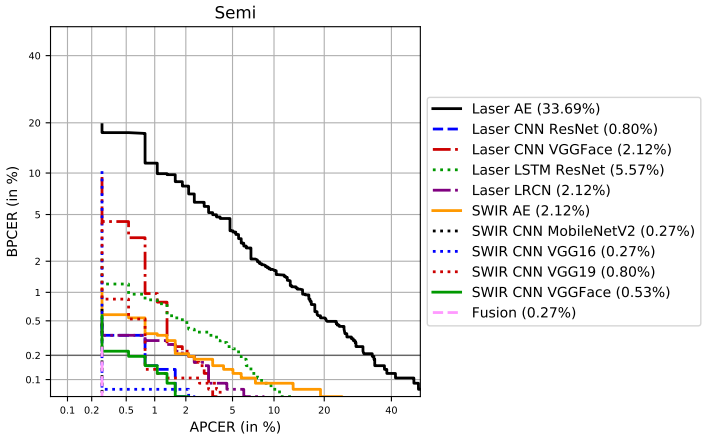



Figure 3.53: DET curves of the *Semi transparent* group and their corresponding $APCER_{0.2}$ values.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | LRCN | Mob.NetV2 | VGG16 | |
| dragonskin | 0 | 3 (6.38%) | 0 | 0 | 0 |
| ecoflex | 0 | 0 | 1 (4.17%) | 0 | 1 (4.17%) |
| silicone | 3 (1.88%) | 5 (3.13%) | 1 (0.63%) | 0 | 0 |
| total | 3 (0.8%) | 8 (2.12%) | 1 (0.27%) | 1 (0.27%) | 1 (0.27%) |

Table 3.21: Summary of APCEs at an $APCER_{0.2}$ on the *Semi* partition.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | VGGFace | VGG16 | VGG19 | |
| opaque | 4 (0.49%) | 2 (0.24%) | 1 (0.12%) | 1 (0.12%) | 0 |
| transparent | 68 (43.31%) | 30 (19.12%) | 22 (14.01%) | 30 (19.12%) | 23 (14.65%) |
| semi | 1 (0.63%) | 16 (10%) | 0 | 0 | 0 |
| total | 73 (6.4%) | 78 (6.84%) | 23 (2.02%) | 31 (2.72%) | 23 (2.02%) |

Table 3.22: Summary of APCEs at an $APCER_{0.2}$ for material group i) partition.

dragonskin, ecoflex, or silicone. The lower number of testing samples results in five times higher error rates compared to the opaque group given the same number of misclassified attack presentations.

So far the results indicate that the PAD performance for training only on fake fingers or on overlays does not differ much towards the baseline partition that includes all PAI species. On the other hand, it is much more challenging to detect unknown transparent overlays than opaque ones. Semi transparent overlays are quiet easy to detect but single errors cause higher error rates due to the small test set. Converting this knowledge to the baseline partition means that its performance is based on the particular shares of (un)challenging PAI groups. An additional insight is that the laser LRCN is not able to correctly classify fakefinger presentations, even though they have no blood movement, when the training set contains only overlay PAIs. Hence, the LRCN model is highly sensitive to its training data which contradicts the ambition of generalisability. As a consequence, combining other PAD methods could further improve the fusion results [176]. However, this equal contribution of laser LRCN and SWIR CNN VGG16 still does a decent job across the different LOO experiments ($1 \times 1^{st}, 3 \times 3^{rd}, 1 \times 9^{th}$).

**Material LOO Groups.** In contrast to the previously evaluated visual LOO groups, the PAIs are additionally grouped based on the utilised material. Due to the varying number of captured samples per PAI species, four material groups were defined in Section 3.2.8:

i) silicone

ii) dragonskin and ecoflex

iii) gelatin, glue, latex, printout, and wax

iv) 3D printed, dental material, playdoh, and silly putty

The LOO results for material group i) are depicted in Figure 3.54. The best results are reported for the fusion and SWIR CNN VGG16 ($APCER_{0.2}$ = 2.02%), while laser AE and LRCN have the highest $APCER_{0.2}$ around 13%. Furthermore, all SWIR algorithms perform better than the best laser algorithm, which is also visible from the APCEs

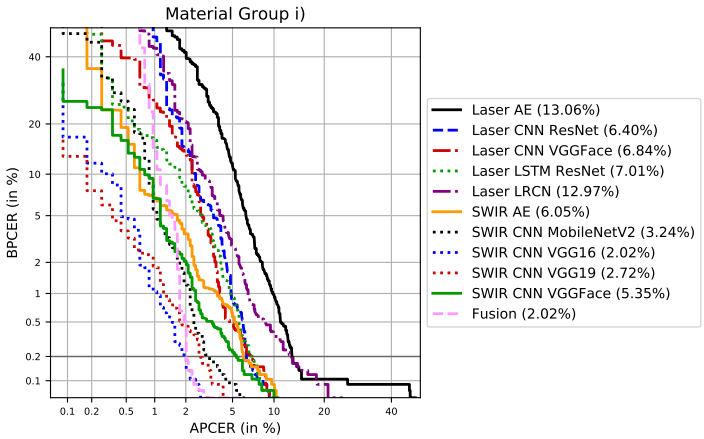



Figure 3.54: DET curves of the material group i) and their corresponding $APCER_{0.2}$ values.

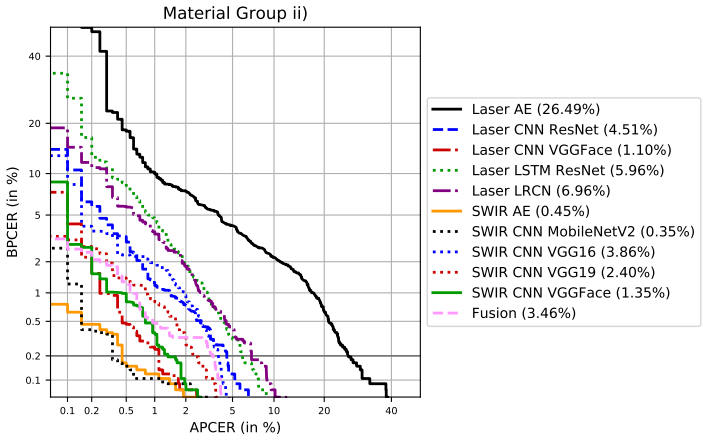



Figure 3.55: DET curves of the material group ii) and their corresponding $APCER_{0.2}$ values.

in Table 3.22, where the particular number of misclassified samples is two to three times higher for the laser CNNs than for the SWIR CNNs. Generally, the majority of undetected silicone presentations are transparent PAIs.

In the next step, material group ii) is evaluated in Figure 3.55. The results show that it seems much easier to detect unknown dragonskin and ecoflex PAIs for particular SWIR algorithms with remarkable low $APCER_{0.2}$ values of 0.35% (MobileNetV2) and 0.45% (AE), followed by laser CNN VGGFace (1.1%). The fusion (3.46%) achieves an average performance on the fifth place. The summary in Table 3.23 reveals that dragonskin PAIs cause the most errors.

Overall better PAD performances are reported for material group iii) in Figure 3.56. The fusion and three additional SWIR CNNs report an $APCER_{0.2}$ of 0.12%, which is equal to one misclassified attack presentation. The laser algorithms (except the AE) report more errors but are generally closer than for other LOO scenarios. As shown in

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | VGGFace | AE | MobileNetV2 | |
| dragonskin | 82 (12.67%) | 19 (2.94%) | 9 (1.39%) | 4 (0.62%) | 45 (6.96%) |
| ecoflex | 8 (0.59%) | 2 (0.15%) | 0 | 3 (0.22%) | 24 (1.78%) |
| total | 90 (4.51%) | 21 (1.1%) | 9 (0.45%) | 7 (0.35%) | 69 (3.46%) |

Table 3.23: Summary of APCEs at an $APCER_{0.2}$ for material group ii) partition.

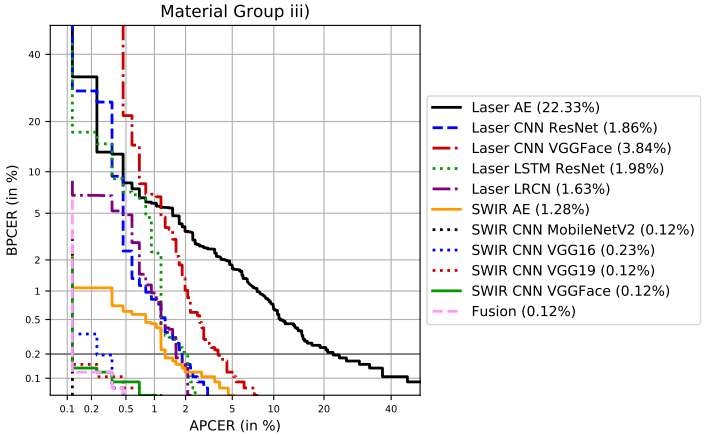



Figure 3.56: DET curves of the material group iii) and their corresponding $APCER_{0.2}$ values.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | C.ResNet | LRCN | Mob.NetV2 | VGG19 | |
| gelatin | 7 (2.33%) | 2 (0.66%) | 0 | 0 | 0 |
| glue | 8 (4.62%) | 11 (6.36%) | 1 (0.58%) | 1 (0.58%) | 1 (0.58%) |
| latex | 1 (0.55%) | 1 (0.55%) | 0 | 0 | 0 |
| total | 16 (1.86%) | 14 (1.63%) | 1 (0.12%) | 1 (0.12%) | 1 (0.12%) |

Table 3.24: Summary of APCEs at an $APCER_{0.2}$ for material group iii) partition.

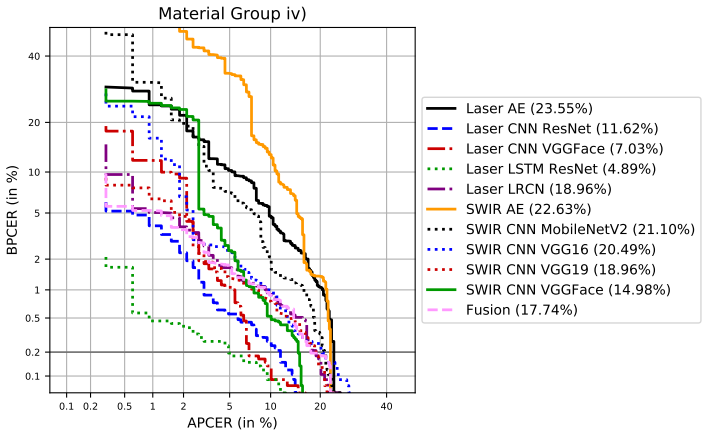



Figure 3.57: DET curves of the material group iv) and their corresponding $APCER_{0.2}$ values.

| PAI | Laser | | SWIR | | Fusion |
|---|---|---|---|---|---|
| | L.ResNet | VGGFace | VGG19 | VGGFace | |
| playdoh | 16 (13.79%) | 8 (6.9%) | 57 (49.14%) | 49 (42.24%) | 58 (50%) |
| silly putty | 0 | 15 (27.27%) | 5 (9.09%) | 0 | 0 |
| total | 16 (4.89%) | 23 (7.03%) | 62 (18.96%) | 49 (14.98%) | 58 (17.74%) |

Table 3.25: Summary of APCEs at an $APCER_{0.2}$ for material group iv) partition.

Table 3.24, the one undetected attack presentation is made from glue, which is also the most difficult PAI group for the laser algorithms.

Finally, the results of material group iv) in Figure 3.57 show an inverted picture. The SWIR methods ($APCER_{0.2}$ between 15% and 23%) are outperformed by the laser algorithms. The laser LSTM ResNet (4.89%) achieves the best result, followed by the laser CNNs VGGFace (7.03%) and ResNet (11.62%). However, the laser LRCN reports an $APCER_{0.2}$ of nearly 19%, which is also reflected in the fusion (17.74%). As known from previous visual LOO experiments, Table 3.25 discloses that those high $APCER_{0.2}$ values are caused by playdoh presentations. Additionally, some silly putty samples are also not detected for specific algorithms.

All in all, the outcome of the material LOO experiments confirms findings from the visual LOO groups that are summarised in the following: *i)* especially transparent overlays are hard to detect when not seen during training, *ii)* except for playdoh, SWIR algorithms seem stronger than laser algorithms for fingerprint PAD, *iii)* the pre-defined fusion of complementary information channels generalises better than its independent components. Regardless of other properties, the most challenging PAI materials are dragonskin, playdoh, and silicone.

### 3.2.10 *Summary*

The camera change within the capture device results in higher resolution images and thus allows a focus on deep learning techniques for fingerprint PAD algorithms. In this context, a new dataset comprising 17,730 bona fide presentations and 4,339 attack presentations from 45 different PAI species has been collected in order to enable fingerprint PAD development. With a focus on the SWIR and laser sequence data, new PAD algorithms have been benchmarked on a unified baseline partition and additionally evaluated regarding their generalisation capabilities.

The tests of the SWIR CNNs showed that the 4-dimensional input processing module is very valuable for fingerprint PAD. When it is trained together with the other CNN blocks, the handcrafted RGB conversion is clearly outperformed in terms of PAD accuracy. Additional experiments on the laser data confirmed the soundness of the CNN approaches (fine-tuned and trained from scratch), which achieved in some cases remarkable results. However, the PAD performance still depends on the network's architecture and the utilised input data. While the MobileNets report good results on SWIR data, both are unsuited to process laser input. Furthermore, the structures of InceptionV3 and Xception seem too deep to successfully apply transfer learning on a small dataset. However, due to their specific architectural design removing some blocks is not as simple as for the MobileNets. Given the laser sequence data, additional LSTM and LRCN approaches are analysed and proved to have the potential to significantly enhance the fingerprint PAD performance. Finally, OC convolutional AEs have been proposed as a general method for anomaly detection. As these models are solely trained on bona fide presentations, they are designed to detect unknown attacks. The reported results demonstrate that the AE benefits from the 4-dimensional SWIR data and is less discriminative for the used laser data. While the SWIR AE achieves nearly perfect results for specific LOO groups, other two-class algorithms are too sensitive towards the training data (i. e., laser LRCN). Hence, these PAD methods do not generalise well, despite their superior performance on the baseline partition. Finally, it can be concluded that fusing laser and SWIR algorithms helps in terms of PAD performance as well as generalisability. While a convenient BPCER can be granted, dragonskin and orange playdoh fake fingers, and especially transparent overlays from two part silicone remain challenging to detect.

## 3.3 FINGERPRINT PAD SUMMARY

As the focus for this Thesis is on hardware-based solutions for fingerprint PAD, it could be shown that complementary information from multiple sensors are beneficial for the classification process. On the

other hand, software-based solutions are limited to one sensor and operate on legacy fingerprint images only. In this context, additional contributions [121] were made by benchmarking a set of fingerprint PAD algorithms on the publicly available LivDet datasets. In this work the Fisher Vector technique is utilised to combine the strengths of local and global features and achieve generalisability towards unknown attacks, cross-sensor captures, and cross-database scenarios. The experimental evaluation is done on the LivDet 2011 to LivDet 2017 datasets and includes seven different feature extractors: Scale Invariant Feature Transform (SIFT), Speed-Up Robust Features (SURF), HOG, LBP, BSIF, Binary Robust Independent Elementary Features (BRIEF), and Oriented FAST and Rotated BRIEF (ORB). The results show that the correct combination of handcrafted features is able to outperform state-of-the-art deep learning algorithms for software-based fingerprint PAD.

# 4 BIOMETRIC INFORMATION PROTECTION

The standard ISO/IEC 24745 on biometric information protection [144] defines three requirements for biometric systems:

- **irreversibility** "To prevent the use of biometric data for any purpose other than originally intended, biometric data shall be processed by irreversible transforms before storage." Given a protected template, it must be impossible to retrieve the original sample.
- **unlinkability** "The stored biometric references should not be linkable across applications or databases." Even with multiple protected templates stemming from the same instance of one data subject, they must not be linkable.
- **renewability** "A biometric reference may need to be changed for a variety of reasons besides compromise." Protected templates can be renewed or revoked without the need for re-enrolment of the data subject.

Furthermore, the biometric recognition performance should not decrease towards unprotected systems and usable transaction times are requested. In this context, the contributions in this Chapter[1] focus on BIP approaches in the encrypted domain based on Homomorphic Encryption (HE) and Garbled Circuits (GCs) in order to address research question RQ3. Both techniques allow privacy-preserving storage and comparison of biometric data without any accuracy loss since computations are directly executed on the ciphertexts. In addition, both concepts are independent of the biometric modality used as long as the respective comparison function is supported.

Additionally, the applied BIP systems utilise post-quantum cryptography [19] to provide long-term security. The latest estimation of the European Union expects efficient quantum computers being able to break current cryptosystems in 2035 [88]. On the other hand, biometric systems are deployed for a timespan of multiple years depending on the application context. While a retention period of three to five years is recommend for publicly operated systems [89, 90], European passports are usually valid for ten years, and some non-governmental access control systems have been running since twelve years [168]. As a consequence, contemporary research should take post-quantum resistant BIP approaches into account to guarantee privacy protection

1 This Chapter is based on our publications [16, 173, 174].

for the biometric templates. The persistence of biometric characteristics is the reason why biometric data needs enhanced protection compared to other authentication methods such as passwords. If protected biometric templates are leaked today, attackers can still reverse them using quantum computers in the future in order to retrieve valid representations. While passwords and tokens can be exchanged to counter previous security incidents, biometric systems pose the risk of impersonation once a utilised characteristic is leaked as the number of biometric instances per characteristic is limited (e. g., one face, two eyes, or ten fingers).

## 4.1 CRYPTOGRAPHIC METHODS FOR BIOMETRIC INFORMATION PROTECTION

Since BIP can be achieved in different ways [13, 37, 213, 238], this Section introduces the cryptographic methods that are relevant for the Thesis contribution such as the utilised HE schemes and Secure Two-Party Computation (STPC).

### 4.1.1 *Homomorphic Encryption*

The concept of Homomorphic Encryption (HE) generally allows computations on ciphertexts without decrypting the content[2]. In particular, HE schemes [1] are based on public-key cryptography (also asymmetric cryptography) with the property that purposeful mathematical operations that are applied on the ciphertext directly correspond to the equivalent operation on the plaintext. For the case of additive and multiplicative operations, the homomorphic properties are universally defined as:

$$Enc(A + B) = Enc(A) \diamond Enc(B) \tag{4.1}$$

$$Enc(A \cdot B) = Enc(A) \circ Enc(B) \tag{4.2}$$

Depending on the utilised HE scheme, the operations $\diamond$ and $\circ$ might vary. In general it holds that an operation $\diamond/\circ$ exists which can be applied to two ciphertexts and returns the encrypted sum/product of both corresponding plaintexts. Depending on the supported functionalities, HE schemes are grouped into the following three types:

Partially Homomorphic Encryption (PHE). These schemes are defined through their limitation to one homomorphic property: addition or multiplication. This might be the case for cryptosystems, where only one operation is mathematically possible.

Somewhat Homomorphic Encryption (SHE). In contrast to PHE, SHE schemes generally allow both operations. However, only a limited number of executions are possible before the result becomes undecryptable noise.

2 This Section is derived from our publication [174].

Fully Homomorphic Encryption (FHE). On the other hand, FHE schemes allow an unlimited number of additions and multiplications. On the contrary, the noise reduction techniques are computationally expensive, which results in longer execution times.

While the restrictions of PHE and SHE might limit the usage in other applications, the biometric comparison function utilises a fixed number and type of operations that are throughout consistent as long as template format and distance computation do not change. Hence, specific HE schemes can be selected according to the biometric use case. Given the additional requirement of post-quantum-security [19], the utilised crypto schemes are introduced in the subsequent Sections. For the whole Chapter, the following abbreviations are used within figures and equations:

- *Enc* - encryption
- *Dec* - decryption
- *sk* - secret key
- *pk* - public key
- *m* - message (*or here:* biometric data to be encrypted)

### 4.1.2 *CKKS Cryptosystem*

The cryptosystem[3] Cheon-Kim-Kim-Song (CKKS) [51] is defined as an approximate HE scheme. In other words, the decryption of a ciphertext is not definite but its precision depends on the system parameters:

$$\mathrm{Dec}_{sk}\left(Enc_{pk}\left(m\right)\right) \neq m \tag{4.3}$$

$$\mathrm{Dec}_{sk}\left(Enc_{pk}\left(m\right)\right) \approx m \tag{4.4}$$

However, with the suggested parameter sets [49], the bounds can be specified in a way that they do not interfere with the approximate decryption. In order to encrypt floating point values, these are separated in *significand* (also: *mantissa*) and *scaling factor* with a *base exponent* using approximate arithmetic. Hence, this representation defines a trade-off between efficiency and accuracy as some decimal places are rounded-off.

