IWBF 2025

Impact and Mitigation of Quality Degradation for Differential Morphing Attack Detection

Torsten Schlett, Christian Rathgeb, Juan E. Tapia, Christoph Busch

da/sec - Biometrics and Security Research Group
Hochschule Darmstadt, Germany / European University of Technology, European Union

Email: {torsten.schlett, christian.rathgeb, juan.tapia-farias, christoph.busch}@h-da.de

Face recognition task

Subject A

Face recognition task under quality degradation

Subject A blurred

Morphing attack detection

Morph or bona fide image?

Morphing attack detection

Subject A

Subject A+B

Subject B

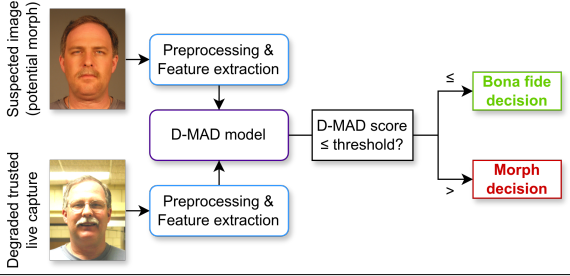
Overview

1. Introduction
2. Dataset
3. Synthetic degradation
4. Quality score impact
5. D-MAD decision impact
6. D-MAD threshold optimization
7. Conclusions

Introduction

The Differential Morphing Attack Detection (**D-MAD**) scenario:

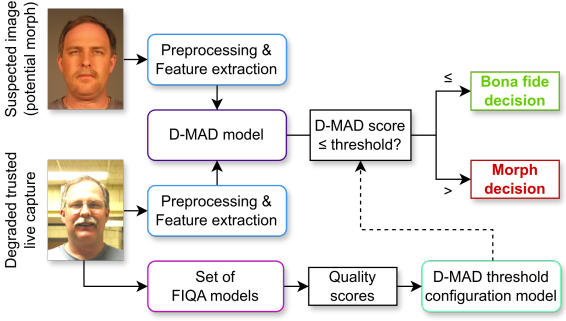
1. There is a previously captured **suspected image**, which could be a morph.
   But this image can be assumed to be of **good quality** (e.g. in a passport).
2. And there is a current **trusted live capture**, which is assumed to be bona fide (not a morph).
   But the quality of this image **may be degraded** by environmental factors (e.g. lighting).
3. Both images are processed by the D-MAD model, which outputs a scalar **D-MAD score**.
4. If the D-MAD score is above a set **threshold**, the suspected image is assumed to be a morph.



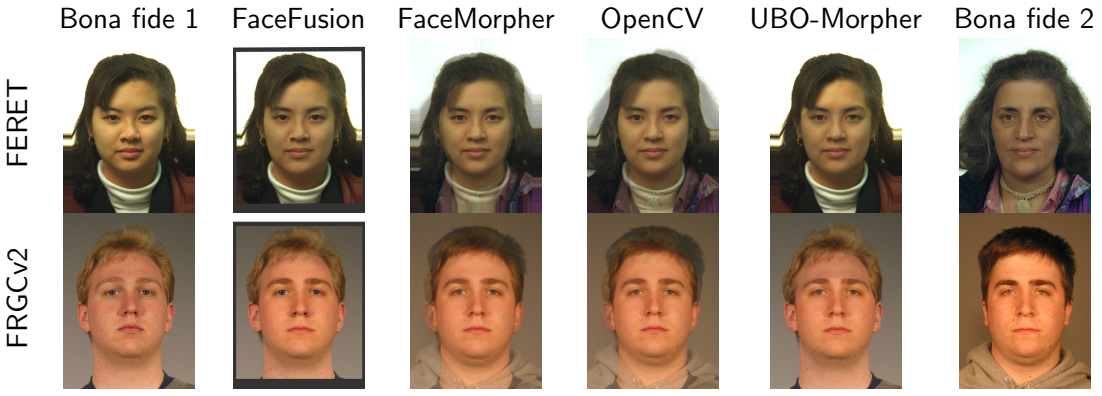
Introduction

A set of Face Image Quality Assessment (**FIQA**) algorithms can be used to compute a corresponding set of **quality scores**. The presented work investigates

- various trusted live capture quality degradations' **impact on quality scores**,
- various trusted live capture quality degradations' **impact on D-MAD decisions**,
- and the training of a **D-MAD decision threshold configuration model** that outputs a D-MAD threshold based on a set of input quality scores per trusted live capture.