This conversion is part of the encoding step, which is illustrated in Figure 4.1. The resulting polynomial can then be encrypted to the ciphertext and in order to retrieve the plaintext, the decrypted polynomial needs to be decoded again. As long as two ciphertexts have an identical *scaling factor*, it is possible to add or multiply them. However, CKKS supports additional homomorphic operations such as rescaling, rotations, and complex conjugation, where rescaling

3 This Section is derived from our publication [174].

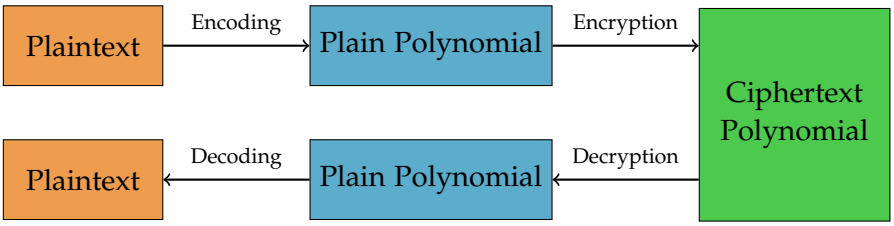



Figure 4.1: Overview of the components from the HE schemes.

is required after multiplications. In order to support an unlimited number of multiplications, the CKKS scheme was adjusted to become fully homomorphic [50]. The post-quantum security of the CKKS scheme is based on the shortest vector problem [193].

### 4.1.3 *BFV Cryptosystem*

The Brakerski/Fan-Vercauteren (BFV) [93] cryptosystem[4] is working on integer inputs and part of the community standard on homomorphic encryption [8]. As it is based on Brakerski's fully homomorphic scheme [31], the main contribution is a transmission from the Learning With Errors (LWE) [244] to the ring-LWE (RLWE) problem [193], which can be further reduced to the shortest vector problem by a quantum algorithm. In this regard, the authors port the fully HE from Brakerski to a simple somewhat HE scheme with RLWE. Subsequently, bootstrapping is added to reduce the noise and make the scheme fully homomorphic.

Bootstrapping was first introduced by Gentry [102] and is a method to lower the noise level before it exceeds the noise bounds, which would lead to undecryptable ciphertexts. However, as bootstrapping is expensive, its usage should be limited to its necessity. While multiplications in the encrypted domain significantly increase the noise level, additions can be executed nearly noise-free.

The general overview in Figure 4.1 also holds for BFV: the plaintext is mapped to a polynomial representation before encryption, which also needs to be decoded after the decryption. Finally, BFV is optimised to speed up the computation compared to Brakerski's original scheme and the authors further provide particular parameters to achieve a specific security level.

### 4.1.4 *NTRU Cryptosystem*

The lattice-based cryptosystem[5] N-th degree truncated polynomial ring (NTRU) [133] stands out since it is more efficient than other

4 This Section is derived from our publication [174].
5 This Section is derived from our publication [173].

public key cryptosystems such as RSA [246], Elgamal [82], or elliptic curves [204]. Indeed, its low memory requirements and computational efficiency [130] are comparable to symmetric cryptography, which allows for application scenarios with embedded or mobile devices. Hence, biometric authentication systems do not require high-end hardware to process encrypted data. Besides, NTRU currently is a finalist in the NIST post-quantum cryptography standardisation effort[6] [5, 6].

As a SHE scheme, NTRU generally supports both additive and multiplicative homomorphic operations, but a combination of those is not possible. In order to successfully decrypt an encrypted product, the exact number of applied multiplications is necessary, whereas encrypted sums can be decrypted by default. Thus, the utilised distance function is limited to either additions or multiplications. However, the system's parameters can be selected such that the decryption in $\mathbb{Z}_2 \,/\, \left(X^N - 1\right)$ automatically performs a modulo-2 operation. This has the advantage that one addition of two protected binary templates directly results in the XOR of both and the Hamming weight (HW) of this is equal to the Hamming distance (HD) between the templates.

The security of NTRU is based on the same RLWE problem [193] as the aforementioned cryptosystems, which defines the post-quantum resistance for lattice-based cryptography.

### 4.1.5 *Secure Two-Party Computation*

In general, Secure Two-Party Computation (STPC) is a concept[7] that is able to evaluate all mathematical functions that can be efficiently described, while preventing the parties to learn the opposite input. Hence, in the area of biometrics, the comparison of two templates is possible in a privacy-preserving way. In this context, biometric applications, independent of the modality, were used to benchmark several STPC approaches [22, 23, 36, 222, 251]. Although these solutions achieve efficient privacy-preserving comparisons, BIP is rarely granted as the templates themselves are classically available in plaintext.

Hence, secret sharing [164] is added to achieve secure database storage. This technique allows to share a biometric template with two non-colluding servers, such that one party on its own cannot learn any information from its share. Given a template $t$, the secret shares $\langle t \rangle_1$ and $\langle t \rangle_2$ can be computed such that reconstruction of $t$ is only possible when all shares are present. During enrolment, all references are shared between the corresponding servers $S_1$ and $S_2$. In the same way, the shares of the probe are computed and send to the servers. Subsequently, the STPC protocol receives in total four shares (two per template) to securely calculate the distance between probe and reference.

6 https://csrc.nist.gov/News/2020/pqc-third-round-candidate-announcement

7 This Section is derived from our publication [16].

The distance computation based on secret shares instead of full templates results in a computationally overhead compared to the classical STPC version, but fulfils the requirements for BIP. However, as STPC is able to utilise symmetric cryptography such as AES [62], it is usually more efficient than HE approaches. The applied STPC architecture follows the outsourced STPC design of Kamara and Raykova [160].

### 4.1.6 *Note on Randomness*

The proposed BIP methods within this Thesis only serve as proof-of-concept implementations[8]. While the utilised Pseudo Random Number Generators (PRNGs) are able to generate sufficient randomness for this purpose [153], hardware-based randomness is found to be cryptographically more secure. Hence, it is advised to exchange the PRNGs for real-world applications.

## 4.2 SYSTEM DESIGN FOR BIP

In addition to the desired post-quantum security, further design decisions[9] are derived based on knowledge gained from related work. Many publications in the area of BIP utilising HE require the client to store the secret key. However, as soon as the subjects needs a secret key in addition to a biometric characteristic, the advantage of using biometrics is lost and a de facto two-factor authentication system is in place. Regardless of the discussion whether two-factor authentication is relevant, the approaches in this Thesis focus on privacy protection of solely biometric systems without further dependencies on another factor. Therefore, a two server architecture is required as the secret decryption key can neither be stored next to the encrypted database nor at client side. Hence, the *honest but curious* model [127] is applied were the involved parties stick to the protocol but try to learn as much information as possible from the processed data. In order to grant full privacy protection, it is mandatory that the servers do not learn the plaintext representation of a template nor does the client learn something about the database. Therefore, these factors need to be taken into account when specifying the protocol.

Furthermore, some published architectures moved the distance computation to the client. This turned out a bad idea, as in this case the *honest but curious* model works only in theory. Since the client's only interest is to get successfully authenticated, it must be expected that the client deviates from the protocol and manipulates the computation. In fact, the client does not need to compute the distance but can simply encrypt a distance value that gets accepted by

8 This Section is derived from our publication [16].
9 This Section is derived from our publications [16, 173, 174].

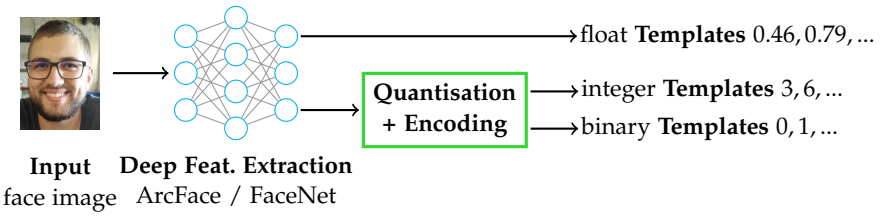



Figure 4.2: Pre-processing pipelines for face template extractions.

the decision threshold. Therefore, the proposed protocols require that the distance computation is done at server side.

Generally, the non-colluding two-server architecture can be considered a realistic setting, where one party offers privacy-preserving computation services. This independent provider assists in the authentication process and has an economic incentive to honestly follow the protocol as preserving the trust of the contracting authority is its business model.

Finally, it is possible to secure *honest but curious* models versus malicious adversaries, who deviate from the protocol, by the cost of additional computations. This was shown independent of the biometric modality in [15] using HE and in general for STPC methods [188]. However, as there are no published studies on post-quantum security, this Thesis continues in the *honest but curious* model.

## 4.3 BENCHMARKING POST-QUANTUM-SECURE HE SCHEMES

Although BIP based on HE is generally independent of the biometric modality used, the three HE schemes are benchmarked[10] in the context of a face verification application [151]. Given their different input requirements (float, integer, binary), the evaluation includes the biometric recognition performance, transaction time, template size, and cryptographic security. While CKKS and BFV were already evaluated in a biometric identification setting for face recognition [77], this work additionally considers the NTRU cryptosystem for a benchmark under equal conditions.

### 4.3.1 *Proposed Scheme*

The systems builds upon two deep feature extraction algorithms, ArcFace [73] and FaceNet [256], to process the facial input images. Both generate a feature vector of 512 floating point values by default as depicted in Figure 4.2. As these templates can only be processed by CKKS HE, subsequent quantisation and encoding provide additional conversions to integer and binary templates to allow usage of BFV and NTRU as well. The quantisation follows the approach of Drozdowski

10 This Section is based on our publication [174].

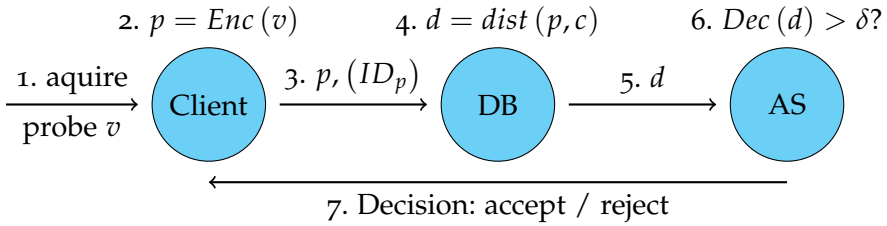



Figure 4.3: Homomorphically secured verification steps for the BIP system.

et al. [80], where the full feature space of the templates is equally split in four parts. Hence, the encoding process simply replaces the float values with the corresponding number of the respective quantisation part to create integer templates. Subsequently, the integers can be transformed to binary representations. However, it is important to maintain minimal distances between neighbouring areas and maximal distances for parts that do not share a common threshold. Hence, the Linearly Separable Subcode (LSSC) [186] is utilised to retrieve three binary digits from each integer.

For unprotected systems, the Euclidean distance is computed for float and integer templates. But since CKKS and BFV do not support calculating the square root in the encrypted domain, the squared Euclidean distance can be used in the same way, as the relation of the comparison scores is not affected by this. On the other hand, the HD is the most efficient method to compare binary plaintext templates. In this regard, the NTRU parameters can be set to automatically perform a modulo-2 operation during decryption of the protected distance. Hence, the XOR of probe and reference requires only one addition in the encrypted domain.

Taking into account the specified system design (Section 4.2), the steps for a biometric verification in the BIP system are illustrated in Figure 4.3. The architecture consists of client, database (DB) server, and authentication server (AS) and the particular transactions are defined as follows [174]:

1. The client captures the biometric characteristics and pre-processes the data, resulting in a probe feature vector $v$.

2. The client encrypts $v$ with the public key to get the protected probe $p$.

3. The encrypted probe $p$ is sent to the DB server. In a biometric verification scenario, the client additionally transfers an ID claim.

4. DB computes the distance $d$ between probe $p$ and reference(s) $c_i$ in the encrypted domain.

5. This encrypted distance $d$ is forwarded to the AS.

6. AS decrypts $d$ using the secret key and compares the result with a decision threshold. Alternatively, AS could also sort all computed distances in biometric identification mode.

7. The final accept/reject decision is revealed to the client.

Step 7 can also be split such that the decision is send to the DB, which then forwards it to the client, such that AS and client do not need to communicate directly. Due to the two-server architecture, this system operates in the *honest but curious* model, implying that parties (especially the servers) do not derive from the protocol to compromise the subject's privacy. In particular, DB and AS do not collude to decrypt complete templates. Summarising the process, the client sends it encrypted probe to the DB server, which computes the distance between two templates in the encrypted domain. Being unable to validate the protected distance, it is forwarded to the AS, which utilises the secret key for decryption in order to compare it to the decision threshold. Hence, AS receives no information about the corresponding templates and cannot collect sensitive data. In addition, transmission channels can be protected against external attackers by usage of TLS.

### 4.3.2 *Experimental Evaluation*

The face images used for the experiments are a subset of the FERET database [229] with frontal orientation towards the camera. This dataset consists of 6,963 samples from 563 subjects in a controlled environment. Reproducibility of the results is achieved by utilising publicly available models for feature extraction (ArcFace [73] and FaceNet [256]) in combination with open source crypto repositories. The C++ implementations of CKKS and BFV are part of the freely available Microsoft SEAL HE library[11] [257] and the NTRU parameters were adjusted in [173] and published as Python3 code[12].

Furthermore, a commodity notebook (Intel Core i7 2.7 GHz CPU and 16 GB DDR4 RAM) was used to measure the transaction times within a virtualised single-core Linux. While C++ is generally faster than native Python, the NTRU version is executed with Pypy3[13], which speeds up Python programs. Since the different template representations are expected to affect the biometric performance, this is measured in terms of biometric verification and rank-1 biometric identification scores. The genuine verifications include all mated comparison trials, while the impostor scores are computed between the first sample of each subject to the first sample of all other subjects. On the other hand, a closed set identification scenario is considered to validate the rank-1

11 https://github.com/Microsoft/SEAL
12 https://github.com/dasec/iris-he-ntru-btp
13 https://pypy.org

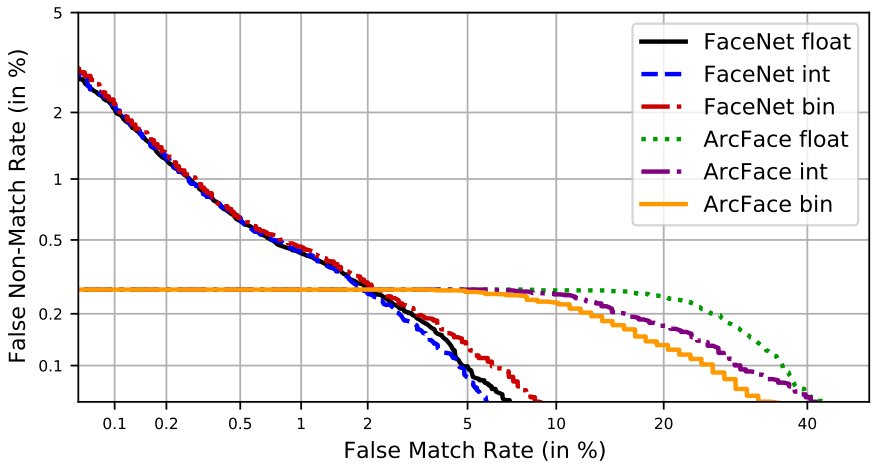



Figure 4.4: Verification performance of all template types in terms of FMR and FNMR. BIP systems are identically to unprotected systems.

| Rank-1 (%) | float | integer | binary |
|---|---|---|---|
| ArcFace | 99.03 | 98.98 | 99.03 |
| FaceNet | 98.50 | 98.42 | 98.36 |

Table 4.1: Identification performance of all template types in terms of rank-1 accuracy.

performance of all mated comparisons. As all HE schemes compute the same distances as unprotected systems, the biometric accuracy does not change in the BIP systems.

#### 4.3.2.1 *Biometric Performance Evaluation*

The DET curves of the biometric verification scenario are plotted in Figure 4.4. It can be observed that despite different template representations, the overall performance is preserved. In fact, the plot shows almost no variance in terms of False Non-Match Rate (FNMR) for application-relevant thresholds of False Match Rate (FMR)s below 2%. The only differences are for the two features extraction methods. On the other hand, Table 4.1 reveals minor divergence for the rank-1 biometric identification rates across the three template types. Hence, the impact of quantisation is measurable but generally negligible as the biometric performance stays consistent in all cases.

#### 4.3.2.2 *Transaction Time and File Sizes*

For all HE schemes the higher security levels result in increased transaction times and file sizes. However, only the outcomes for 128 bits of security are discussed in the following as the relative speed-up between the HE schemes remains the same.

| 128 bits Security | CKKS (float) | BFV (int) | NTRU (bin) |
|---|---|---|---|
| Key generation (ms) | 779 (±4) | 255 (±5) | 362 (±84) |
| Encryption (ms) | 6 (±2) | 76 (±1) | 27 (±5) |
| Comparison (ms) | 3,391 (±10) | 618 (±26) | 23 (±3) |

Table 4.2: Transaction times in terms of median and standard deviation. The comparison includes distance computation, decryption, and deriving the final decision.

| 128 bits Security | CKKS (float) | BFV (int) | NTRU (bin) |
|---|---|---|---|
| Keys | 99 MB | 12 MB | 6 KB |
| Template | 516 KB | 132 KB | 5.5 KB |

Table 4.3: Key and template sizes for the different HE schemes.

The timing results of relevant transactions are shown in Table 4.2. Generating the HE keys is a one time effort during system setup and is completed within one second for all cryptosystems. The encryption is executed once for each reference enrolment and again for each incoming probe in biometric verification and biometric identification applications. CKKS needs 6 ms, BFV 76 ms, and NTRU 27 ms to encrypt one template. The comparison consists of computing the distance between two protected templates, decrypting the result, and deriving the final decision. One comparison procedure with encrypted float features (3,391 ms) is five times slower than for integer templates (618 ms), which again is outranked by 25 times on binary representations (23 ms). As the comparison is measured between two templates it reflects the biometric verification and for biometric identifications the captured times are multiplied with the number of enrolled references. However, as the DB is already protected, the probe encryption affects the biometric identification times much less than a biometric verification. While CKKS needs significantly more time for the comparison than both other HE schemes, its encryption is the fastest. In the case of NTRU, each block encryption requires a new random polynomial and these generations seem to be much more efficient in CKKS. This is also reflected in the higher deviations within the key generation, where multiple random polynomials are created. Additionally, the NTRU keys might be discarded (and re-generated) if they do not comply with a particular structure. Moreover, also the disk usage varies as summarised on Table 4.3, where supporting floats requires more storage than integers and binaries. Hence, nearly 100 MB key material is generated for CKKS, 12 MB for BFV, and 6 KB for NTRU. Aligning with these sizes, CKKS templates require 516 KB, BFV 132 KB, and NTRU 5.5 KB.

Based on the observations of single instances, an emulated DB storage of 1,000 subjects requires around 500 MB in the CKKS scheme, 130 MB with BFV, and 6 MB with NTRU. Besides, timing a biometric identification on 1,000 enrolled references runs about one hour with CKKS encryption, close to twelve minutes for BFV, and 23 seconds for NTRU.

#### 4.3.2.3 *Security Analysis*

All three HE schemes achieve *irreversibility* with post-quantum security [19] for long term privacy protection. Moreover, a random factor during encryption ensures *unlinkable* ciphertexts even for identical plaintexts. Generally, *renewability* is possible in the specified BIP design by exchanging the key pair and re-encrypting the database. Re-enrolment is not necessary as the client uses only the public key.

### 4.3.3 *Summary*

This work confirms previous observations [77, 80] on face recognition that transforming float templates into integer and binary representations is possible without significant loss of biometric performance. On the other hand, this is the most important step to utilise more efficient BIP techniques. Furthermore, all three evaluated HE schemes (CKKS, BFV, and NTRU) comply with the ISO/IEC 24745 [144] requirements on *irreversibility*, *unlinkability*, and *renewability*. Additionally, the post-quantum security guarantees long-term privacy protection. Most importantly, it could be shown that biometric verifications in combination with BIP are executable in real time for integer and binary face templates. On the other hand, the used off-the-shelf hardware reaches it limitations for efficient biometric identifications. Here, only the NTRU system remains usable to some extend with reasonable transaction times.

## 4.4 EFFICIENT HOMOMORPHIC ENCRYPTION WITH WORKLOAD REDUCTION

In order to counter the computational overhead from HE, biometric systems can additionally make use of workload reduction techniques [79]. In this Section[14] an early decision strategy for the comparison of NTRU-protected iris-codes [69] is presented to speed up biometric verification and biometric identification scenarios.

In the context of iris recognition, Hollingsworth et al. [134] analysed the stability of the bits across multiple samples. As particular bit areas appear more fragile, a weighted comparison was proposed to improve the biometric recognition accuracy. With the aim of template size re-

14 This Section is based on our publication [173].

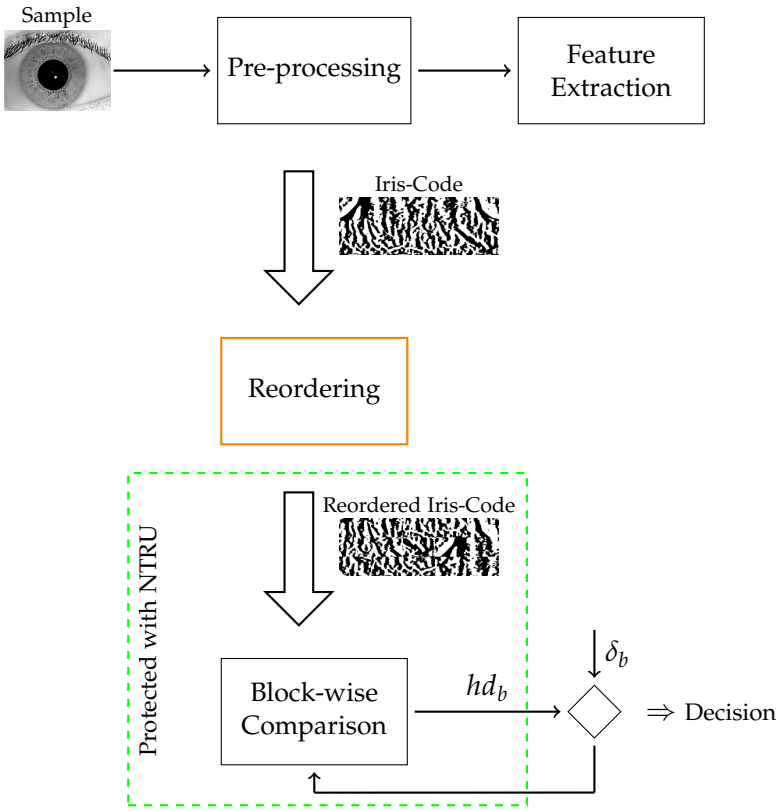



Figure 4.5: The ordinary iris-code is extracted from the pre-processed sample. The proposed modifications consists of reordering, encryption, and a block-wise comparison to accelerate the process.

duction, Gentile et al. [101] study the relevance of different parts of the iris-code. By removing irrelevant information, the shorter templates led to faster comparisons. Following these observations, Rathgeb et al. [239] sorted the most relevant bits to the beginning of the iris-code. Subsequently, an early rejection strategy was applied for block-wise comparisons in order to speed up biometric identification. Unlikely candidates were excluded from the following block comparisons, thus reducing the computational complexity.

This combination of rearrangement and early decision is adjusted in this work to improve the efficiency of biometric verification and biometric identification in the encrypted domain.

### 4.4.1 *Proposed Scheme*

The proposed system is illustrated in Figure 4.5. After the iris-code is extracted from the sample, the most significant areas are sorted to the front. The following block-by-block comparison allows early decisions based on intermediate HDs. The privacy is preserved by encrypting the templates directly after reordering and computing the block distances using the homomorphic properties of NTRU. An optimal sorting of the iris-code is expected to alleviate the accuracy

degradation of the early decision strategy. The efficiency of this enhanced BIP system is benchmarked against a regular comparison in the encrypted domain. Additionally, the accuracy degradation of the early decisions is observed. For this purpose, first an unprotected system is used to derive particular operation thresholds for the full system as well as block-based thresholds. All three versions support biometric verification and biometric identification modes and start with the original iris-code.

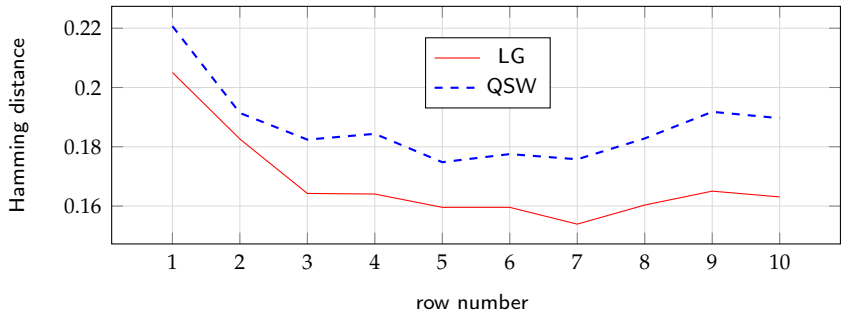
#### 4.4.1.1 *Feature Extraction*

The freely available *Iris Toolkit*[15][240] is used to process the samples and extract the iris-code. In particular, two different feature extraction algorithms are selected: Log-Gabor (LG) [202] and Quadratic Spline Wavelet (QSW) [194]. Both algorithms split the iris sample into ten circular rows, where the first one is closest to the pupil and the last one the outer ring near the sclera. Each row is represented by 512 bits, resulting in an iris-code of 5,120 bits. Finally, the probe iris-code is shifted left and right in order to compensate slight rotations during the capture process. By applying fixed $\pm 8$ circular bit shifts, a total of 17 probe variations are compared to the reference and only the minimum HD counts.

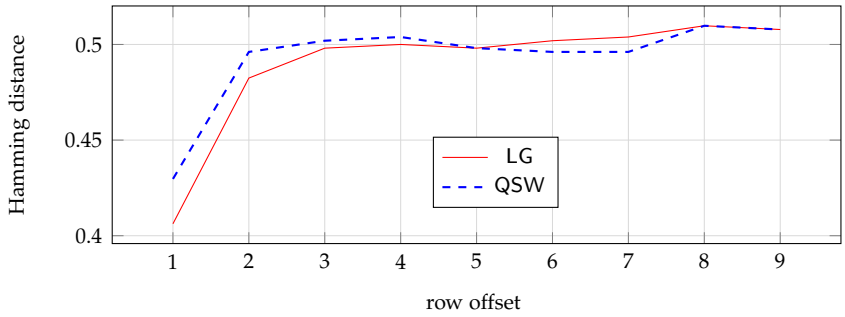
#### 4.4.1.2 *Bit Analysis and Reordering*

The motivation to force early decisions is the acceleration of biometric recognitions with HE. However, when the decision should remain accurate while comparing only a fraction of the template, it is of interest to compare the most discriminative information first. Hence, this Section combines own experiments with relevant finding in the literature in order to find a suitable iris-code arrangement for this purpose. When looking at the distinct rows of the iris-code, Hollingsworth et al. [134] observed that the pupil's dilation can affect the first rows. Besides, finding the delimitation between iris and sclera challenges the last rows. These findings can generally be confirmed for both feature extractors (Figure 4.6a), whereas the pupil appears much more troubling than the sclera. On the other hand, the middle rows indicate a higher stability and thus are more relevant in the new order. In addition, the next experiment looks at the correlation between rows of the same sample. The results in Figure 4.6b reveal that adjacent rows are closer, while distant ones have an average HD of 0.5. This means that information from neighbouring rows is not as discriminative and a row offset of two or more rows is recommended. In combination with the first experiment, the first blocks should be filled with the middle rows, the next blocks with the outer rows, and lastly the inner ones, while trying to split adjacent information.

15 USIT – University of Salzburg Iris Toolkit: http://www.wavelab.at/sources/




(a) Average Hamming distance per row across mated comparison trials.




(b) Correlation between different rows within the same sample.

Figure 4.6: Average Hamming distances to show stability (a) and correlation (b) of the rows in the iris-code.

However, the stability within rows varies as well as highlighted by Broussard et al. [38]. This is based on the fact that the eyelids are located above and below the eye and partly occlude those areas during some capture processes as visualised in Figure 4.7. Hence, the left and right sectors contain more consistent bits. These findings are not statistically analysed but simply confirmed by manually reviewing some samples of the utilised database. All in all, the absolute priority starts with the left and right sectors of the middle rows and finishes with the inner rows of the top and bottom sectors. The exact ordering is publicly available within the source code of this project[16].

#### 4.4.1.3 *Unprotected Comparison*

The steps for an unprotected biometric verification are depicted in Figure 4.8.

1. The client acquires the probe feature vector $v$.
2. The probe $v$ is sent to the server, for a biometric verification an ID claim is appended as well.
3. The server computes the HD between probe $v$ and reference $r^{ID}$.
4. The authentication request is either accepted or rejected.

16 https://github.com/dasec/iris-he-ntru-btp

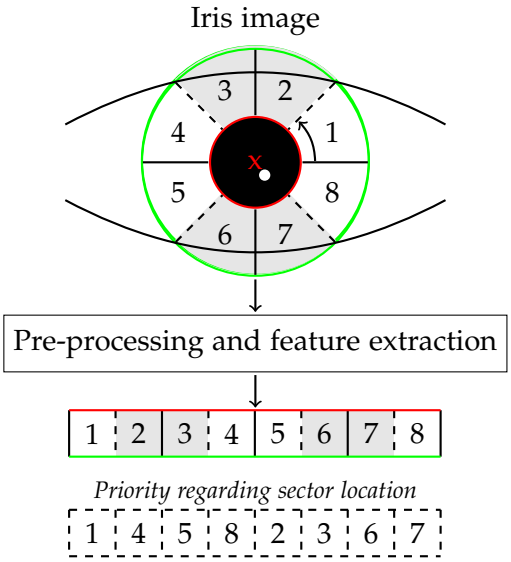



Figure 4.7: Sectors of the iris across rows. Eyelids may cover top and bottom sectors while left and right sectors remain mostly visible.

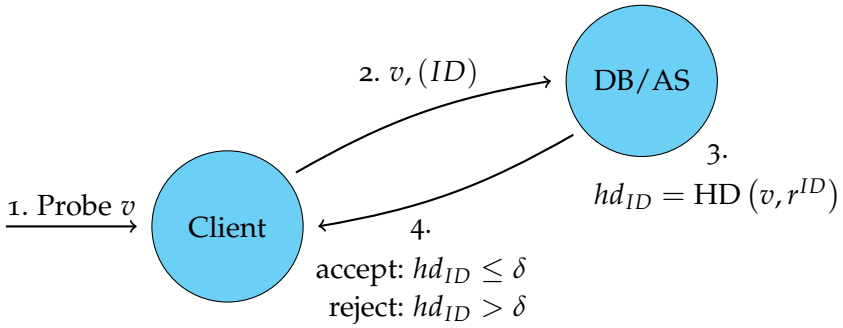



Figure 4.8: Verification steps for the unprotected system.

During biometric identification, the server computes the HDs from the probe to all enrolled references and returns the ID of the most similar candidate.

#### 4.4.1.4 *Proposed Improvement*

The proposed early decision strategy requires that the iris-codes of probe and reference are reordered prior to the comparison. In order to prevent unnecessary computations, it is suggested to store reordered references in the DB. In contrast to the baseline system, the server computes the HD of one block $b$ at a time and compares it to two thresholds $\delta_a$ and $\delta_r$, with $\delta_a < \delta_r$. The three possible outcomes can be defined as:

$$\text{Option} = \left\{ \begin{array}{ll} \text{accept}, & \text{for } hd_b \leq \delta_a \\ \text{reject}, & \text{for } hd_b > \delta_r \\ b = b + 1, & \text{for } \delta_a < hd_b \leq \delta_r \end{array} \right\} \tag{4.5}$$

Thus the biometric verification can be early accepted or rejected if the HD is outside of the decision boundaries or the next block $b+1$ is compared for cases where no early decision is possible. In the latter case, the thresholds $\delta_a$ and $\delta_r$ are adjusted to take the so far computed HD into account. In case the full iris-code is compared, the baseline threshold is applied to derive the final decision.

On the other hand, the optimised biometric identification computes the HDs of one probe block to the same block of all references. Those distances are then sorted and the most unlikely candidates are discarded. Hence, only a specified share $K$ of the most similar references is considered for the next block comparison. As a consequence, the number of future block comparisons decreases after each compared block, which is a significant speed-up in contrast to a naive biometric identification approach. As the efficiency gain is based on the remaining share $K$, the DB size $d$, the number of blocks $B$, and the applied circular shifts $s$, the particular number of block comparisons $Z$ can be calculated as follows:

$$Z = s \cdot \sum_{b=0}^{B-1} \lfloor d \ \cdot K^b \rfloor \tag{4.6}$$

#### 4.4.1.5 *Protected Comparison*

In contrast to the unprotected system, DB and AS are separated on two non-colluding servers since storing the secret decryption key next to the encrypted DB grants no protection for a leaked DB. The crucial step to use NTRU for BIP is to split the iris-code into blocks according to the specified maximum message length for encryption. The adjusted biometric verification steps of the baseline BIP system are depicted in Figure 4.9.

1. The probe feature vector $v$ is acquired.
2. The client encrypts the plaintext probe.
3. The encrypted probe $p$ and a potential ID claim are send to the DB.
4. Probe $p$ and reference $c$ are added in the encrypted domain.

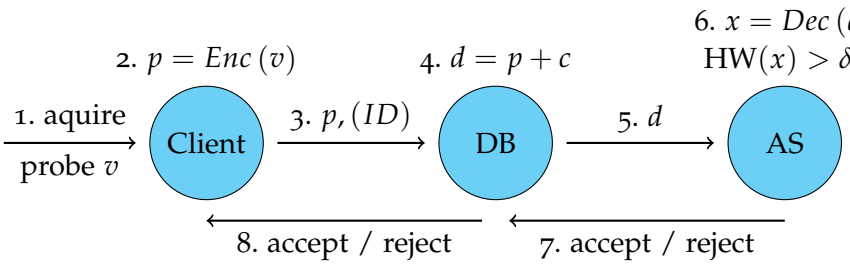



Figure 4.9: Verification steps for the baseline BIP system.

5. The encrypted sum $d$ is send to AS.
6. The decryption of the sum results in the XOR of probe and reference. The HW is then compared to a threshold $\delta$ to derive a decision.

7./8. The decision is forwarded to the client.

While this baseline version compares full iris-codes of 5,120 bits, the optimised system evaluates the HD after each block comparison based on the proposed early decision strategy as illustrated in Figure 4.10.

1. The probe feature vector $v$ is acquired.
2. The client reorders the bits of the probe iris-code and encrypts it.
3. The encrypted probe $p$ and a potential ID claim are send to the DB.
4. The DB adds one block $b$ of probe $p$ and reference $c$.
5. The encrypted sum $d_b$ is send to AS.
6. The decryption of the sum results in the XOR of the probe and reference block. The HW is then compared to two threshold $\delta_a$ and $\delta_r$ to see whether the authentication attempt can be accepted or rejected or whether the next block $b$ needs to be compared as well.
7. The decision is send to the DB, who either forwards it to the client or adds the next blocks.
8. The client receives the final decision.

The biometric identification works analogous to this as follows: the DB computes the sums of one probe block with the same block of all references. These sums are then processed by AS, which sorts the resulting HDs. The specified share $K$ decides how many references

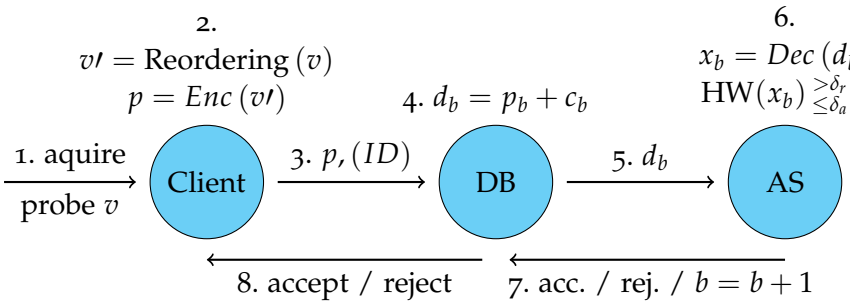



Figure 4.10: Verification steps for the optimised BIP system.

are kept for the next block comparison. The DB receives the information which references are most similar so far and proceeds with the addition of the next block. This interchange continues until all blocks are compared or only one reference is left.

### 4.4.2 *Experimental Evaluation*

The IITD iris database [182] with 2,240 samples from 224 subjects is selected to evaluate the proposed scheme. Each eye is captured five times and the NIR images have a resolution of $320 \times 240$ pixels. Since genetically related irises are as unique as unrelated ones [70], the dataset contains 448 different classes. The decision thresholds are derived from the samples of the first 24 subjects, the following 150 subjects are enrolled in the system, while the last 50 subjects are considered impostors who attack the system. This separation allows unbiased training and testing to evaluate the proposed efficiency enhancements.

The timing was tracked on a single-core Python3 application running on an Intel Xeon(R) 3.50GHz CPU with 16 GB RAM. Each function was executed 25 times with Pypy3[17] and the median transaction time is reported to exclude outliers. The NTRU parameters of the Python implementation[18] were adjusted to automatically perform a modulo-2 operation during decryption, thus enabling XOR behaviour for homomorphic additions. Finally, the full source code is published[19] to make this BIP scheme available. The implementation comprises four selectable security levels (112, 128, 196, or 256 bits) [132], which utilise increasing block sizes for higher security levels. Hence, dissimilar number of bits are compared within one block, less information in lower security levels and more information in higher security levels.

#### 4.4.2.1 *Accuracy Evaluation*

According to ISO/IEC 19795-1 [149], the biometric verification performance is presented in FMR and FNMR based on all mated comparisons of the enrolled subjects and in addition on all non-mated comparisons from the impostor set. Based on the thresholds that were fixed with help of the training set, the accuracy degradation of the applied early decision strategy can be measured. Additionally, the biometric identification accuracy is analysed in a closed-set scenario based on the rank-1 performance. The open-set scenario is not considered as HE generally maintains the baseline performance and the focus of this work is the evaluation of acceleration possibilities in the encrypted domain.

17 Pypy: https://pypy.org/features.html
18 NTRUEncrypt Python: https://github.com/logannc/pyNTRUEncrypt
19 NTRU Iris BIP Source Code: https://github.com/dasec/iris-he-ntru-btp

| | baseline system | NTRU security level (bits) 112 | 128 | 192 | 256 |
|---|---|---|---|---|---|
| **FMR** (%) | 0.11 | 0.52 | 0.45 | 0.36 | 0.39 |
| **FNMR** (%) | 1.60 | 1.90 | 1.87 | 2.10 | 1.83 |

Table 4.4: FMR and FNMR for the baseline (left) and block-optimised (right) verification scenario.

| $K$ | NTRU security level (bits) 112 | 128 | 192 | 256 |
|---|---|---|---|---|
| 100.00 % | 98.30 | 98.30 | 98.30 | 98.30 |
| 50.00 % | 97.93 | 97.95 | 98.08 | 98.08 |
| 25.00 % | 97.48 | 97.57 | 97.82 | 97.90 |
| 12.50 % | 97.25 | 97.33 | 97.55 | 97.58 |
| 6.25 % | 97.00 | 97.07 | 97.33 | 97.47 |

Table 4.5: Varying rank-1 identification rates (%) for selected shares $K$ that are kept after each block comparison.

The biometric verification results for the baseline and optimised versions are shown in Table 4.4. The decision thresholds were selected to grant FMR = FNMR = 1% in the training set, but the results deviates already for the baseline version. While the FMR turns out much lower (0.11%), the FNMR with 1.6% is slightly higher. As expected, the early decisions increase both error rates and show varying performance for different block sizes. Generally, higher security levels allow the comparison of more information within one block and thus are closer to the baseline accuracy.

The influence of different $K$ values on the rank-1 identification rate is summarised in Table 4.5. The baseline system ($K$ = 100%) scores the best result with 98.3% and discarding references through early rejection decreases the rank-1 identification rate to a minimum of 97% when keeping only the best 6.25% of references after each block comparison in a security level of 112 bits. The other results align between those two boundaries with better rank-1 scores for higher $K$ values and bigger block sizes. Hence, the biometric identification scenario keeps a stable accuracy while benefiting from a significant reduction of computational effort.

#### 4.4.2.2 *Computational Complexity*

As defined in Eq. (4.6) for biometric identification, the particular number of block comparisons depends on the database size $d$, the number of blocks per template $B$, the number of circular shifts $s$, and

| $K$ | NTRU security level (bits) | | | |
|---|---|---|---|---|
| | **112** | **128** | **192** | **256** |
| 100.00 % | 66,300 | 61,200 | 45,900 | 35,700 |
| 50.00 % | 10,132 | 10,132 | 10,132 | 10,081 |
| 25.00 % | 6,766 | 6,766 | 6,766 | 6,766 |
| 12.50 % | 5,797 | 5,797 | 5,797 | 5,797 |
| 6.25 % | 5,423 | 5,423 | 5,423 | 5,423 |

Table 4.6: Number of required block comparisons for the baseline system and the enhanced versions with security levels of {112, 128, 192, 256} bits given factor $K$, $d = 300$, $s = 17$, and $B = \{13, 12, 9, 7\}$.

the specified share $K$ of blocks that are kept for the next comparison. Since the block sizes, and thus the total number of blocks, depend on the NTRU security level, Table 4.6 presents the required block comparisons for the different version. Compared to the baseline system ($K = 100\%$), the early rejection strategy is able to significantly reduce the number of block comparisons and consequently accelerate the execution. In the next step, the transaction times of the main functions are measured and the medians from 25 runs are depicted in Table 4.7. It generally holds that utilising higher security levels requires more time for execution.

The *Key generation* takes up to one second and thus does not impact the setup phase of the BIP system. The template encryption is split into references and probes since the reference templates (300 in this case) are only encrypted once during deployment and for the probe all 17 shifts are encrypted before comparison. While the one-time effort with maximal 18 seconds is negligible, the probe encryption is required for every authentication attempt and hence affect the biometric verification much more than a biometric identification. Furthermore, with 0.5 to 1 second the probe encryption is far slower than a comparison in the encrypted domain, where the baseline verification is finished within 0.1 and 0.14 seconds. Since the transaction time of the probe encryption needs to be added, the early decision strategy does not really accelerate the biometric verification scenario.