Introduction



Cf. “Deep Face Representations for Differential Morphing Attack Detection” figure 4.
See section III.D. for details.

Synthetic degradation

To investigate the image degradation of the trusted live captures,
we opted for **synthetic degradation of four defect types** in this work,
since this allows for a clearly **controlled degradation** of all base images.

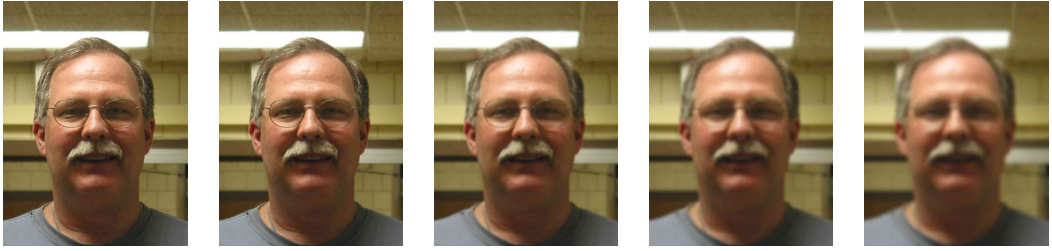
In particular, we utilized the synthetic degradation previously employed in the
NIST FATE Quality SIDD report 2024-04-26,
which correspond to environmental factors:

- Corresponding to camera focus / motion:
  -  **Gaussian blur** (called “Resolution” in the NIST report)
  -  **Motion blur**
- Corresponding to lighting:
  -  **Overexposure**
  -  **Underexposure**

Synthetic degradation

Defect type: Gaussian blur

Based on NIST FATE Quality SIDD report 2024-04-26, where this is called “Resolution”, this uses ImageMagick’s “convert -gaussian-blur 0×*(severity)*”:



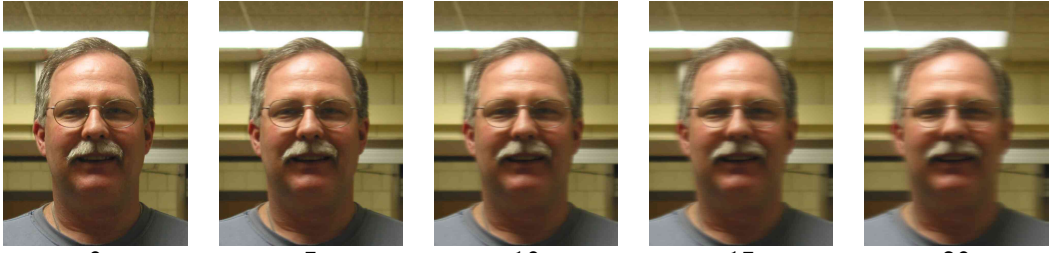
0 1 3 5 7

Degradation severity steps (equal to the ImageMagick setting)

Synthetic degradation

Defect type: Motion blur

Based on NIST FATE Quality SIDD report 2024-04-26,
this uses ImageMagick's "convert -motion-blur 0x*(severity)*":



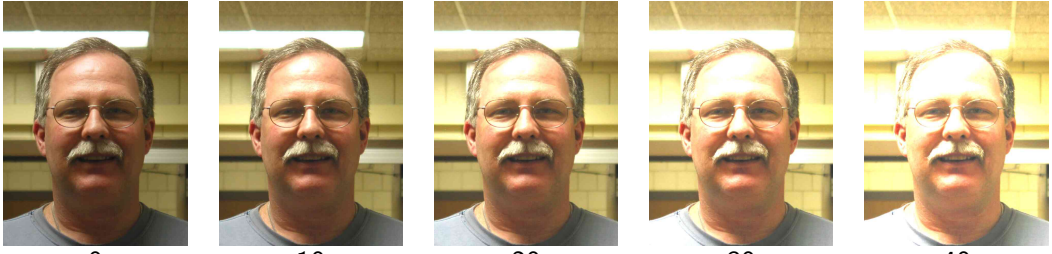
0 5 10 15 20

Degradation severity steps (equal to the ImageMagick setting)

Synthetic degradation

Defect type: Overexposure

Based on NIST FATE Quality SIDD report 2024-04-26,
this uses ImageMagick's "convert -brightness-contrast *(severity)*x*(severity)*":



0 10 20 30 40

Degradation severity steps (equal to the ImageMagick setting)

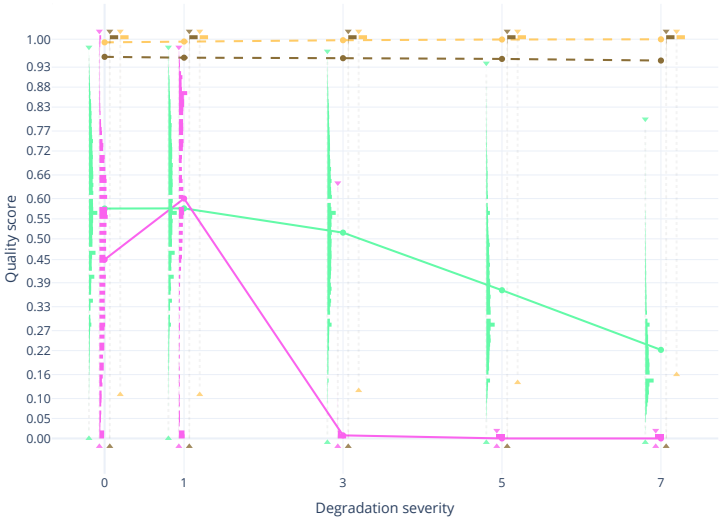
Quality score impact

How does the trusted live capture image degradation impact the quality score output for FIQA algorithms?

- The used FIQA algorithms are parts of the OFIQ project:
  `https://github.com/BSI-OFIQ/OFIQ-Project`
  (OFIQ stands for "Open Source Face Image Quality")
- OFIQ is the reference implementation for the next edition of ISO/IEC 29794-5.
- Specifically these OFIQ measures were selected, based on the assumption that these are the most relevant ones for the investigated defect types:

Defect type	**OFIQ measure**	Type
All	**Unified**	CNN (MagFace)
Gaussian blur	**Sharpness**	Hand-crafted and random forest
Motion blur	**Sharpness**	Hand-crafted and random forest
Overexposure	**Over-Exposure-Prevention**	Hand-crafted
Underexposure	**Under-Exposure-Prevention**	Hand-crafted