On the other hand, the enhancements for biometric identification scenarios are much more impactful. While the baseline identification for 300 enrolled references takes between five and nine minutes based on the security level, it can be reduced to 25 and 83 seconds, respectively, when keeping only 6.25% of the most similar references after each block comparison. Hence, an efficiency gain of 92% or 85% can be observed for a loss of at most 1.3% rank-1 identification rate.

| system function | NTRU security level (bits) | | | |
|---|---|---|---|---|
| | **112** | **128** | **192** | **256** |
| **Key generation** | 0.302 | 0.359 | 0.662 | 1.012 |
| **Reference encryption** | 11.900 | 12.881 | 16.142 | 18.201 |
| **Probe encryption** | 0.549 | 0.602 | 0.794 | 0.952 |
| **Verification** | | | | |
| **Baseline** | 0.091 | 0.109 | 0.120 | 0.138 |
| **Enhanced** | 0.038 | 0.041 | 0.043 | 0.048 |
| **Identification** | | | | |
| $K$ = **100.00%** | 303.483 | 334.376 | 449.229 | 545.353 |
| $K$ = **50.00%** | 46.290 | 55.261 | 98.994 | 154.068 |
| $K$ = **6.25%** | 24.738 | 29.530 | 52.894 | 82.640 |

Table 4.7: Median transaction times in seconds for relevant functions of the BIP system in baseline and enhanced mode.

#### 4.4.2.3 *Security Analysis*

Since the identical NTRU cryptosystem is used as in the previous work, all three ISO/IEC 24745 [144] requirements (*irreversibility*, *unlinkability*, and *renewability*) are fulfilled. Furthermore, the block-based decision strategy does not pose an additional risk since the suggested block sizes of the HE scheme are used for this. Hence, BIP is fully granted.

### 4.4.3 *Summary*

This work evaluated an early decision strategy to accelerate the comparison of homomorphically encrypted iris templates in biometric verification and biometric identification scenarios. In order to support this, the iris-code was reordered to concentrate the most stable information in the beginning and mitigate the accuracy degradation. Given the different security levels, the trade-off between efficiency and security becomes visible since lower security levels are faster in execution. However, breaking 192 or 256 bits encryption security is already much harder than attacking the iris-code itself. According to Daugman [70] an iris-code code has around 250 bits of entropy and guessing approximately 70% correct is sufficient to force a match, which is less effort than breaking 192 or 256 bits encryption security. Therefore, those systems provide more security than the iris-code itself.

The loss of accuracy based on the early decision strategy is not of importance since the biometric verification is already fast enough in the baseline version ($\approx$ 1 second) and the biometric identification is

still too slow to operate in an authentication application and thus providing a list of $m$ candidates (rank-$m$) could further converge the performance in the direction of the baseline system. Additionally, it is possible to trade storage for efficiency. Since the probe encryption is the most time-consuming part during biometric verification, the $\pm 8$ circular shifts can be applied to the references (which requires more storage) in order to accelerate the probe encryption by a factor of 17.

## 4.5 POST-QUANTUM-SECURE TWO PARTY COMPUTATION FOR BIP

In contrast to previous STPC approaches with classical encryption [284], this work[20] presents the first application of post-quantum STPC in the area of BIP. Based on the recently published implementation of post-quantum STPC [41], iris-codes are stored and compared in the encrypted domain. The system's efficiency is then benchmarked to a classical STPC protocol based on [290] and the post-quantum secure NTRU HE implementation [173] described in Section 4.4. As for the previous BIP contributions, all utilised software is freely available to grant full reproducibility of the results.

### 4.5.1 *Proposed Scheme*

The proposed scheme is aligned to the NTRU-based iris BIP system in order to grant full comparability. Analogously, the same iris-codes [69] are used which were extracted with the LG algorithm [202] from the freely available *Iris Toolkit* [240]. The resulting 5,120 bits iris-codes are then secret shared across two servers, which are able to compute the distance in the encrypted domain following the specified STPC protocol. While references are kept like this, $\pm 8$ circular shifts are applied to the probes in order to compensate a possible tilted head during the capture process. Thus a total of 17 probe templates are later compared to the enrolled reference. Figure 4.11 illustrates the corresponding processing steps of the BIP system. In contrast to the HE BIP system (Section 4.4), always the full iris-code is used for comparison.

#### 4.5.1.1 *Secret Sharing*

Based on the security assumption that both servers do not collude, secret sharing guarantees BIP according to ISO/IEC IS 24745 [144], since it is impossible to retrieve the original template without access to all shares. The secret sharing of the enrolment is shown in Figure 4.12.

20 This Section is based on our publication [16].

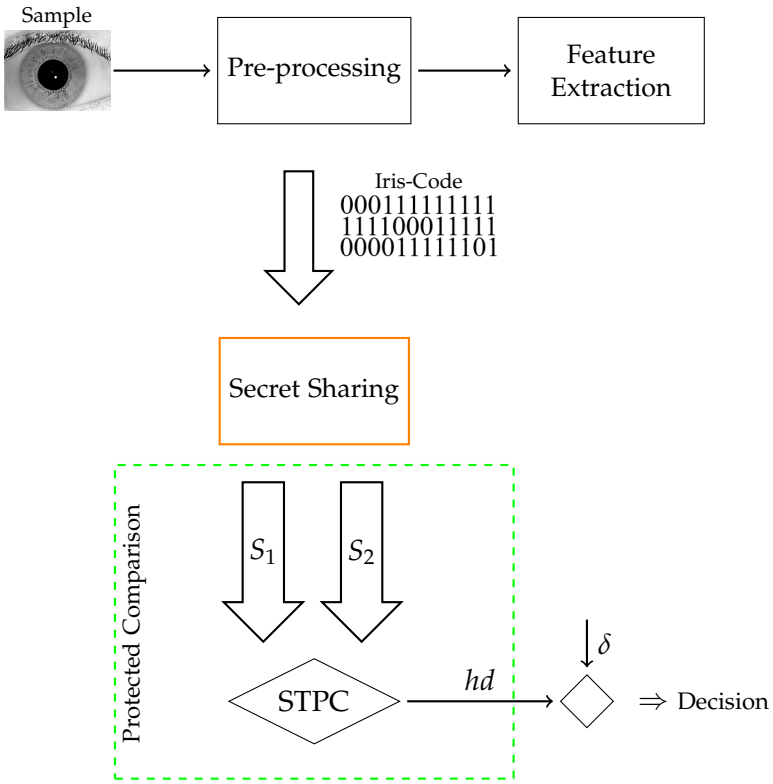



Figure 4.11: STPC system overview building upon the basic feature extraction steps. Template protection is granted through secret sharing and privacy-preserving comparisons are done using STPC.

1. Given the reference template $t_r$, the client generates a random 5,120 bit string $\langle t_r \rangle_1$ in order to compute the second share as:

$$\langle t_r \rangle_2 = t_r \oplus \langle t_r \rangle_1 \tag{4.7}$$

2. The shares $\langle t_r \rangle_1$ and $\langle t_r \rangle_2$ are send to the servers.

In fact, using a random value with the bit-wise Boolean XOR resembles the cryptographic one-time pad [17] at first sight. However, since the references are not exchanged after each comparison, the random value remains the same and is used multiple times throughout its life span. After all, this does not effect the security in the given setup as long as both server do not exchange their shares. Even a malicious server is unable to retrieve any information about the original template from its own share. Hence, post-quantum security is achieved due to the information-theoretic-secure secret sharing model.

#### 4.5.1.2 *Unprotected Comparison*

Given the secret sharing for template splitting, first the unprotected biometric verification process is described. Based on the probe iris-code $t_p$, the client creates its secret shares using a fresh random $\langle t_p \rangle_1$

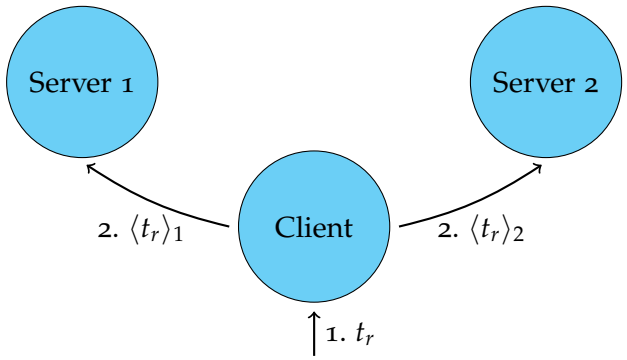



Figure 4.12: Secret sharing of the enrolment process.

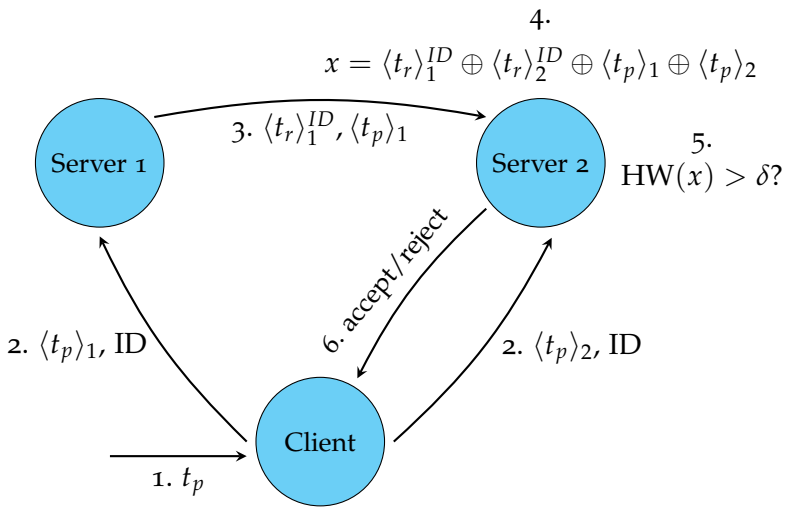



Figure 4.13: Unprotected comparison steps based on secret shared templates.

for each of the 17 circular shifted probe templates. In accordance to the enrolment, the corresponding second share $\langle t_p \rangle_2$ is computed from the XOR of original template and random value:

$$\langle t_p \rangle_2 = t_p \oplus \langle t_p \rangle_1 \tag{4.8}$$

A biometric verification scenario for the example of one probe template is illustrated in Figure 4.13. In a real application, all 17 shifts are compared to the reference and the lowest HD decides whether the authentication attempt is accepted or rejected.

1. The client computes the secret shares $\langle t_p \rangle_2$ and $\langle t_p \rangle_1$ from the original probe template.
2. The shares are distributed to the servers with the ID claim.
3. One server forwards its corresponding shares of probe and reference to the other party.

4. This party now holds all shares in order to compute the distance. First, the server XORs all shares, which resembles the XOR of the original probe and reference.

5. Then, the HW of the result can be compared to the decision threshold $\delta$.

6. The decision is send to the client.

However, during the distance computation (step 4) the original probe as well as the original reference template can be reconstructed from the server holding all shares:

$$t_r^{ID} = \langle t_r \rangle_1^{ID} \oplus \langle t_r \rangle_2^{ID} \qquad t_p = \langle t_p \rangle_1 \oplus \langle t_p \rangle_2 \tag{4.9}$$

This is the point where STPC protocols help to compute the distance without revealing the own input to the other party.

#### 4.5.1.3 *Protected Comparison*

The proposed scheme applies STPC on top of the secret sharing to guarantee secure storage and comparison for biometric templates. In this context, a baseline version with classical cryptographic security is implemented based on the EMP-Toolkit [290]. Additionally, the PQ-MPC implementation [41] is used to grant long-term post-quantum security. Following Yao's GCs protocol [296], the two servers are able to compute the HD between probe and reference in the encrypted domain. Hence, disclosing information to the other party can be prevented by utilising STPC.

The structure of classical and post-quantum STPC systems is identical since both versions operate on top of the secret shared database. Since the secret sharing is already post-quantum secure, this has the advantage that the classical STPC protocol can simply be exchanged by its post-quantum successor. This is especially interesting for transition periods since the EMP processing is less complex and thus faster. In contrast to permanent storages, live biometric verification processing is not yet endangered by future quantum computers [88]. However, attackers might still record the network traffic in order to break classical encryption in the future. Hence, from the security perspective, it is recommended to already use post-quantum cryptography wherever possible. As depicted in Figure 4.14, the verification steps are modified to achieve a privacy-preserving comparison of the template shares.

1. The client computes the secret shares $\langle t_p \rangle_2$ and $\langle t_p \rangle_1$ from the original probe template.

2. The shares are distributed to the servers with the ID claim.

3. Each server computes the XOR of its probe and corresponding reference shares.

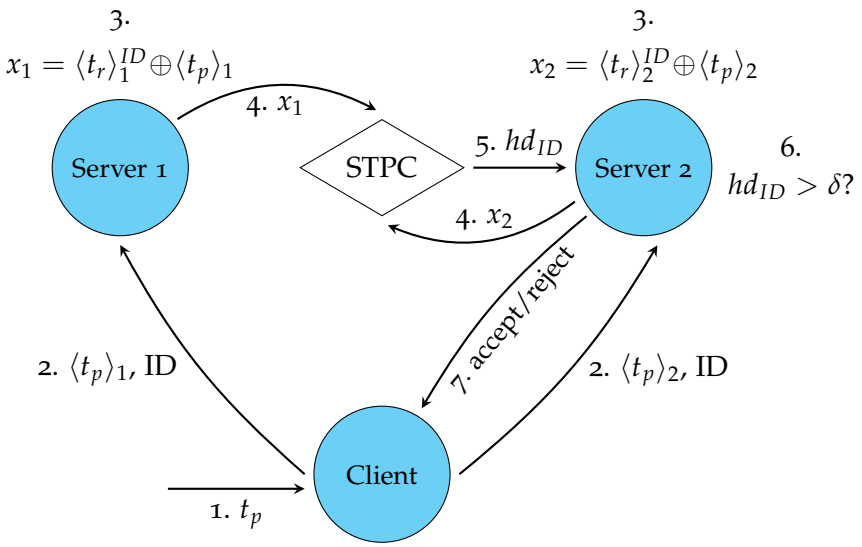



Figure 4.14: Protected comparison steps based on secret shared templates.

4. Those results serve then as input to the STPC protocol, which computes the HD of $x_1$ and $x_2$ in the encrypted domain.
5. The HD is send to one server.
6. This server compares the distance to the decision threshold $\delta$.
7. The final decision is disclosed to the client.

Additionally, it is also possible to perform the threshold comparison within the STPC protocol, which requires slight computational overhead.

### 4.5.2 *Experimental Evaluation*

The same IITD Iris Database [182] is used for the experiments as in the NTRU HE scheme (Section 4.4.2). The BIP system is implemented in C++ in order to use the freely available toolkits EMP [290] and PQ-MPC [41] for the STPC. Finally, measuring transaction times is done on a commodity notebook with 16 GB DDR4 RAM and an Intel Core 2.7 GHz CPU.

In order to compute the FMR and FNMR for biometric verification scenarios, all possible mated comparisons are calculated and additionally the first sample of each iris is compared to the first sample of all other instances. The resulting DET plot in Figure 4.15 proves that the biometric performance is not affected by the implemented STPC system since all versions work on the full iris-code.

#### 4.5.2.1 *Transaction Time*

The focus of this work is the runtime benchmark of the classical EMP Toolkit [290], the PQ-MPC library [41], and the post-quantum secure

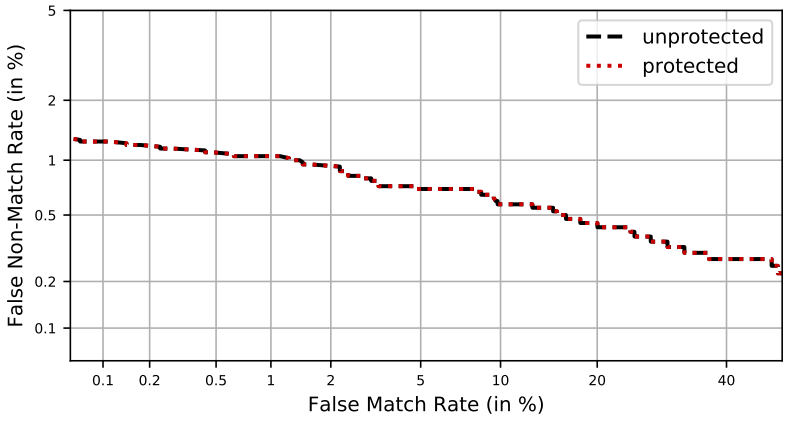



Figure 4.15: Using STPC, the verification performance of unprotected and protected systems are identical.

NTRU HE implementation [173]. Additionally, the three systems are evaluated using the same security level of 128 bits. For STPC based on GCs, this security parameter specifies the length of its internal labels. While classical 128-bit security is achieved for EMP, the post-quantum protocol requires a security parameter of 256 bits due to Grover's quantum algorithm [28]. The NTRU cryptosystem supports multiple security levels based on the selected system parameters [132].

The different system functions were executed 25 times and the median transaction times are presented in Table 4.8. The one-time efforts secret sharing and reference encryption are measured under the common name of *database setup*. As secret sharing is used for both STPC implementations, they require the same time for references and probe. In accordance to the NTRU BIP system, the *probe pre-processing* operates on all 17 circular shifts. Next, the *connection setup*, which initialises the STPC, is nearly ten times faster for the EMP version and also significantly slower than the actual *comparison*. Combining these results reveals that one biometric verification takes about 44 ms in EMP, half a second in PQ-MPC, and one second with NTRU. Since *pre-processing* and *connection setup* are executed only once for a biometric identification, this process requires 97 ms for EMP, nearly 20 seconds for PQ-MPC, and around 5.5 minutes with NTRU for a database with 300 enrolled references.

Focussing on the post-quantum secure implementations, the STPC version is twice as efficient as the real time biometric verification of NTRU HE. Furthermore, a biometric identification requires only 6% of the time when comparing the full iris-code of 5,120 bits, thus STPC becomes interesting for those scenarios as well. The even faster transaction times of EMP can be an attractive alternative during the transition into the quantum age.

| System function | EMP | PQ-MPC | NTRU (128) |
|---|---|---|---|
| **Database setup** | | 2731.48 | 13240.0 |
| **Probe pre-processing** | | 38.97 | 602.0 |
| **Connection setup** | 4.71 | 400.56 | – |
| **Comparison** | 0.18 | 63.38 | 109.0 |
| **Verification** | 43.86 | 502.90 | 711.0 |
| **Identification** (300 tmpl.) | 97.38 | 19,452.33 | 334,376.0 |

Table 4.8: Median transaction times in milliseconds for the classically and post-quantum protected system, compared to NTRU HE scheme.

#### 4.5.2.2 *Security Analysis*

The evaluated STPC systems are fully compliant to the ISO/IEC IS 24745 [144] requirements, in the case of PQ-MPC even against adversaries with access to quantum computers. The secret sharing guarantees *irreversibility* based on information-theoretic security using XOR and random values. Furthermore, the comparison in the encrypted domain through STPC prevents reconstructing the templates outside the secure storage. On the other hand, a fresh random factor during the secret sharing grants *unlinkability* of template stemming from the same instance. *Renewability* works in the same way by simply using a new random value for secret sharing. Consequently, the proposed system fulfils all BIP specifications in a post-quantum secure manner.

### 4.5.3 *Summary*

This work investigated BIP for iris-codes based on STPC and benchmarked the efficiency with the previously introduced NTRU HE system. The results show that while keeping the same post-quantum security level, the STPC approach requires only 50% of the execution time for biometric verifications and astonishing 6% for the same biometric identification. Since the classical STPC version builds upon the same post-quantum secure secret sharing, a fast alternative is available for live comparisons during the transition phase and can be seamlessly exchanged with its post-quantum secure successor. Finally, the concept of STPC is applicable for other biometric modalities as well. Furthermore in contrast to most HE schemes, STPC is also able perform the threshold comparison in the encrypted domain, which requires a slight overhead compared to the distance computation. Due to the usage of publicly available dataset and STPC libraries, the results are fully reproducible.

## 4.6 BIP SUMMARY

All in all, it can be concluded that BIP for biometric verification applications is executable in real time. Even the usage of post-quantum secure cryptography [19] for long term privacy protection does not create impractical computational overhead. On the other hand, specific workload reduction techniques [79] can be combined with BIP in order to accelerate biometric identification scenarios, but there is still room for improvement. However, in both cases it is possible to preserve the privacy in a way that leaked templates pose no risk to the subject nor the system. Due to the estimation of quantum computers being available in the near future [88], it is recommend to consider post-quantum secure BIP starting today.