Quality score impact

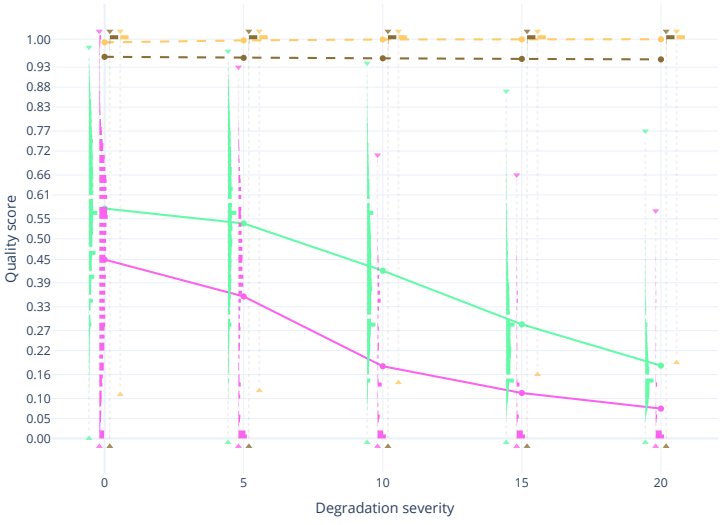


Gaussian blur

Related
- Unified (MagFace)
- Sharpness

Unrelated
- Under-Exposure-Prevention
- Over-Exposure-Prevention

Quality score impact

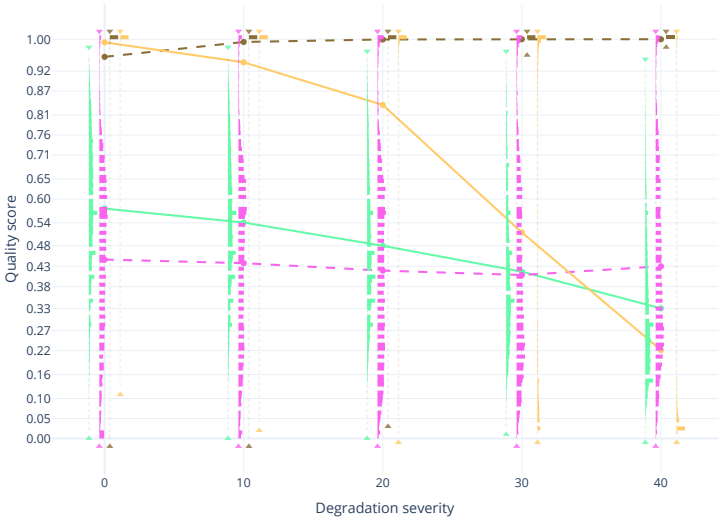


Motion blur

Related
- Unified (MagFace)
- Sharpness

Unrelated
- Under-Exposure-Prevention
- Over-Exposure-Prevention

Quality score impact



Overexposure

Related
- Unified (MagFace)
- Over-Exposure-Prevention

Unrelated
- Sharpness
- Under-Exposure-Prevention

Quality score impact

Expected conclusions from the quality score impact results:

- The OFIQ measures that were expected to be **related** to a defect type also responded to that defect type, while the ones expected to be **unrelated** did not respond as much.
- For instance, the unified OFIQ measure (MagFace) responded to all degradations.

Unexpected conclusions from the quality score impact results:

- OFIQ's **sharpness measure** responded more to **Gaussian blur** than to **motion blur**.
  - Note that the used OFIQ version did not have a motion blur measure, which could be added in the future.
- And for **Gaussian blur** the **sharpness measure** yielded somewhat higher quality scores for the degradation severity step 1 than for step 0, which probably should not happen.
  - This indicates that the sharpness measure could be improved for the next OFIQ version.
  - There also was a sharp fall-off to 0 quality scores at and beyond degradation severity step 3, but this may not necessarily be a functional issue for real use cases.

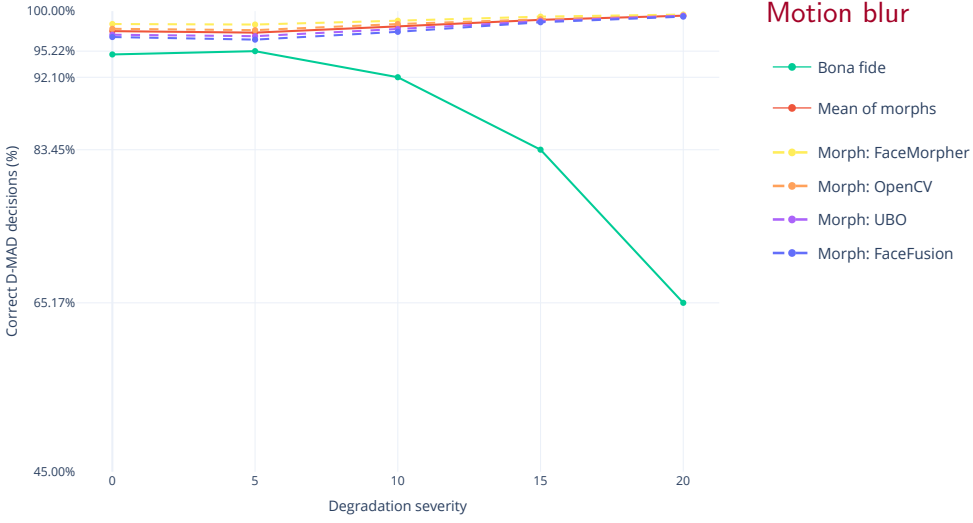
D-MAD decision impact

How does the trusted live capture image degradation impact the correct D-MAD decision percentages for bona fide and morph cases?

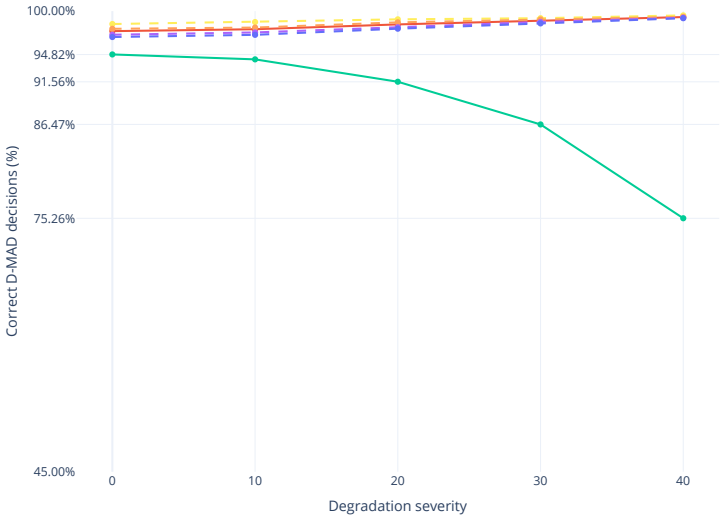
The following plots are similar to the prior quality score impact plots, with these differences:

- The Y-axis shows correct D-MAD decision percentages, **i.e. higher is always better**.
  - So in the perfect case all values would be at 100%.
  - This is in contrast to the prior quality score plots, where the quality scores related to the defect type should decrease as the degradation severity steps increase.
- The different curves correspond to cases that are either
  - known to be **bona fide** (i.e. the correct D-MAD decision is "bona fide"),
  - or known to be a **morph** (i.e. the correct D-MAD decision is "morph"), whereby the morph case curves are separated by the dataset's four different morph types, plus a corresponding mean curve.

D-MAD decision impact



D-MAD decision impact



Overexposure

- Bona fide
- Mean of morphs
- Morph: FaceMorpher
- Morph: OpenCV
- Morph: UBO
- Morph: FaceFusion

D-MAD decision impact

Conclusions from the D-MAD impact results:

- Unsurprisingly, both blur and exposure defects can substantially affect D-MAD.
  - For the investigated D-MAD model, the trusted live capture **degradation led to higher D-MAD scores**, hence why the correct D-MAD decision percentage **improved slightly for morph cases** but **worsened substantially for the bona fide cases**.
- Underexposure had a stronger D-MAD impact than overexposure.
  -  But this may not be too surprising since the underexposure effect appears to be visibly more pronounced as well, despite lower settings.

Conclusions from the D-MAD impact results **in relation to prior FIQA impact results**:

-  **Motion blur had a stronger D-MAD impact** than **Gaussian blur** of a visually roughly similar magnitude.
- Yet OFIQ's **sharpness measure responded more to Gaussian blur** than motion blur.
  - This further indicates that the next OFIQ version could benefit from an improved sharpness measure or an additional motion blur measure.

D-MAD optimization potential

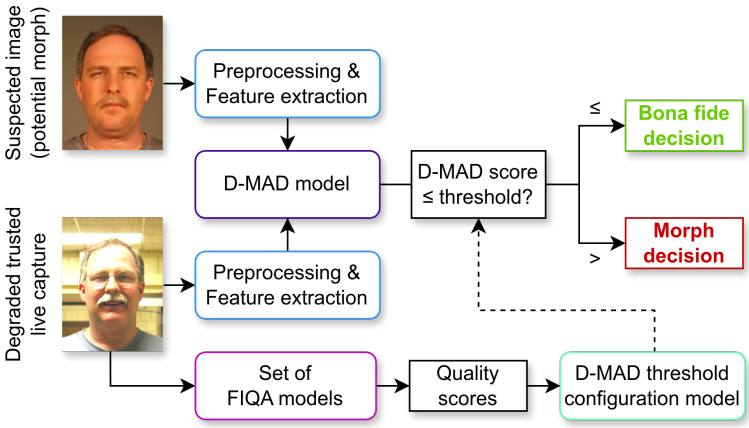
The previous slides' results use the D-MAD system's default decision threshold value (0.5).