Generally, the BIP methods based on HE and STPC are independent of the biometric modality. In this context, additional contributions are made in the area of voice biometrics. The approach in [212] utilises the Paillier HE scheme [224] in order to protect the voice templates as well as the vendor models that are required for speaker recognition. However, due to the limitations of this scheme, the distance computation is done at client side, which allows a malicious client to attack the process and get authenticated without being enrolled in the system. This issue is then fixed in [284] by replacing HE with STPC using the ABY framework [71]. Hence, this BIP approach preserves the same privacy level, while securing the system against malicious clients, and additionally comes with an enormous speed-up compared to the previous HE version. Conclusively, [213] provides an extensive overview of BIP for voice data. Starting with legislation and standardisation efforts, speaker recognition and cryptographic methods for BIP are explained, to finally present a technology survey of state-of-the-art privacy solutions for speech and speaker recognition.

# 5 CONCLUSIONS AND FUTURE WORK

This Thesis investigated methods for security enhancement and privacy protection for biometric systems in the context of fingerprint PAD and post-quantum secure BIP. In this Chapter, the own contributions are summarised, the research questions are answered, and finally topics for future work are discussed.

In the area of fingerprint PAD, a data collection with a new camera-based capture device was done in order to develop fingerprint PAD algorithms based on finger vein images, four selected SWIR wavelengths, and a laser sequence. The experimental evaluation showed that vein-based PAD methods are strong against full fake fingers, but cannot detect thin overlays that do not block the bona fide vein pattern behind. On the other hand, the LSCI technique reveals blood movement within the tissues beneath the skin. Hence, this liveness indicator is suited for PAD as long as the PAI is thick enough that the laser cannot penetrate it. In addition, the SWIR PAD methods operate on four frames of different wavelengths and thus show the best robustness. While all bona fide skin types reflect this illumination in a similar way, also orange playdoh is nearly indistinguishable. Therefore, multiple fusion schemes were evaluated to enhance the PAD performance based on complementary information channels.

Given the promising results especially for LSCI and SWIR algorithms, the capture device was updated to acquire higher resolution images in the SWIR spectrum. Based on a more extensive data collection, the PAD development focussed on deep learning approaches for laser and SWIR data. In this context, different CNN architectures were benchmarked and additional LSTM as well as LRCN methods used to process the temporal information. Finally, the best-performing PAD algorithms were further analysed regarding their generalisation capabilities for unknown attacks and directly set side by side with the developed one-class autoencoders. Overall, the results indicate that particular PAD algorithms perform well on the baseline partition that includes all captured PAI species, while a fusion of complementary data generalises better towards unknown scenarios.

In the area of BIP, the evaluated approaches enable biometric verifications in real time, while providing post-quantum security. On the other hand, the biometric identification can gain efficiency when combined with suited workload reduction techniques. However, it is important to maintain the cryptographic security while doing so. Generally, the proposed HE and STPC schemes are applicable independent of the biometric modality and fully preserve the biometric accuracy of

unprotected systems. Due to the possibility of attacking biometric systems now and break the protection mechanisms with future quantum computers, this Thesis suggests to implement post-quantum secure BIP wherever possible.

All in all, both areas of PAD and BIP are vital for biometric systems. Without security mechanisms, the bound between data subject and corresponding identity can be compromised. On the other hand, the trust in the system, and thus the will to use it, can only be achieved by privacy-preserving handling of the personal sensitive data. Based on these findings, the research questions can be answered.

## 5.1 RQ1: FINGERPRINT CAPTURE DEVICE

> *Which type of data needs to be captured for reliable fingerprint presentation attack detection?*

Due to the wide variety of (known) PAI species, it is impossible to name specific data types in order to answer this question. Generally it holds that the captured data needs to unite bona fide presentations to separate this class from attack presentations. Otherwise it is impossible to operate a convenient system and the new capture device is unsuited to replace legacy devices. On the other hand, the design of the capture device is to be taken into account as well. While e. g. capacitive sensors accept only conductive PAI species, camera-based devices are generally able to capture every visible material. In the context of this Thesis it was shown that it is beneficial to capture complementary information in order to enhance the PAD performance compared to a single data type. Hence, the challenge is to find different sensors that acquire a homogeneous bona fide group, which remains distinguishable from as many PAI species as possible. With the focus on the tested hardware-based approaches, the following sub-questions are taken into account as well.

> *What type of sensors are included in the capture device?*

The camera-based capture device (Section 3.1.1 and Section 3.2.1) combines two cameras with multiple distinct illuminations:

- Basler acA1300-60gm (400 nm - 1000 nm): sensitive for visible and NIR spectra
- Hamamatsu InGaAs / Xenics Bobcat 320 (1200 nm - 1700 nm): sensitive for the SWIR spectrum
- Visible (white) light to capture the fingerprint
- NIR (940 nm) back-illumination to capture the finger veins
- SWIR LEDs in 1200 nm, 1300 nm, 1450 nm, and 1550 nm for PAD

- 1310 nm laser illumination for temporal blood movement (PAD)

*Does the captured data require particular pre-processing?*

For the camera-based capture device, it is mandatory to extract the RoI from the image to focus on the finger data and ignore the periphery of the sample, which might be influenced by construction elements of the capture device. Furthermore, pre-processing is required in order to combine the information from multiple frames. Given the four captured SWIR wavelengths, it is possible to process them one by one or to link the information (e. g., spectral signature, 3D conversion). On the other hand, the captured laser sequence comprises only subtle changes within its single frames but can be processed in a way to reveal the blood movement over time. Finally, specific pre-processing is needed to extract the finger vein pattern from a finger vein photo. Hence, the data could be used without pre-processing but in order to obtain meaningful results, particular methods extract more relevant information.

*Is this system still compatible with legacy fingerprint sensors?*

Within this Thesis it was shown that legacy compatibility is generally possible based on the captured fingerphotos. As depicted in Figure 3.2, a commercial software (Neurotechnology Verifinger SDK) is able to extract the fingerprint and locate the minutiae points, which could then be successfully matched with a sample captured by an optical device. No further pre-processing was done for this test except for cropping the full capture to only include the upper phalanx. Additionally, the image had to be flipped (left to right) in order to be compliant to other touch-based capture devices with e. g. capacitive or optical sensors. However, a large scale experiment on compatibility is not part of this Thesis since the focus is on fingerprint PAD based on data acquired with different sensors. Furthermore, the setup is similar to other studies, which reported promising results for legacy compatibility of fingerphotos to touch-based fingerprints [187]. Additionally, the utilised capture devices provide a controlled environment with steady illumination and a fixed finger slot, which averts most challenges of collecting touchless fingerprints.

After answering the sub-questions, a response to research question RQ1 can be summarised as follows:

*The combination of SWIR and laser data allows the development of reliable fingerprint PAD methods. While bona fide presentations generally appear very similar, a large set of attack presentations can be successfully detected, which enables convenient and secure biometric systems.*

## 5.2 RQ2: PRESENTATION ATTACK DETECTION CLASSIFIERS

> *Which machine learning classifiers aid the detection of attack presentations while keeping the false alarm rate low?*

Similar to research question *RQ1*, the answer to this question is not universally valid for all fingerprint PAD approaches. Within this Thesis, a series of handcrafted classifiers (e. g., SVM) were evaluated. While those are definitely useful in terms of efficient training and can be trained to maintain a convenient BPCER (given suited features), their weakness lies on subtle differences between bona fide presentations and attack presentations. However, since those similarities are based on the particular utilised sensing technique, the same classifier might perform much better on different input data. On the other hand, one-class classifiers have the advantage that all PAI species are considered unknown attacks and thus the classifiers are expected to be more robust to future attack presentations, without the need to re-train the model. However, since these classifiers are trained on bona fide data only, PAI species that appear similar to bona fide presentations are most likely not detected. Hence, the performance highly depends on the captured data type and the specific PAI species. Generally, one-class PAD algorithms are always worth exploring as they can never be biased towards particular PAI species. Finally, based on the camera-based capture device, CNN-based PAD algorithms (CNN, LSTM, LRCN) perform well, whereas there are significant differences due to the CNN architectures. However, as they are designed to work on images, they are especially suited to process the different photos of the utilised capture device. All in all, different data requires different processing and classification techniques.

> *Does the combination of classifier and PAD data require further pre-processing of the data?*

Again this depends on the combination of data and classifier. The architecture of deep learning networks can be modified to directly process the specific input data. The 4D CNN is able to deal with all four SWIR wavelengths simultaneously and the LRCN can handle a full image sequence to connect temporal relations. On the other hand, handcrafted classifiers require a fixed 1-dimensional input vector and thus need further pre-processing in the case of a camera-based capture device. With the usage of different sensors (i. e., measuring single values on specific points) the pre-processing step is not necessarily required.

*In the context of this specific camera-based capture device, especially CNNs are a powerful tool to process and classify the captured photos. The advantage of using deep learning techniques for fingerprint PAD is the ability to learn subtle differences between bona fide presentations and particular attack presentations.*

## 5.3 RQ3: EFFICIENT BIOMETRIC INFORMATION PROTECTION

*Which concepts are suited for the protection of biometric systems while allowing real time efficiency?*

The evaluated BIP systems in this Thesis have shown that current post-quantum secure cryptography is fast enough to perform real time biometric verifications on commodity hardware. Additionally, the proposed methods are based on HE and STPC and thus generally independent of the biometric modality. However, the utilised cryptographic methods come with particular strengths such as processing float, integer, or binary inputs and hence are suited for different applications. On the other hand, biometric identifications are more complex and require significantly more time. In this context, workload reduction techniques [79] can be exploited to accelerate the process while maintaining BIP.

*All in all, long term privacy protection is possible using post-quantum secure cryptography. While biometric verifications in the encrypted domain achieve real time efficiency, further research is required to accelerate biometric identifications based on protected data.*

## 5.4 FUTURE WORK

Based on the contributions within this Thesis, future research lines arise. Given the two research areas of fingerprint PAD and BIP, the most relevant ones are summarised in this Section.

*Presentation Attack Detection*

- The most interesting part would be to have *large publicly available datasets for unified PAD benchmarks*. The LivDet team[1] organises biannual competitions and releases data collections, but they focus on software-based PAD methods and few specific PAI species with high relevance for the selected capture devices. However, it remains very challenging to benchmark different hardware-based PAD approaches, which could be solved by corporate data collection events. For example, the LivDet data collections could be extended in a way that academia and industry can bring their own capture devices. As a result, the same data (bona fide presentations and attack presentations) are captured with multiple devices, thus establishing a fair benchmark criterion. Additionally, these collected data could be published in accordance with GDPR. This would allow further competition and result in an overall better PAD performance as can be observed for software-based PAD methods since the start of the LivDet competitions.

1 https://livdet.diee.unica.it/ (previous competitions: https://livdet.org)

- Especially for fingerprint PAD, *once-class PAD algorithms* are of special interest due to the wide variety of known and unknown PAI species. In this context, further research could build upon the proposed methods or contribute completely different approaches in order to improve the detection accuracy. In the case of the proposed convolutional autoencoder, another loss function could aid to detect additional/other attack presentations.

- Finally, a more general remark; the fingerprint PAD evaluations in this Thesis highlight among others that some methods achieve remarkable results for a particular scenario but do not generalise well for unknown attacks. Hence, all future works definitely should investigate the generalisation capabilities towards unknown scenarios such as unknown attacks, cross database, and if possible cross capture device.

*Biometric Information Protection*

- With biometric verification scenarios achieving real time efficiency in the proposed BIP systems, the next step is to focus on *efficient solutions for biometric identification in the encrypted domain*. In this context, different cryptographic methods might come with less computational overhead. Additionally, further workload reduction methods can be exploited in combination with BIP. An interesting approach could be to reduce the template size (while maintaining the biometric accuracy) and batch multiple templates into one ciphertext in order to reduce the number of comparisons similar to [85].

- Another focus should be the evaluation of the proposed BIP methods given malicious servers, which might deviate from the protocol to obtain further information. However, when adjusting the specific solutions, the provided post-quantum security should not be affected. Due to the expected computational overhead, also the real time efficiency needs to be re-evaluated.

# a

# APPENDIX

The appendix mainly includes further Figures, which are additionally summarised in Table a.1.

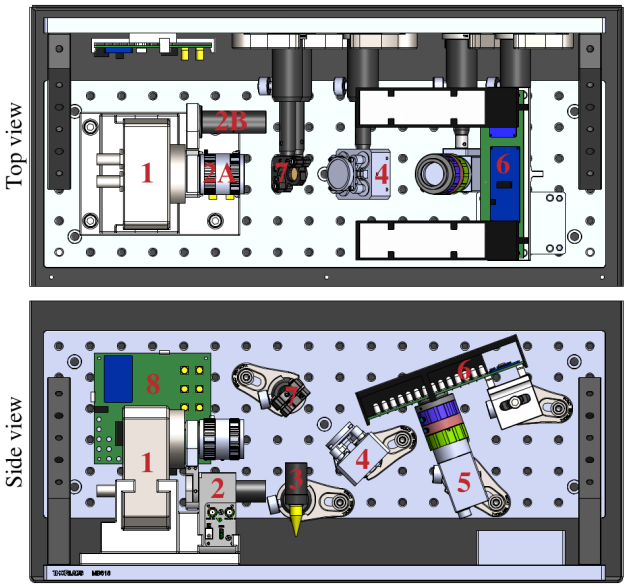
| Figure | Description |
|---|---|
| a.1 | Example materials and fingerprint PAIs |
| a.2 | Fingerprint capture device (version 1) |
| a.3 | Spectral remission intensities of skin and PAI materials |
| a.4 | Fingerprint capture device (version 2) |

Table a.1: Listing of figures in the appendix.

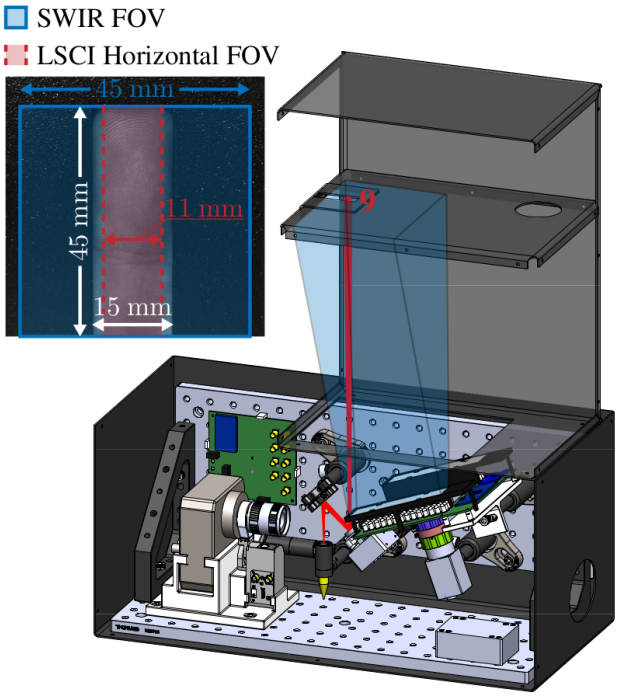

(a) A set of PAI materials.

(b) Some fingerprint PAIs.

Figure a.1: Example materials (a) to fabricate fingerprint PAIs (b).




(a) Lower compartment of the capture device.




(b) Field of view (FOV) and light paths.

Figure a.2: Fingerprint capture device (version 1) as illustrated and described in [137]. (a) Top and side view of the lower compartment of the capture device: (1) InGaAs area image sensor housed inside a detector head; (2) lens flipper holding a 25 mm SWIR (2A) and an LSCI (2B) lens; (3) laser module at 1310 nm mounted on beam collimator; (4) Fast Steering Mirror (FSM); (5) visible and NIR camera; (6) illumination board with LEDs controlled by FT232H and PCA9685; (7) mirror for laser; (8) controller board (for laser, FSM and lens flipper control). (b) Full capture device, with the cover removed, depicting the finger slot (9) and the light paths for multi-spectral light (blue) and laser light (red), as well as the FOV for SWIR and LSCI data.

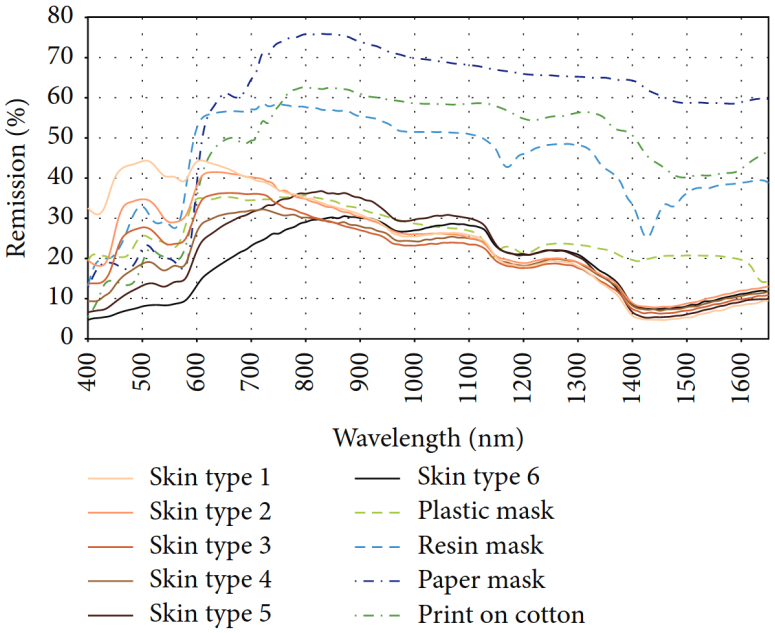



Figure a.3: Spectral remission intensities of skin and different PAI materials as illustrated in [271]. The SWIR spectrum is in the range of 1200 nm to 1700 nm, which is the area where different skin types align.

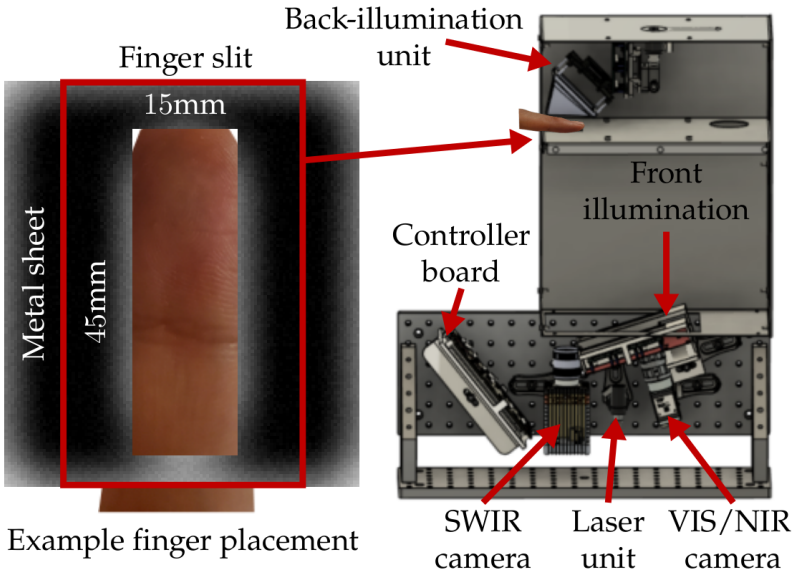



Figure a.4: Fingerprint capture device (version 2) as illustrated in [268]. The InGaAs camera was replaced and only one lens is used now. Hence, the lens flipper and mirror could also be removed. Additional technicals details are included in [267].