What if an optimal threshold were chosen for each degradation severity step per defect type?

- "Optimal" in our analysis refers to the minimization of the worst (lowest) of the correct decision percentages across the curve types (bona fide & 4 morphs).
- These optimal thresholds can yield substantially better decision percentages.

Threshold configuration model

How well can we train a model to output D-MAD thresholds specific to a trusted live capture, based only on the set of quality scores for that trusted live capture?

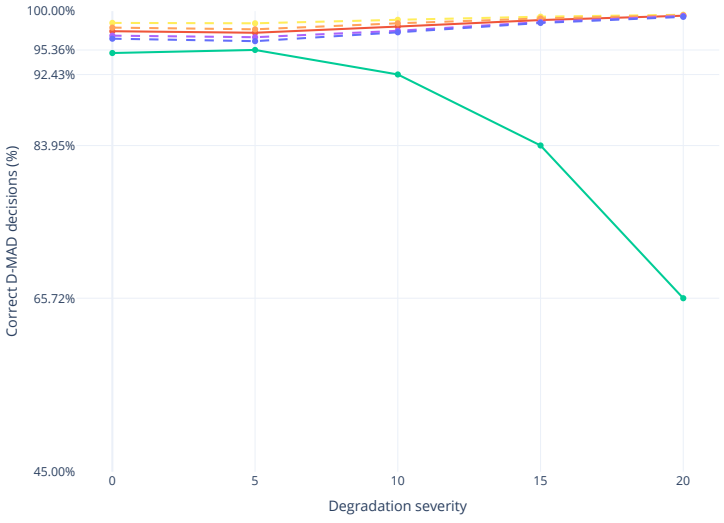


Threshold configuration model

Dataset split: Approximately 10% training data, 10% validation data, 80% test data.
Model input: The four relevant OFIQ components.
Training target: Thresholds as shown in "D-MAD optimization potential" (X-axis labels).
Used model type: *sklearn.ensemble.HistGradientBoostingRegressor* with *max_iter* 200.
(Training time below 1s, prediction time below 100ms for the entire validation data.)

- The *HistGradientBoostingRegressor* was selected after evaluating alternative models on the basis of the validation data.

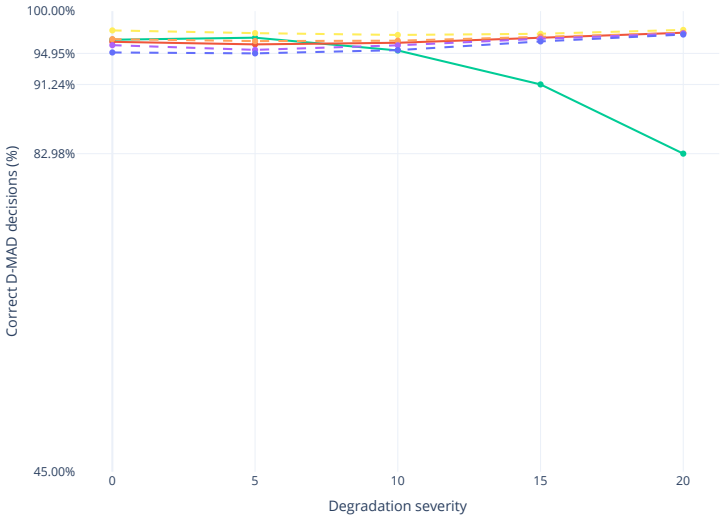
Threshold configuration model



Motion blur **without** threshold model

- Bona fide
- Mean of morphs
- Morph: FaceMorpher
- Morph: OpenCV
- Morph: UBO
- Morph: FaceFusion

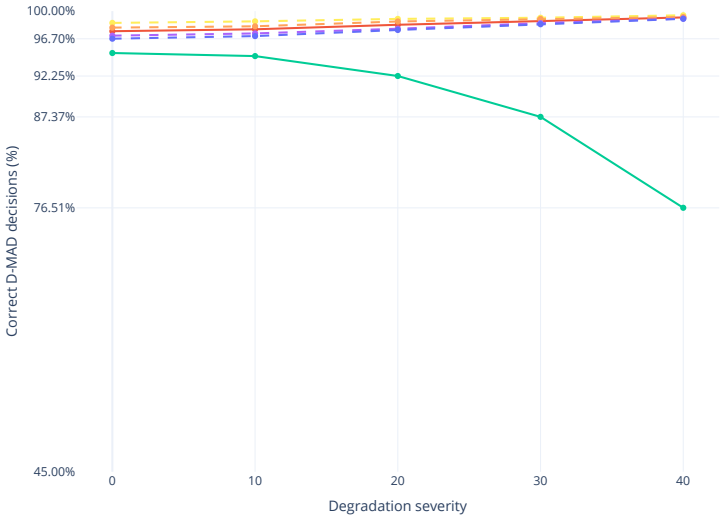
Threshold configuration model



Motion blur
with threshold model
(worst correct-% +17.26)

- Bona fide
- Mean of morphs
- Morph: FaceMorpher
- Morph: OpenCV
- Morph: UBO
- Morph: FaceFusion

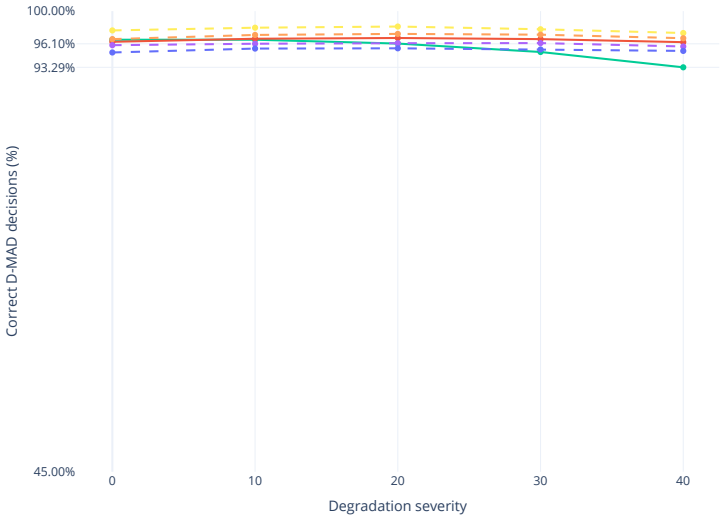
Threshold configuration model



Overexposure **without** threshold model

- Bona fide
- Mean of morphs
- Morph: FaceMorpher
- Morph: OpenCV
- Morph: UBO
- Morph: FaceFusion

Threshold configuration model



Overexposure
with threshold model
(worst correct-% +16.78)

- Bona fide
- Mean of morphs
- Morph: FaceMorpher
- Morph: OpenCV
- Morph: UBO
- Morph: FaceFusion

Conclusions

Roughly in order from the least to the most interesting conclusion:

- Unsurprisingly, both blur and exposure defects can substantially affect D-MAD.
- Underexposure had a stronger D-MAD impact than overexposure.
  - But this may not be too surprising since the underexposure effect appears to be visibly more pronounced as well, despite lower settings.
- The unified OFIQ measure (MagFace) responded to all degradations, as expected.
  - This adds to the evidence that this is a sensible unified model choice.
- Motion blur had a stronger D-MAD impact than roughly comparable Gaussian blur.
- Yet OFIQ's sharpness measure responded more to Gaussian blur than to motion blur.
  - This indicates that the next OFIQ version could benefit from an improved sharpness measure or an additional motion blur measure.
- **A simple/lightweight threshold configuration model**, that only used OFIQ's unified, sharpness, under- and over-exposure-prevention assessments as input, **was able to substantially improve the worst-case D-MAD decision percentages**.
  - This is a promising result for future D-MAD research, since this indicates that the impact of the investigated defect types can be mitigated substantially.

Thank you for your attention!

Questions?