# GLOSSARY

| | |
|---|---|
| APCER | Attack Presentation Classification Error Rate. "proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario" [147]. xvii, 8, 35, 38, 40–46, 77, 82, 133, 134 |
| $APCER_{0.2}$ | APCER at a fixed operation point BPCER = 0.2%, i. e. 1/500 bona fide presentations is misclassified (convenient system). 8, 63, 68, 70, 72–74, 77, 78, 80–82, 84, 86, 88, 90 |
| attack presentation | "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system" [147]. An attack presentation to the capture device to either conceal the own identity or impersonate someone else. 4, 5, 7, 8, 11, 14, 15, 19, 26, 29, 30, 33–35, 38–40, 42, 45–48, 50, 56, 59–61, 63, 65, 73, 75, 77, 82, 86, 88, 90, 124–128, 133 |
| biometric identification | "process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual" [146]. 1, 2, 16, 18, 99, 101–105, 108, 109, 111–115, 120–123, 127, 128 |
| biometric verification | "process of confirming a biometric claim through biometric comparison" [146]. 1, 2, 16, 100–105, 107, 109, 111–123, 127, 128 |
| bona fide presentation | "interaction of the biometric capture subject and the biometric data capture subsystem in the fashion intended by the policy of the biometric system" [147]. A normal or genuine presentation. 4, 5, 7, 8, 11, 12, 14, 15, 26, 29, 33–35, 38–40, 45, 47–50, 56, 58–63, 65, 77, 82, 90, 91, 124–127, 133 |
| BPCER | Bona fide Presentation Classification Error Rate. "proportion of bona fide presentations incorrectly classified as attack presentations in a specific scenario" [147]. xvii, 8, 35, 36, 38–47, 71, 91, 126, 133, 134 |
| $BPCER_{0.2}$ | BPCER at a fixed operation point APCER = 0.2%, i. e. 1/500 attack presentations is misclassified (high security scenario). 8, 80 |

D-EER — Detection Equal Error Rate. PAD operation point where APCER = BPCER. xviii, 8, 43, 45, 46, 78, 80

enrolment — "act of creating and storing a biometric enrolment data record in accordance with an enrolment policy" [146]. 1, 97, 103, 115, 117

FMR — False Match Rate. "proportion of zero-effort impostor attempt samples falsely declared to match the compared non-self template" [149]. xviii, 111, 112, 119

FNMR — False Non-Match Rate. "proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user supplying the sample" [149]. xviii, 111, 112, 119

LSCI image — single image resulting from the LSCI preprocessing of the 1,000 frames laser speckle sequence as descriped in Section 3.1.2.1. 23, 34, 35, 49

PAD — Presentation Attack Detection."automated determination of a presentation attack" [145]. xix, 4–9, 11, 12, 14, 15, 19, 21–24, 26, 27, 29–31, 33–36, 38–43, 46–51, 54, 56, 58–63, 65–68, 70, 72–75, 77–82, 84, 86–88, 90, 91, 123–128, 134

PAI — Presentation Attack Instrument. "biometric characteristic or object used in a presentation attack" [145]. For instance, a printed face photo, a contact lens, or a silicone fingerprint overlay. xix, 7, 14, 19–21, 29–31, 38, 45, 47, 48, 65, 66, 71, 73, 77, 81, 82, 86–88, 90, 123

PAI species — "class of presentation attack instruments created using a common production method and based on different biometric characteristics" [147]. 8, 9, 11, 12, 14, 15, 19, 21, 26, 30, 31, 34, 35, 38, 42, 46, 47, 61, 63, 65, 66, 73, 74, 77, 86, 87, 90, 123, 124, 126–128, 133

probe — "biometric sample or biometric feature set input to an algorithm for use as the subject of biometric comparison to a biometric reference(s)" [146]. 1, 2, 17, 97, 98, 100, 101, 103, 105, 107–111, 113, 115–118, 120

reference — "one or more stored biometric samples, biometric templates or biometric models attributed to a biometric data subject and used as the object for biometric comparison" [146]. 1, 2, 17–19, 97, 98, 100, 103–105, 107–113, 115–118, 120

# BIBLIOGRAPHY


[1] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti. "A Survey on Homomorphic Encryption Schemes: Theory and Implementation." In: *ACM Computing Surveys (CSUR)* 51.4 (2018), pp. 1–35.

[2] A. Adler. "Sample Images Can Be Independently Restored From Face Recognition Templates." In: *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*. Vol. 2. IEEE. 2003, pp. 1163–1166.

[3] S. Agarwal, A. Rattani, and C. R. Chowdary. "AILearn: An Adaptive Incremental Learning Model for Spoof Fingerprint Detection." In: *arXiv preprint arXiv:2012.14639* (2020).

[4] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey. "Recent Advances in Homomorphic Encryption: A Possible Future for Signal Processing in the Encrypted Domain." In: *IEEE Signal Processing Magazine* 30.2 (2013), pp. 108–117.

[5] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, Y. Liu, C. Miller, D. Moody, R. Peralta, et al. *Status Report on the First Round of the NIST Post-Quantum Cryptography Standardization Process*. Tech. rep. National Institute of Standards and Technology (NIST), 2019.

[6] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, Y. Liu, C. Miller, D. Moody, R. Peralta, et al. *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process*. Tech. rep. National Institute of Standards and Technology (NIST), 2020.

[7] W. A. Alberto Torres, N. Bhattacharjee, and B. Srinivasan. "Privacy-preserving Biometrics Authentication Systems Using Fully Homomorphic Encryption." In: *Intl. Journal of Pervasive Computing and Communications* 11.2 (2015), pp. 151–168.

[8] M. R. Albrecht, M. Chase, H. Chen, J. Ding, S. Goldwasser, S. Gorbunov, S. Halevi, J. Hoffstein, K. Laine, K. E. Lauter, et al. *Homomorphic Encryption Standard*. Tech. rep. Toronto, Canada: HomomorphicEncryption.org, 2018.

[9] A. Alharbi, H. Zamzami, and E. Samkri. "Survey on Homomorphic Encryption and Address of New Trend." In: *Intl. Journal of Advanced Computer Science and Applications (IJACSA)* 11.7 (2020), pp. 618–626.

[10] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel. "Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments." In: *Proc. Intl. Conf. on Machine Learning (ICML)*. 2017.

[11] A. Anjos, L. El-Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. "Bob: A Free Signal Processing and Machine Learning Toolbox for Researchers." In: *Proc. ACM Intl. Conf. on Multimedia*. 2012, pp. 1449–1452.

[12] M. Barni, T. Bianchi, D. Catalano, M. Di Raimondo, R. D. Labati, P. Failla, D. Fiore, R. Lazzeretti, V. Piuri, A. Piva, et al. "A Privacy-compliant Fingerprint Recognition System Based on Homomorphic Encryption and Fingercode Templates." In: *IEEE Intl. Conf. on Biometrics: Theory Applications and Systems (BTAS)*. IEEE. 2010, pp. 1–7.

[13] M. Barni, G. Droandi, and R. Lazzeretti. "Privacy Protection in Biometric-Based Recognition Systems: A Marriage Between Cryptography and Signal Processing." In: *IEEE Signal Processing Magazine* 32.5 (2015), pp. 66–76.

[14] M. Barni, G. Droandi, R. Lazzeretti, and T. Pignata. "SEMBA: Secure Multi-Biometric Authentication." In: *IET Biometrics* 8.6 (2019), pp. 411–421.

[15] A. Bassit, F. Hahn, J. Peeters, T. Kevenaar, R. Veldhuis, and A. Peter. "Biometric Verification Secure Against Malicious Adversaries." In: *arXiv preprint arXiv:2101.10631* (2021).

[16] P. Bauspieß, J. Kolberg, D. Demmler, J. Krämer, and C. Busch. "Post-Quantum Secure Two-Party Computation for Iris Biometric Template Protection." In: *Proc. IEEE Workshop on Information Forensics and Security (WIFS)*. 2020, pp. 1–6.

[17] S. M. Bellovin. "Frank Miller: Inventor of the One-Time Pad." In: *Cryptologia* 35.3 (2011), pp. 203–222.

[18] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 2013.

[19] D. J. Bernstein and T. Lange. "Post-quantum Cryptography." In: *Nature* 549.7671 (2017), pp. 188–194.

[20] T. Bianchi, S. Turchi, A. Piva, R. D. Labati, V. Piuri, and F. Scotti. "Implementing Fingercode-based Identity Matching in the Encrypted Domain." In: *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*. IEEE. 2010, pp. 15–21.

[21] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. "Security Evaluation of Biometric Authentication Systems under Real Spoofing Attacks." In: *IET Biometrics* 1.1 (2012), pp. 11–24.

[22] M. Blanton and M. Aliasgari. "Secure Outsourced Computation of Iris Matching." In: *Journal of Computer Security* 20.2-3 (2012), pp. 259–305.

[23] M. Blanton and P. Gasti. "Secure and Efficient Protocols for Iris and Fingerprint Identification." In: *Proc. European Symposium on Research in Computer Security*. Springer, 2011, pp. 190–209.

[24] C. Blundo, E. De Cristofaro, and P. Gasti. "EsPRESSo: Efficient Privacy-Preserving Evaluation of Sample Set Similarity." In: *Data Privacy Management and Autonomous Spontaneous Security*. Springer, 2012, pp. 89–103.

[25] D. A. Boas and A. K. Dunn. "Laser Speckle Contrast Imaging in Biomedical Optics." In: *Journal of Biomedical Optics* 15.1 (2010).

[26] V. N. Boddeti. "Secure Face Matching Using Fully Homomorphic Encryption." In: *Proc. Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*. IEEE. 2018, pp. 1–10.

[27] D. Boneh, E. Goh, and K. Nissim. "Evaluating 2-DNF Formulas on Ciphertexts." In: *Proc. Theory of Cryptography Conf.* Springer. 2005, pp. 325–341.

[28] X. Bonnetain, M. Naya-Plasencia, and A. Schrottenloher. "Quantum Security Analysis of AES." In: *IACR Trans. on Symmetric Cryptology* 2019.2 (2019), pp. 55–93.

[29] L. Bottou. "Large-scale Machine Learning with Stochastic Gradient Descent." In: *Proc. Intl. Conf. on Computational Statistics (COMPSTAT)*. Springer, 2010, pp. 177–186.

[30] A. P. Bradley. "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." In: *Pattern Recognition* 30.7 (1997), pp. 1145–1159.

[31] Z. Brakerski. "Fully Homomorphic Encryption without Modulus Switching from Classical GapSVP." In: *Annual Cryptology Conf.* Springer. 2012, pp. 868–886.

[32] Z. Brakerski, C. Gentry, and V. Vaikuntanathan. "(Leveled) Fully Homomorphic Encryption without Bootstrapping." In: *ACM Trans. on Computation Theory (TOCT)* 6.3 (2014), pp. 1–36.

[33] Z. Brakerski and V. Vaikuntanathan. "Fully Homomorphic Encryption from Ring-LWE and Security for Key Dependent Messages." In: *Proc. Annual Cryptology Conf.* Springer. 2011, pp. 505–524.

[34] R. Breithaupt, C. Sousedik, and S. Meissner. "Full Fingerprint Scanner Using Optical Coherence Tomography." In: *Proc. Intl. Workshop on Biometrics and Forensics (IWBF)*. 2015, pp. 1–6.

[35] D. J. Briers. "Laser Speckle Contrast Imaging for Measuring Blood Flow." In: *Optica Applicata* 37.1-2 (2007), pp. 139–152.

[36] J. Bringer, H. Chabanne, M. Favre, A. Patey, T. Schneider, and M. Zohner. "GSHADE: Faster Privacy-Preserving Distance Computation and Biometric Identification." In: *Proc. ACM Workshop on Information Hiding and Multimedia Security*. 2014, pp. 187–198.

[37] J. Bringer, H. Chabanne, and A. Patey. "Privacy-preserving Biometric Identification Using Secure Multiparty Computation: An Overview and Recent Trends." In: *IEEE Signal Processing Magazine (SPM)* 30.2 (2013), pp. 42–52.

[38] R. P. Broussard, L. R. Kennell, and R. W. Ives. "Identifying Discriminatory Information Content within the Iris." In: *SPIE Defense and Security Symposium*. 2008, pp. 1–11.

[39] Bundesamt für Sicherheit in der Informationstechnik. *BSI Technical Guideline TR-03121-3 Biometrics for Public Sector Applications - Part 3: Application Profiles, Function Modules and Processes - Volume 1: Border Control*. Version 5.1. 2020.

[40] Bundesamt für Sicherheit in der Informationstechnik. *BSI Technical Guideline TR-03121-3 Biometrics for Public Sector Applications - Part 3: Application Profiles, Function Modules and Processes - Volume 2: Enrolment Scenarios for Identity Documents*. Version 5.1. 2020.

[41] N. Büscher, D. Demmler, N. P. Karvelas, S. Katzenbeisser, J. Krämer, D. Rathee, T. Schneider, and P. Struck. "Secure Two-Party Computation in a Quantum World." In: *Proc. Intl. Conf. on Applied Cryptography and Network Security (ACNS)*. Springer, 2020, pp. 461–480.

[42] V. Bushaev. "Understanding RMSprop - Faster Neural Network Learning." In: *Towards Data Science* (2018).

[43] P. Campisi. *Security and Privacy in Biometrics*. Vol. 24. Springer, 2013.

[44] K. Cao and A. K. Jain. "Learning Fingerprint Reconstruction: From Minutiae to Image." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 10.1 (2014), pp. 104–117.

[45] R. Cappelli, A. Lumini, D. Maio, and D. Maltoni. "Fingerprint Image Reconstruction From Standard Templates." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)* 29.9 (2007), pp. 1489–1503.

[46] A. Cavoukian and A. Stoianov. "Biometric Encryption: The New Breed of Untraceable Biometrics." In: *Biometrics: Theory, Methods, and Applications*. Ed. by N. V. Boulgouris, K. N. Plataniotis, and E. Micheli-Tzanakou. IEEE Press, 2010. Chap. 26, pp. 655–715.

[47] T. F. Chan, G. H. Golub, and R. J. LeVeque. "Updating Formulae and a Pairwise Algorithm for Computing Sample Variances." In: *Proc. Intl. Conf. on Computational Statistics (COMPSTAT)*. Springer. 1982, pp. 30–41.

[48] Y. Chen, X. S. Zhou, and T. S. Huang. "One-class SVM for Learning in Image Retrieval." In: *Proc. Intl. Conf. on Image Processing (ICIP)*. Vol. 1. IEEE. 2001, pp. 34–37.

[49] J. H. Cheon, K. Han, A. Kim, M. Kim, and Y. Song. "A Full RNS Variant of Approximate Homomorphic Encryption." In: *Proc. Intl. Conf. on Selected Areas in Cryptography*. Springer. 2018, pp. 347–368.

[50] J. H. Cheon, K. Han, A. Kim, M. Kim, and Y. Song. "Bootstrapping for Approximate Homomorphic Encryption." In: *Proc. Intl. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*. Springer. 2018, pp. 360–384.

[51] J. H. Cheon, A. Kim, M. Kim, and Y. Song. "Homomorphic Encryption for Arithmetic of Approximate Numbers." In: *Proc. Intl. Conf. on the Theory and Application of Cryptology and Information Security*. Springer. 2017, pp. 409–437.

[52] H. Choi, R. Kang, K. Choi, and J. Kim. "Aliveness Detection of Fingerprints Using Multiple Static Features." In: *Proc. of World Academy of Science, Engineering and Technology*. Vol. 22. 2007.

[53] F. Chollet. "Xception: Deep Learning with Depthwise Separable Convolutions." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 1251–1258.

[54] François Chollet et al. *Keras*. https://keras.io. 2015.

[55] T. Chugh, K. Cao, and A. Jain. "Fingerprint Spoof Buster: Use of Minutiae-Centered Patches." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 13.9 (2018), pp. 2190–2202.

[56] T. Chugh and A. Jain. "Fingerprint Spoof Detector Generalization." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 16 (2020), pp. 42–55.

[57] T. Chugh and A. K. Jain. "Fingerprint Presentation Attack Detection: Generalization and Efficiency." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2019, pp. 1–8.

[58] T. Chugh and A. K. Jain. "OCT Fingerprints: Resilience to Presentation Attacks." In: *arXiv preprint arXiv:1908.00102* (2019).

[59] J. S. Chung, A. Nagrani, and A. Zisserman. "VoxCeleb2: Deep Speaker Recognition." In: *Proc. Interspeech*. 2018, pp. 1086–1090.

[60] C. Cortes and V. Vapnik. "Support-vector Networks." In: *Machine Learning* 20.3 (1995), pp. 273–297.

[61] T. Cover and P. Hart. "Nearest Neighbor Pattern Classification." In: *IEEE Trans. on Information Theory* 13.1 (1967), pp. 21–27.

[62] J. Daemen and V. Rijmen. *The Design of Rijndael*. Vol. 2. Springer, 2002.

[63] N. Dalal and B. Triggs. "Histograms of Oriented Gradients for Human Detection." In: *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2005, pp. 886–893.

[64] I. Damgård, M. Geisler, and M. Krøigaard. "Efficient and Secure Comparison for On-Line Auctions." In: *Proc. Australasian Conf. on Information Security and Privacy*. Springer. 2007, pp. 416–430.

[65] I. Damgård, M. Geisler, and M. Krøigaard. "A Correction to 'Efficient and Secure Comparison for On-Line Auctions'." In: *Intl. Journal of Applied Cryptography* 1.4 (2009), pp. 323–324.

[66] I. Damgård, M. Keller, E. Larraia, V. Pastro, P. Scholl, and N. P. Smart. "Practical Covertly Secure MPC for Dishonest Majority - Or: Breaking the SPDZ Limits." In: *Proc. European Symposium on Research in Computer Security*. Springer. 2013, pp. 1–18.

[67] I. Damgård, V. Pastro, N. Smart, and S. Zakarias. "Multiparty Computation from Somewhat Homomorphic Encryption." In: *Proc. Annual Cryptology Conf.* Springer. 2012, pp. 643–662.

[68] L. N. Darlow, L. Webb, and N. Botha. "Automated Spoof-detection for Fingerprints Using Optical Coherence Tomography." In: *Applied Optics* 55.13 (2016), pp. 3387–3396.

[69] J. Daugman. "How Iris Recognition Works." In: *IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)* 14.1 (2004), pp. 21–30.

[70] J. Daugman and C. Downing. "Epigenetic Randomness, Complexity and Singularity of Human Iris Patterns." In: *Proc. of the Royal Society of London. Series B: Biological Sciences* 268.1477 (2001), pp. 1737–1740.

[71] D. Demmler, T. Schneider, and M. Zohner. "ABY - A Framework for Efficient Mixed-Protocol Secure Two-Party Computation." In: *Proc. Network and Distributed System Security Symposium (NDSS)*. 2015.

[72] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. "ImageNet: A Large-Scale Hierarchical Image Database." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2009, pp. 248–255.

[73] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 4690–4699.

[74] Y. Ding and A. Ross. "An Ensemble of One-Class SVMs for Fingerprint Spoof Detection Across Different Fabrication Materials." In: *Proc. Intl. Workshop on Information Forensics and Security (WIFS)*. IEEE. 2016, pp. 1–6.

[75] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. "Long-term Recurrent Convolutional Networks for Visual Recognition and Description." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 2625–2634.

[76] M. Drahansky, M. Dolezel, J. Michal, E. Brezinova, J. Yim, et al. "New Optical Methods for Liveness Detection on Fingers." In: *BioMed Research International* 2013 (2013), pp. 1–11.

[77] P. Drozdowski, N. Buchmann, C. Rathgeb, M. Margraf, and C. Busch. "On the Application of Homomorphic Encryption to Face Identification." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2019, pp. 1–8.

[78] P. Drozdowski, S. Garg, C. Rathgeb, M. Gomez-Barrero, D. Chang, and C. Busch. "Privacy-Preserving Indexing of IrisCodes with Cancelable Bloom Filter-based Search Structures." In: *Proc. European Signal Processing Conf. (EUSIPCO)*. IEEE, 2018, pp. 1–5.

[79] P. Drozdowski, C. Rathgeb, and C. Busch. "Computational Workload in Biometric Identification Systems: An Overview." In: *IET Biometrics* 8.6 (2019), pp. 351–368.

[80] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch. "Benchmarking Binarisation Schemes for Deep Face Templates." In: *Proc. Intl. Conf. on Image Processing (ICIP)*. IEEE, 2018, pp. 1–5.

[81] D. D. Duncan and S. J. Kirkpatrick. "Can Laser Speckle Flowmetry be made a Quantitative Tool?" In: *Journal of the Optical Society of America* 25.8 (2008), pp. 2088–2094.

[82] T. Elgamal. "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms." In: *Workshop on the Theory and Application of Cryptographic Techniques*. Springer. 1984, pp. 10–18.

[83] J. J. Engelsma, K. Cao, and A. K. Jain. "RaspiReader: Open Source Fingerprint Reader." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)* 41.10 (2018), pp. 2511–2524.

[84] J. J. Engelsma and A. K. Jain. "Generalizing Fingerprint Spoof Detector: Learning a One-class Classifier." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2019, pp. 1–8.

[85] J. J. Engelsma, A. K. Jain, and V. N. Boddeti. "HERS: Homomorphically Encrypted Representation Search." In: *arXiv preprint arXiv:2003.12197* (2020).

[86] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft. "Privacy-Preserving Face Recognition." In: *Proc. Intl. Symposium on Privacy Enhancing Technologies Symposium (PETS)*. Springer. 2009, pp. 235–253.

[87] M. Espinoza and C. Champod. "Using the Number of Pores on Fingerprint Images to Detect Spoofing Attacks." In: *Proc. Intl. Conf. on Hand-Based Biometrics*. IEEE. 2011, pp. 1–5.

[88] EU Parliament. *EU Quantum Manifesto: A New Era of Technology*. 2016.

[89] European Commission. URL: https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system_en.

[90] European Data Protection Supervisor. *Report on Logging to the SIS II at National Level*. 2018.

[91] European Parliament. *EU Regulation 2016/679 (General Data Protection Regulation)*. 2016.

[92] B. Fan, Z. Wang, and F. Wu. *Local Image Descriptor: Modern Approaches*. Springer, 2015.

[93] J. Fan and F. Vercauteren. "Somewhat Practical Fully Homomorphic Encryption." In: *IACR Cryptology ePrint Archive* 2012 (2012), p. 144.

[94] J. Feng and A.K. Jain. "FM Model Based Fingerprint Reconstruction from Minutiae Template." In: *Proc. Intl. Conf. on Biometrics (ICB)*. Vol. 5558. Lecture Notes in Computer Science. Springer, 2009, pp. 544–553.

[95] J. Fierrez, A. Morales, and J. Ortega-Garcia. "Biometrics Security." In: *Encyclopedia of Cryptography, Security and Privacy*. Ed. by S. Jajodia, P. Samarati, and M. Yung. Springer Berlin Heidelberg, 2019, pp. 1–3.

[96] T. B. Fitzpatrick. "The Validity and Practicality of Sun-Reactive Skin Types I Through VI." In: *Archives of Dermatology* 124.6 (1988), pp. 869–871.

[97] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia. "A High Performance Fingerprint Liveness Detection Method based on Quality Related Features." In: *Future Generation Computer Systems* 28.1 (2012), pp. 311–321.

[98] J. Galbally, S. Marcel, and J. Fierrez. "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition." In: *IEEE Trans. on Image Processing* 23.2 (2014), pp. 710–724.

[99] J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia. "On the Vulnerability of Face Verification Systems to Hill-Climbing Attacks." In: *Pattern Recognition* 43.3 (2010), pp. 1027–1038.

[100] J. Galbally, A. Ross, M. Gomez-Barrero, J. Fierrez, and J. Ortega-Garcia. "Iris Image Reconstruction From Binary Templates: An Efficient Probabilistic Approach Based on Genetic Algorithms." In: *Computer Vision and Image Understanding* 117.10 (2013), pp. 1512–1525.

[101] J. E. Gentile, N. Ratha, and J. Connell. "SLIC: Short-Length Iris Codes." In: *Proc. Intl. Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*. IEEE. 2009, pp. 1–5.

[102] C. Gentry. "A Fully Homomorphic Encryption Scheme." PhD thesis. Standford University, 2009.

[103] C. Gentry and S. Halevi. "Implementing Gentry's Fully-Homomorphic Encryption Scheme." In: *Proc. Intl. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*. Springer. 2011, pp. 129–148.

[104] L. Ghiani, D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli, and S. A. Schuckers. "Review of the Fingerprint Liveness Detection (LivDet) Competition Series: 2009 to 2015." In: *Image and Vision Computing* 58 (2017), pp. 110–128.

[105] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli, and S. A. Schuckers. "LivDet 2013 Fingerprint Liveness Detection Competition 2013." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2013, pp. 1–6.

[106] I. Goicoechea-Telleria. "Evaluation of Presentation Attack Detection under the Context of Common Criteria." PhD thesis. Universidad Carlos III de Madrid, 2019.

[107] M. Gomez-Barrero, J. Fierrez, J. Galbally, E. Maiorana, and P. Campisi. "Implementation of Fixed-Length Template Protection based on Homomorphic Encryption with Application to Signature Biometrics." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 191–198.

[108] M. Gomez-Barrero, J. Galbally, A. Morales, M. A. Ferrer, J. Fierrez, and J. Ortega-Garcia. "A Novel Hand Reconstruction Approach and Its Application to Vulnerability Assessment." In: *Information Sciences* 268 (2014), pp. 103–121.

[109] M. Gomez-Barrero, J. Galbally, A. Morales, and J. Fierrez. "Privacy-Preserving Comparison of Variable-Length Data with Application to Biometric Template Protection." In: *IEEE Access* 5.1 (2017), pp. 8606–8619.

[110] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Towards Fingerprint Presentation Attack Detection Based on Short Wave Infrared Imaging and Spectral Signatures." In: *Proc. Norwegian Information Security Conf. (NISK)*. 2018.

[111] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Towards Multi-Modal Finger Presentation Attack Detection." In: *Proc. Intl. Conf. on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2018, pp. 547–552.

[112] M. Gomez-Barrero, J. Kolberg, and C. Busch. "Multi-Modal Fingerprint Presentation Attack Detection: Looking at the Surface and the Inside." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2019, pp. 1–8.

[113] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, and J. Fierrez. "Multi-Biometric Template Protection Based on Homomorphic Encryption." In: *Pattern Recognition* 67 (2017), pp. 149–163.

[114] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez. "Unlinkable and Irreversible Biometric Template Protection based on Bloom Filters." In: *Information Sciences* 370–371 (2016), pp. 18–32.

[115] M. Gomez-Barrero, C. Rathgeb, J. Galbally, J. Fierrez, and C. Busch. "Protected Facial Biometric Templates Based on Local Gabor Patterns and Adaptive Bloom Filters." In: *2014 22nd Intl. Conf. on Pattern Recognition (ICPR)*. 2014, pp. 4483–4488.

[116] M. Gomez-Barrero, C. Rathgeb, G. Li, R. Raghavendra, J. Galbally, and C. Busch. "Multi-Biometric Template Protection based on Bloom Filters." In: *Information Fusion* 42 (2018), pp. 37–50.

[117] M. Gomez-Barrero, R. Tolosana, J. Kolberg, and C. Busch. "Multi-Spectral Short Wave Infrared Sensors and Convolutional Neural Networks for Biometric Presentation Attack Detection." In: *AI and Deep Learning in Biometric Security: Trends, Potential and Challenges*. CRC Press, 2021, pp. 105–132.

[118] L. J. González-Soler, L. Chang, J. Hernández-Palancar, A. Pérez-Suárez, and M. Gomez-Barrero. "Fingerprint Presentation Attack Detection Method Based on a Bag-of-Words Approach." In: *Proc. Iberoamerican Congress on Pattern Recognition*. Springer. 2017, pp. 263–271.

[119] L. J. González-Soler, M. Gomez-Barrero, L. Chang, A. Perez-Suarez, and C. Busch. "On the Impact of Different Fabrication Materials on Fingerprint Presentation Attack Detection." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2019.

[120] L. J. González-Soler, M. Gomez-Barrero, L. Chang, A. Perez-Suarez, and C. Busch. "Fingerprint Presentation Attack Detection based on Local Features Encoding for Unknown Attacks." In: *IEEE Access* 9 (2021), pp. 5806–5820.

[121] L. J. González-Soler, M. Gomez-Barrero, J. Kolberg, L. Chang, A. Pérez-Suárez, and C. Busch. "Local Feature Encoding for Unknown Presentation Attack Detection: An Analysis of Different Local Feature Descriptors." In: *IET Biometrics* (2021), pp. 1–18.

[122] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep Learning*. Vol. 1. 2. MIT Press, 2016.

[123] J. W. Goodman. "Statistical Properties of Laser Speckle Patterns." In: *Laser Speckle and Related Phenomena*. Vol. 9. Springer, 1975, pp. 9–75.

[124] S. A. Grosz, T. Chugh, and A. K. Jain. "Fingerprint Presentation Attack Detection: A Sensor and Material Agnostic Approach." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. 2020.

[125] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. "Biometrics Systems Under Spoofing Attack: An Evaluation Methodology and Lessons Learned." In: *IEEE Signal Processing Magazine* 32.5 (2015), pp. 20–30.

[126] Y. Han, C. Ryu, J. Moon, H. Kim, and H. Choi. "A Study on Evaluating the Uniqueness of Fingerprints Using Statistical Analysis." In: *Proc. Intl. Conf. on Information Security and Cryptology*. Springer. 2004, pp. 467–477.

[127] C. Hazay and Y. Lindell. "Semi-honest Adversaries." In: *Efficient Secure Two-Party Protocols: Techniques and Constructions*. Springer Berlin Heidelberg, 2010, pp. 53–80.

[128] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[129] C. Hengfoss, A. Kulcke, G. Mull, C. Edler, K. Püschel, and E. Jopp. "Dynamic Liveness and Forgeries Detection of the Finger Surface on the Basis of Spectroscopy in the 400–1650nm Region." In: *Forensic Science International* 212.1 (2011), pp. 61–68.

[130] J. Hermans, F. Vercauteren, and B. Preneel. "Speed Records for NTRU." In: *Proc. Conf. Cryptographers' Track at the RSA*. Springer. 2010, pp. 73–88.

[131] T. K. Ho. "Random Decision Forests." In: *Proc. Intl. Conf. on Document Analysis and Recognition*. Vol. 1. IEEE. 1995, pp. 278–282.

[132] J. Hoffstein, J. Pipher, J .M Schanck, J. H. Silverman, W. Whyte, and Z. Zhang. "Choosing Parameters for NTRUEncrypt." In: *Cryptographers' Track at the RSA Conference*. Springer. 2017, pp. 3–18.

[133] J. Hoffstein, J. Pipher, and J. H. Silverman. "NTRU: A Ring-Based Public Key Cryptosystem." In: *Proc. Intl. Algorithmic Number Theory Symposium*. Springer. 1998, pp. 267–288.

[134] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn. "The Best Bits in an Iris Code." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 31.6 (2009), pp. 964–973.

[135] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." In: *arXiv preprint arXiv:1704.04861* (2017).

[136] C. W. Hsu, C. C. Chang, and C. J. Lin. *A Practical Guide to Support Vector Classification*. 2003.

[137] M. Hussein, L. Spinoulas, F. Xiong, and W. Abd-Almageed. "Fingerprint Presentation Attack Detection Using A Novel Multi-Spectral Capture Device and Patch-Based Convolutional Neural Networks." In: *Proc. IEEE Workshop on Information Forensics and Security (WIFS)*. 2018, pp. 1–8.

[138] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo. "A Survey in Presentation Attack and Presentation Attack Detection." In: *Proc. Intl. Carnahan Conference on Security Technology (ICCST)*. IEEE. 2019, pp. 1–13.

[139] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo. "Dynamic Fingerprint Statistics: Application in Presentation Attack Detection." In: *IEEE Access* 8 (2020), pp. 95594–95604.

[140] A. Husseis, J. Liu-Jimenez, and R. Sanchez-Reillo. "Fingerprint Presentation Attack Detection Utilizing Spatio-Temporal Features." In: *Sensors* 21.6 (2021).

[141] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[142] S. Ioffe and C. Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." In: *Proc. Intl. Conf. on Machine Learning (ICML)*. Vol. 37. Proc. of Machine Learning Research. PMLR, 2015, pp. 448–456.

[143] Y. Ishii and M. Takanashi. "Low-cost Unsupervised Outlier Detection by Autoencoders with Robust Estimation." In: *Journal of Information Processing* 27 (2019), pp. 335–339.

[144] ISO/IEC JTC1 SC27 Security Techniques. *ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection*. 2011.

[145] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Information Technology - Biometric presentation attack detection - Part 1: Framework*. 2016.

[146] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 2382-37:2017 Information Technology - Vocabulary - Part 37: Biometrics*. 2017.

[147] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*. 2017.

[148] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC DIS 30107-2. Information Technology - Biometric presentation attack detection - Part 2: Data formats*. International Organization for Standardization. 2017.

[149] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2021. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*. International Organization for Standardization and International Electrotechnical Committee. 2021.

[150] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. "An Identity-authentication System Using Fingerprints." In: *Proc. of the IEEE* 85.9 (1997), pp. 1365–1388.

[151] A. K. Jain and S. Z. Li. *Handbook of Face Recognition*. Vol. 1. Springer, 2011.

[152] A. K. Jain, A. Ross, and U. Uludag. "Biometric Template Security: Challenges and Solutions." In: *Proc. European Signal Processing Conf.* IEEE. 2005, pp. 1–4.

[153] F. James. "A Review of Pseudorandom Number Generators." In: *Computer Physics Communications* 60.3 (1990), pp. 329–344.

[154] H. U. Jang, H. Y. Choi, D. Kim, J. Son, and H. K. Lee. "Fingerprint Spoof Detection Using Contrast Enhancement and Convolutional Neural Networks." In: *Proc. Intl. Conf. on Information Science and Applications (ICISA)*. 2017, pp. 331–338.

[155] X. Jia, X. Yang, K. Cao, Y. Zang, N. Zhang, R. Dai, X. Zhu, and J. Tian. "Multi-scale Local Binary Pattern with Filters for Spoof Fingerprint Detection." In: *Information Sciences* 268 (2014), pp. 91–102.

[156] W. Jian, Y. Zhou, and H. Liu. "Densely Connected Convolutional Network Optimized by Genetic Algorithm for Fingerprint Liveness Detection." In: *IEEE Access* (2020), pp. 2229–2243.

[157] Y. Jiang and X. Liu. "Uniform Local Binary Pattern for Fingerprint Liveness Detection in the Gaussian Pyramid." In: *Hindawi Journal of Electrical and Computer Engineering* 2018 (2018), pp. 1–9.

[158] J. Joyce. *Bayes' theorem*. 2003.

[159] R. Jozefowicz, W. Zaremba, and I. Sutskever. "An Empirical Exploration of Recurrent Network Architectures." In: *Proc. Intl. Conf. on Machine Learning*. 2015, pp. 2342–2350.

[160] S. Kamara and M. Raykova. "Secure Outsourced Computation in a Multi-Tenant Cloud." In: *Proc. IBM Workshop on Cryptography and Security in Clouds*. 2011, pp. 15–16.

[161] O. Kanich, M. Drahansky, and M. Mézl. "Use of Creative Materials for Fingerprint Spoofs." In: *Proc. Intl. Workshop on Biometrics and Forensics (IWBF)*. 2018, pp. 1–8.

[162] J. Kannala and E. Rahtu. "BSIF: Binarized Statistical Image Features." In: *Proc. Intl. Conf. on Pattern Recognition (ICPR)*. 2012, pp. 1363–1366.

[163] C. Karabat, M. S. Kiraz, H. Erdogan, and E. Savas. "THRIVE: Threshold Homomorphic Encryption based Secure and Privacy Preserving Biometric Verification System." In: *EURASIP Journal on Advances in Signal Processing* 2015.1 (2015), pp. 1–18.

[164] E. Karnin, J. Greene, and M. Hellman. "On Secret Sharing Systems." In: *IEEE Trans. on Information Theory* 29.1 (1983), pp. 35–41.

[165] C. Kauba, L. Debiasi, and A. Uhl. "Enabling Fingerprint Presentation Attacks: Fake Fingerprint Fabrication Techniques and Recognition Performance." In: *arXiv preprint arXiv:2012.00606* (2020).

[166] M. Ke, C. Lin, and Q. Huang. "Anomaly Detection of Logo Images in the Mobile Phone Using Convolutional Autoencoder." In: *Proc. Intl. Conf. on Systems and Informatics (ICSAI)*. IEEE. 2017, pp. 1163–1168.

[167] P. Keilbach, J. Kolberg, M. Gomez-Barrero, C. Busch, and H. Langweg. "Fingerprint Presentation Attack Detection using Laser Speckle Contrast Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2018, pp. 1–6.

[168] R. Kessler, O. Henninger, and C. Busch. "Fingerprints, forever young?" In: *Proc. Intl. Conf. on Pattern Recognition (ICPR)*. 2021, pp. 8647–8654.

[169] R. Kiefer, J. Stevens, A. Patel, and M. Patel. "A Survey on Spoofing Detection Systems for Fake Fingerprint Presentation Attacks." In: *Proc. Intl. Conf. on Information and Communication Technology for Intelligent Systems (ICTIS)*. Springer. 2020, pp. 315–334.

[170] S. Kim, B. Park, B. Song, and S. Yang. "Deep Belief Network based Statistical Feature Learning for Fingerprint Liveness Detection." In: *Pattern Recognition Letters* 77 (2016), pp. 58–65.

[171] E. J. Kindt. *Privacy and Data Protection Issues of Biometric Applications*. Vol. 1. Springer, 2016.

[172] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *Proc. Intl. Conf. on Learning Representations (ICLR)*. 2015, pp. 1–13.

[173] J. Kolberg, P. Bauspieß, M. Gomez-Barrero, C. Rathgeb, M. Dürmuth, and C. Busch. "Template Protection based on Homomorphic Encryption: Computationally Efficient Application to Iris-Biometric Verification and Identification." In: *Proc. IEEE Workshop on Information Forensics and Security (WIFS)*. 2019, pp. 1–6.

[174] J. Kolberg, P. Drozdowski, M. Gomez-Barrero, C. Rathgeb, and C. Busch. "Efficiency Analysis of Post-quantum-secure Face Template Protection Schemes based on Homomorphic Encryption." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2020, pp. 1–4.

[175] J. Kolberg, M. Gomez-Barrero, and C. Busch. "Multi-algorithm Benchmark for Fingerprint Presentation Attack Detection with Laser Speckle Contrast Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2019, pp. 1–5.

[176] J. Kolberg, M. Gomez-Barrero, and C. Busch. "On the Generalisation Capabilities of Fingerprint Presentation Attack Detection Methods in the Short Wave Infrared Domain." In: *IET Biometrics* (2021), pp. 1–15.

[177] J. Kolberg, M. Gomez-Barrero, S. Venkatesh, R. Ramachandra, and C. Busch. "Presentation Attack Detection for Finger Recognition." In: *Handbook of Vascular Biometrics*. Springer, 2020, pp. 435–463.

[178] J. Kolberg, M. Grimmer, M. Gomez-Barrero, and C. Busch. "Anomaly Detection with Convolutional Autoencoders for Fingerprint Presentation Attack Detection." In: *Trans. on Biometrics, Behavior, and Identity Science (TBIOM)* 3.2 (2021), pp. 190–202.

[179] J. Kolberg, A. C. Vasile, M. Gomez-Barrero, and C. Busch. "Analysing the Performance of LSTMs and CNNs on 1310 nm Laser Data for Fingerprint Presentation Attack Detection." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. 2020, pp. 1–7.

[180] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet Classification with Deep Convolutional Neural Networks." In: *Advances in Neural Information Processing Systems*. 2012, pp. 1097–1105.

[181] R. Kulkarni and A. Namboodiri. "Secure Hamming Distance based Biometric Authentication." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2013, pp. 1–6.

[182] A. Kumar and A. Passi. "Comparison and Combination of Iris Matchers for Reliable Personal Authentication." In: *Pattern Recognition* 43.3 (2010), pp. 1016–1026.

[183] N. Kumar. "Cancelable Biometrics: A Comprehensive Survey." In: *Artificial Intelligence Review* (2019), pp. 1–44.

[184] P. A. Lachenbruch and M. Goldstein. "Discriminant Analysis." In: *Biometrics* (1979), pp. 69–85.

[185] C. Lee, S. Lee, and J. Kim. "A Study of Touchless Fingerprint Recognition System." In: *Joint IAPR Intl. Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer. 2006, pp. 358–365.

[186] M. H. Lim and A. B. J. Teoh. "A Novel Encoding Scheme for Effective Biometric Discretization: Linearly Separable Subcode." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 35.2 (2012), pp. 300–313.

[187] C. Lin and A. Kumar. "Matching Contactless and Contact-based Conventional Fingerprint Images for Biometrics Identification." In: *IEEE Trans. on Image Processing (TIP)* 27.4 (2018), pp. 2008–2021.

[188] Y. Lindell and B. Pinkas. "Secure Two-Party Computation via Cut-and-Choose Oblivious Transfer." In: *Journal of Cryptology (JoC)* 25.4 (2012), pp. 680–722.

[189] F. Liu, G. Liu, and X. Wang. "High-accurate and Robust Fingerprint Anti-spoofing System Using Optical Coherence Tomography." In: *Expert Systems with Applications* 130 (2019), pp. 31–44.

[190] F. Liu, H. Liu, W. Zhang, G. Liu, and L. Shen. "One-Class Fingerprint Presentation Attack Detection Using Auto-Encoder Network." In: *IEEE Trans. on Image Processing (TIP)* 30 (2021), pp. 2394–2407.

[191] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen. "From BoW to CNN: Two Decades of Texture Representation for Texture Classification." In: *Intl. Journal of Computer Vision (IJCV)* 127.1 (2019), pp. 74–109.

[192] Y. Luo, S. C. Sen-ching, and S. Ye. "Anonymous Biometric Access Control Based on Homomorphic Encryption." In: *Proc. Intl. Conf. on Multimedia and Expo (ICME)*. IEEE. 2009, pp. 1046–1049.

[193] V. Lyubashevsky, C. Peikert, and O. Regev. "On Ideal Lattices and Learning with Errors Over Rings." In: *Journal of the ACM (JACM)* 60.6 (2013), pp. 1–35.

[194] L. Ma, T. Tan, Y. Wang, and D. Zhang. "Efficient Iris Recognition by Characterizing Key Local Variations." In: *IEEE Trans. on Image Processing* 13.6 (2004), pp. 739–750.

[195] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain. "On the Reconstruction of Face Images from Deep Face Templates." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)* 41.5 (2018), pp. 1188–1202.

[196] N. Manivanan, S. Memon, and W. Balachandran. "Automatic Detection of Active Sweat Pores of Fingerprint Using Highpass and Correlation Filtering." In: *Electronics letters* 46.18 (2010), pp. 1268–1269.

[197] E. Marasco and A. Ross. "A Survey on Antispoofing Schemes for Fingerprint Recognition Systems." In: *ACM Computing Surveys (CSUR)* 47.2 (2014), pp. 1–36.

[198] E. Marasco and C. Sansone. "On the Robustness of Fingerprint Liveness Detection Algorithms against New Materials used for Spoofing." In: *BIOSIGNALS*. Vol. 8. 2011, pp. 553–555.

[199] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans. *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Springer, 2019.

[200] G. M. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers. "First International Fingerprint Liveness Detection Competition - LivDet 2009." In: *Proc. Intl. Conf. on Image Analysis and Processing (ICIAP)*. 2009, pp. 12–23.

[201] M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia, and J. A. Siguenza. "Hill-Climbing and Brute-Force Attacks on Biometric Systems: A Case Study in Match-on-Card Fingerprint Verification." In: *Proc. Intl. Carnahan Conference on Security Technology*. IEEE. 2006, pp. 151–159.

[202] L. Masek. "Recognition of Human Iris Patterns for Biometric Identification." MA thesis. University of Western Australia, 2003.

[203] S. Memon, N. Manivannan, and W. Balachandran. "Active Pore Detection for Liveness in Fingerprint Identification System." In: *Proc. of Telecommunications Forum (TELFOR)*. IEEE. 2011, pp. 619–622.

[204] V. S. Miller. "Use of Elliptic Curves in Cryptography." In: *Proc. Conf. on the Theory and Application of Cryptographic Techniques*. Springer. 1985, pp. 417–426.

[205] H. Mirzaalian, M. Hussein, and W. Abd-Almageed. "On the Effectiveness of Laser Speckle Contrast Imaging and Deep Neural Networks for Detecting Known and Unknown Fingerprint Presentation Attacks." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2019, pp. 1–8.

[206] N. Miura, A. Nagasaka, and T. Muyatake. "Extraction of Finger-vein Patterns using Maximum Curvature Points in Image Profiles." In: *IEICE Trans. on Information and Systems* 90.8 (2007), pp. 1185–1194.

[207] Y. Moolla, L. Darlow, A. Sharma, A. Singh, and J. van der Merwe. "Optical Coherence Tomography for Fingerprint Presentation Attack Detection." In: *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Ed. by S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans. Springer, 2019, pp. 49–70.

[208] V. Mura, L. Ghiani, G. L. Marcialis, F. Roli, D. Yambay, and S. Schuckers. "LivDet 2015 Fingerprint Liveness Detection Competition 2015." In: *Proc. Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*. IEEE. 2015, pp. 1–6.

[209] V. Mura, G. Orrù, R. Casula, A. Sibiriu, G. Loi, P. Tuveri, L. Ghiani, and G. L. Marcialis. "LivDet 2017 Fingerprint Liveness Detection Competition 2017." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2018, pp. 297–302.

[210] K. Nandakumar and A. K. Jain. "Biometric Template Protection: Bridging the Performance Gap Between Theory and Practice." In: *IEEE Signal Processing Magazine* 32.5 (2015), pp. 88–100.

[211] I. Natgunanathan, A. Mehmood, Y. Xiang, G. Beliakov, and J. Yearwood. "Protection of Privacy in Biometric Data." In: *IEEE Access* 4 (2016), pp. 880–892.

[212] A. Nautsch, S. Isadskiy, J. Kolberg, M. Gomez-Barrero, and C. Busch. "Homomorphic Encryption for Speaker Recognition: Protection of Biometric Templates and Vendor Model Parameters." In: *Proc. Speaker Odyssey*. 2018, pp. 16–23.

[213] A. Nautsch, A. Jiménez, A. Treiber, J. Kolberg, C. Jasserand, E. Kindt, H. Delgado, M. Todisco, M. A. Hmani, A. Mtibaa, et al. "Preserving Privacy in Speaker and Speech Characterisation." In: *Computer Speech & Language* 58 (2019), pp. 441–480.

[214] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. "Iris Recognition with off-the-shelf CNN Features: A Deep Learning Perspective." In: *IEEE Access* 6 (2017), pp. 18848–18855.

[215] S. B. Nikam and S. Agarwal. "Texture and Wavelet-based Spoof Fingerprint Detection for Fingerprint Biometric Systems." In: *Proc. Intl. Conf. on Emerging Trends in Engineering and Technology*. IEEE. 2008, pp. 675–680.

[216] O. Nikisins, A. George, and S. Marcel. "Domain Adaptation in Multi-Channel Autoencoder based Features for Robust Face Anti-Spoofing." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2019, pp. 1–8.

[217] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel. "On Effectiveness of Anomaly Detection Approaches against Unseen Presentation Attacks in Face Anti-Spoofing." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2018, pp. 75–81.

[218] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. "Fingerprint Liveness Detection Using Convolutional Neural Networks." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 11.6 (2016), pp. 1206–1213.

[219] T. Ojala, M. Pietikäinen, and D. Harwood. "A Comparative Study of Texture Measures with Classification Based on Featured Distributions." In: *Pattern Recognition* 29.1 (1996), pp. 51–59.

[220] H. Olkkonen and P. Pesola. "Gaussian Pyramid Wavelet Transform for Multiresolution Analysis of Images." In: *Graphical Models and Image Processing* 58.4 (1996), pp. 394–398.

[221] G. Orrù, R. Casula, P. Tuveri, C. Bazzoni, G. Dessalvi, M. Micheletto, L. Ghiani, and G. L. Marcialis. "LivDet in Action - Fingerprint Liveness Detection Competition 2019." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2019, pp. 1–6.

[222] M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich. "SCiFI - A System for Secure Face Identification." In: *Proc. IEEE Symposium on Security and Privacy*. IEEE. 2010, pp. 239–254.

[223] C. Paar and J. Pelzl. "Message Authentication Codes (MACs)." In: *Understanding Cryptography*. Springer, 2010, pp. 319–330.

[224] P. Paillier. "Public-key Cryptosystems Based on Composite Degree Residuosity Classes." In: *Proc. Intl. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*. Springer. 1999, pp. 223–238.

[225] O. M. Parkhi, A. Vedaldi, and A. Zisserman. "Deep Face Recognition." In: *British Machine Vision Association* (2015), pp. 1–12.

[226] V. M. Patel, N. K. Ratha, and R. Chellappa. "Cancelable Biometrics: A Review." In: *IEEE Signal Processing Magazine* 32 (2015), pp. 54–65.

[227] C. Patsakis, van J. Rest, M. Choraś, and M. Bouroche. "Privacy-Preserving Biometric Authentication and Matching via Lattice-Based Encryption." In: *Proc. Intl. Workshop on Data Privacy Management*. Springer. 2015, pp. 169–182.

[228] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[229] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. "The FERET Evaluation Methodology for Face-Recognition Algorithms." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 22.10 (2000), pp. 1090–1104.

[230] B. Pinkas, T. Schneider, N. P. Smart, and S. C. Williams. "Secure Two-Party Computation is Practical." In: *Proc. Intl. Conf. on the Theory and Application of Cryptology and Information Security*. Springer. 2009, pp. 250–267.

[231] R. Plesh, K. Bahmani, G. Jang, D. Yambay, K. Brownlee, T. Swyka, P. Johnson, A. Ross, and S. Schuckers. "Fingerprint Presentation Attack Detection utilizing Time-Series, Color Fingerprint Captures." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2019, pp. 1–8.

[232] X. Qian, X. S. Hua, P. Chen, and L. Ke. "PLBP: An Effective Local Binary Patterns Texture Descriptor with Pyramid Representation." In: *Pattern Recognition* 44.10 (2011), pp. 2502–2515.

[233] J. R. Quinlan. "Induction of Decision Trees." In: *Machine Learning* 1.1 (1986), pp. 81–106.

[234] A. Radford, L. Metz, and S. Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." In: *arXiv preprint arXiv:1511.06434* (2015).

[235] R. Raghavendra, K. Raja, J. Surbiryala, and C. Busch. "A low-cost Multimodal Biometric Sensor to capture Finger Vein and Fingerprint." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. 2014.

[236] S. D. Rane, W. Sun, and A. Vetro. "Secure Distortion Computation Among Untrusting Parties Using Homomorphic Encryption." In: *Proc. IEEE Intl. Conf. on Image Processing (ICIP)*. IEEE. 2009, pp. 1485–1488.

[237] C. Rathgeb and A. Uhl. "Attacking Iris Recognition: An Efficient Hill-Climbing Technique." In: *Proc. Intl. Conf. on Pattern Recognition (ICPR)*. IEEE. 2010, pp. 1217–1220.

[238] C. Rathgeb and A. Uhl. "A Survey on Biometric Cryptosystems and Cancelable Biometrics." In: *EURASIP Journal on Information Security* 2011.1 (2011), pp. 1–25.

[239] C. Rathgeb, A. Uhl, and P. Wild. "Incremental Iris Recognition: A Single-algorithm Serial Fusion Strategy to Optimize Time Complexity." In: *Proc. Intl. Conf. on Biometrics: Theory Applications and Systems (BTAS)*. IEEE. 2010, pp. 1–6.

[240] C. Rathgeb, A. Uhl, P. Wild, and H. Hofbauer. "Design Decisions for an Iris Recognition SDK." In: *Handbook of Iris Recognition*. second. Advances in Computer Vision and Pattern Recognition. Springer, 2016, pp. 359–396.

[241] A. Rattani and R. Derakhshani. "On Fine-Tuning Convolutional Neural Networks for Smartphone based Ocular Recognition." In: *Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB)*. IEEE. 2017, pp. 762–767.

[242] A. Rattani, W. J. Scheirer, and A. Ross. "Open Set Fingerprint Spoof Detection Across Novel Fabrication Materials." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 10.11 (2015), pp. 2447–2460.

[243] P. V. Reddy, A. Kumar, S. M. K. Rahman, and T. S. Mundra. "A New Antispoofing Approach for Biometric Devices." In: *IEEE Trans. on Biomedical Circuits and Systems (TBioCAS)* 2.4 (2008), pp. 328–337.

[244] O. Regev. "On Lattices, Learning with Errors, Random Linear Codes, and Cryptography." In: *Journal of the ACM (JACM)* 56.6 (2009), pp. 1–40.

[245] D. A. Reynolds. "Gaussian Mixture Models." In: *Encyclopedia of Biometrics* 741 (2009), pp. 659–663.

[246] R. L. Rivest, A. Shamir, and L. Adleman. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." In: *Communications of the ACM* 21.2 (1978), pp. 120–126.

[247] C. Roberts. "Biometric Attack Vectors and Defences." In: *Computers & Security* 26.1 (2007), pp. 14–25.

[248] P. J. Rousseeuw and M. Hubert. "Robust Statistics for Outlier Detection." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1.1 (2011), pp. 73–79.

[249] R. K. Rowe, K. A. Nixon, and P. W. Butler. "Multispectral Fingerprint Image Acquisition." In: *Advances in Biometrics: Sensors, Algorithms and Systems*. Springer London, 2008, pp. 3–23.

[250] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. "ImageNet Large Scale Visual Recognition Challenge." In: *Intl. Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252.

[251] A. R. Sadeghi, T. Schneider, and I. Wehrenberg. "Efficient Privacy-preserving Face Recognition." In: *Proc. Intl. Conf. on Information Security and Cryptology*. Springer. 2009, pp. 229–244.

[252] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. "Image Classification with the Fisher Vector: Theory and Practice." In: *Intl. Journal of Computer Vision (IJCV)* 105.3 (2013), pp. 222–245.

[253] M. Sandhya and M. V. N. K. Prasad. "Biometric Template Protection: A Systematic Literature Review of Approaches and Modalities." In: *Biometric Security and Privacy* (2017), pp. 323–370.

[254] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 4510–4520.

[255] S. B. Sandouka, Y. Bazi, and N. Alajlan. "Transformers and Generative Adversarial Networks for Liveness Detection in Multitarget Fingerprint Sensors." In: *Sensors* 21.3 (2021).

[256] F. Schroff, D. Kalenichenko, and J. Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 815–823.

[257] *Microsoft SEAL (release 3.2)*. https://github.com/Microsoft/SEAL. Microsoft Research, Redmond, WA. Feb. 2019.

[258] J. Senarathna, A. Rege, N. Li, and N. V. Thakor. "Laser Speckle Contrast Imaging: Theory, instrumentation and applications." In: *IEEE Reviews in Biomedical Engineering* 6 (2013), pp. 99–110.

[259] N. K. Shaydyuk and T. Cleland. "Biometric Identification via Retina Scanning with Liveness Detection using Speckle Contrast Imaging." In: *Proc. Intl. Carnahan Conf. on Security Technology (ICCST)*. IEEE, 2016, pp. 1–5.

[260] K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." In: *Proc. Intl. Conf. on Learning Representations (ICLR)*. 2015.

[261] J. M. Singh, A. Madhun, G. Li, and R. Ramachandra. "A Survey on Unknown Presentation Attack Detection for Fingerprint." In: *Proc. Intl. Conf. on Intelligent Technologies and Applications (INTAP)*. 2020, pp. 189–202.

[262] J. Sivic and A. Zisserman. "Efficient Visual Search of Videos Cast as Text Retrieval." In: *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)* 31.4 (2008), pp. 591–606.

[263] S. Skansi. "Autoencoders." In: *Introduction to Deep Learning*. Springer, 2018, pp. 153–163.

[264] K. Soomro, A. R. Zamir, and M. Shah. "A Dataset of 101 Human Action Classes from Videos in the Wild." In: *Center for Research in Computer Vision* 2.11 (2012).

[265] C. Sousedik and C. Busch. "Presentation Attack Detection Methods for Fingerprint Recognition Systems: A Survey." In: *IET Biometrics* 3.1 (2014), pp. 1–15.

[266] C. Sousedik and C. Busch. "Volumetric Fingerprint Data Analysis Using Optical Coherence Tomography." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. IEEE Computer Society, 2014.

[267] L. Spinoulas, M. Hussein, D. Geissbühler, J. Mathai, O. G. Almeida, G. Clivaz, S. Marcel, and W. AbdAlmageed. "Multispectral Biometrics System Framework: Application to Presentation Attack Detection." In: *IEEE Sensors Journal* (2021).

[268] L. Spinoulas, H. Mirzaalian, M. Hussein, and W. AbdAlmageed. "Multi-Modal Fingerprint Presentation Attack Detection: Evaluation On A New Dataset." In: *Trans. on Biometrics, Behavior, and Identity Science (TBIOM)* (2021).

[269] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. "Striving for Simplicity: The All Convolutional Net." In: *Proc. Intl. Conf. on Learning Representations (ICLR)*. (workshop track). 2015.

[270] H. Steiner, A. Kolb, and N. Jung. "Reliable Face Anti-Spoofing using Multispectral SWIR Imaging." In: *Proc. Intl. Conf. on Biometrics (ICB)*. 2016, pp. 1–8.

[271] H. Steiner, S. Sporrer, A. Kolb, and N. Jung. "Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification." In: *Journal of Sensors* 2016 (2016), pp. 1129–1144.

[272] J. Steinman, M. Kunze, S. Barton, A. Sas, C. Zenz, and G. Kranner. *Tanz der Vampire: Das Musical; Gesamtaufnahme*. Polygram, 1998.

[273] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." In: *Proc. AAAI Conf. on Artificial Intelligence*. 2017, pp. 4278–4284.

[274] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the Inception Architecture for Computer Vision." In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2818–2826.

[275] B. Tan, A. Lewicke, D. Yambay, and S. Schuckers. "The Effect of Environmental Conditions and Novel Spoofing Methods on Fingerprint Anti-spoofing Algorithms." In: *Proc. Intl. Workshop on Information Forensics and Security (WIFS)*. IEEE. 2010, pp. 1–6.

[276] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. "A Survey on Deep Transfer Learning." In: *Proc. Intl. Conf. on Artificial Neural Networks*. Springer. 2018, pp. 270–279.

[277] Y. Tang, F. Gao, J. Feng, and Y. Liu. "FingerNet: An Unified Deep Network for Fingerprint Minutiae Extraction." In: *Proc. Intl. Joint Conf. on Biometrics (IJCB)*. IEEE. 2017, pp. 108–116.

[278] D. M. J. Tax. "One-class Classification: Concept Learning in the Absence of Counter-examples." PhD thesis. TU Delft, 2002.

[279] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia. "Biometric Presentation Attack Detection: Beyond the Visible Spectrum." In: *IEEE Trans. on Information Forensics and Security (TIFS)* 15 (2019), pp. 1261–1275.

[280] R. Tolosana, M. Gomez-Barrero, J. Kolberg, A. Morales, C. Busch, and J. Ortega. "Towards Fingerprint Presentation Attack Detection Based on Convolutional Neural Networks and Short Wave Infrared Imaging." In: *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. 2018, pp. 1–5.

[281] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. "Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics." In: *IEEE Access* 6 (2018), pp. 5128–5138.

[282] A. Toosi, A. Bottino, S. Cumani, P. Negri, and P. L. Sottile. "Feature Fusion for Fingerprint Liveness Detection: A Comparative Study." In: *IEEE Access* 5 (2017), pp. 23695–23709.

[283] A. Toosi, S. Cumani, and A. Bottino. "CNN Patch-Based Voting for Fingerprint Liveness Detection." In: *Proc. Intl. Joint Conf. on Computational Intelligence (IJCCI)*. 2017, pp. 158–165.

[284] A. Treiber, A. Nautsch, J. Kolberg, T. Schneider, and C. Busch. "Privacy-preserving PLDA Speaker Verification Using Outsourced Secure Computation." In: *Speech Communication* 114 (2019), pp. 60–71.

[285] M. Turk and A. Pentland. "Eigenfaces for Recognition." In: *Journal of Cognitive Neuroscience* 3.1 (1991), pp. 71–86.

[286] A. Uhl, C. Busch, S. Marcel, and R. Veldhuis. *Handbook of Vascular Biometrics*. Springer Nature, 2020.

[287] M. Upmanyu, A. M. Namboodiri, K. Srinathan, and C. Jawahar. "Efficient Biometric Verification in Encrypted Domain." In: *Proc. Intl. Conf. on Biometrics (ICB)*. Springer. 2009, pp. 899–908.

[288] P. G. Vaz, A. Humeau-Heurtier, E. Figueiras, C. Correia, and J. Cardoso. "Laser Speckle Imaging to Monitor Microvascular Blood Flow: A Review." In: *IEEE Reviews in Biomedical Engineering* 9 (2016), pp. 106–120.

[289] S. Wang and J. Hu. "Alignment-free Cancelable Fingerprint Template Design: A Densely Infinite-to-one Mapping (DITOM) Approach." In: *Pattern Recognition* 45.12 (2012), pp. 4129–4137.

[290] X. Wang, A. J. Malozemoff, and J. Katz. *EMP-Toolkit: Efficient MultiParty Computation Toolkit*. https://github.com/emp-toolkit. 2016.

[291] S. Wold, K. Esbensen, and P. Geladi. "Principal Component Analysis." In: *Chemometrics and Intelligent Laboratory Systems* 2.1-3 (1987), pp. 37–52.

[292] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. "LivDet 2011 Fingerprint Liveness Detection Competition 2011." In: *Proc. Intl. Conf. on Biometrics (ICB)*. IEEE. 2012, pp. 208–215.

[293] D. Yambay, L. Ghiani, G. L. Marcialis, F. Roli, and S. Schuckers. "Review of Fingerprint Presentation Attack Detection Competitions." In: *Handbook of Biometric Anti-Spoofing*. Springer, 2019, pp. 109–131.

[294] W. Yang, S. Wang, K. Yu, J. J. Kang, and M. N. Johnstone. "Secure Fingerprint Authentication with Homomorphic Encryption." In: *Proc. Digital Image Computing: Techniques and Applications (DICTA)*. 2020, pp. 1–6.

[295] W. Yang, S. Wang, G. Zheng, J. Chaudhry, and C. Valli. "ECB4CI: An Enhanced Cancelable Biometric System for Securing Critical Infrastructures." In: *The Journal of Supercomputing* 74.10 (2018), pp. 4893–4909.

[296] A. C. Yao. "How to Generate and Exchange Secrets." In: *Proc. Annual Symposium on Foundations of Computer Science (SFCS)*. IEEE. 1986, pp. 162–167.

[297] M. Yasuda, T. Shimoyama, J. Kogure, K. Yokoyama, and T. Koshiba. "Packed Homomorphic Encryption Based on Ideal Lattices and Its Application to Biometrics." In: *Proc. Intl. Conf. on Availability, Reliability, and Security (ARES)*. Springer. 2013, pp. 55–74.

[298] M. Yasuda, T. Shimoyama, J. Kogure, K. Yokoyama, and T. Koshiba. "New Packing Method in Somewhat Homomorphic Encryption and Its Applications." In: *Security and Communication Networks* 8.13 (2015), pp. 2194–2213.

[299] S. Ye, Y. Luo, J. Zhao, and S. Cheung. "Anonymous Biometric Access Control." In: *EURASIP Journal on Information Security* 2009.1 (2009), pp. 1–17.

[300] J. Zhu, H. Zou, S. Rosset, and T. Hastie. "Multi-class AdaBoost." In: *Statistics and its Interface* 2.3 (2009), pp. 349–360.

# DECLARATION

Hiermit erkläre ich, die vorgelegte Arbeit zur Erlangung des akademischen Grades *Doktor der Naturwissenschaften (Dr. rer.nat.)* mit dem Titel

*Security Enhancement and Privacy Protection for Biometric Systems*

selbstständig und ohne unerlaubte fremde Hilfe sowie nur mit den angegebenen Hilfen angefertigt zu haben. Alle wörtlichen oder sinngemäß aus veröffentlichten Schriften entnommenen Textstellen und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Die Grundsätze guter wissenschaftlicher Praxis sind eingehalten. Ich habe bisher noch keinen Promotionsversuch unternommen.

*Darmstadt, 8. Juli 2021*

Jascha Kolberg

## COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*".
The diagrams in this thesis were created using *TIKZ, PGFPlots* by Christian Feuersänger.