NTNU | Norwegian University of Science and Technology

AI Supporting the Evaluation of Face Recognition Technology

Haoyu Zhang, Christoph Busch
Norwegian Biometrics Laboratory
{haoyu.zhang, christoph.busch}@ntnu.no

Overview

- 1. **Motivation**
- 2. Methodology & Database
- 3. Experiments & Results
  - 3.1 Evaluating non-mated samples
  - 3.2 Evaluating mated samples
- 4. Conclusions
- 5. Future Works

1. Motivation

- Need:
  - to develop, train, test, evaluate recognition algorithms in large scale systems.
- Problem:
  - access to data, amount and privacy
- Potential solution:
  - generate synthetic identities
- Aim of this work:
  - analyse if synthetic generated (face) samples provide similar characteristic to the bona-fide ones.
  - test quality and comparison scores distributions.

Example of synthetic data

BiGAN [Donahue2016])

StyleGAN2 [Karras2020]

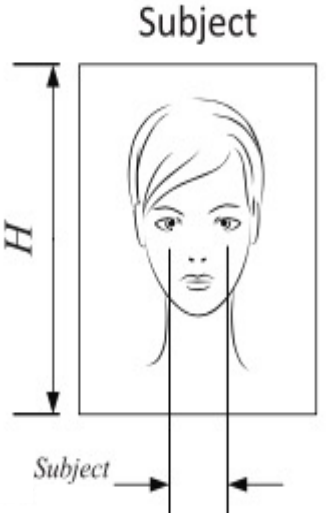
1. Motivation

- Biometric Sample Quality
  - Standard ISO/IEC 29794-5 to be aligned with both
    - ISO/IEC 19794-5:2011
    - ISO/IEC 39794-5:2019
  - https://www.iso.org/standard/81005.html
- Definitions
  - Unified quality score - FaceQnet (JRC)
  - Capture-related quality elements
  - Subject-related quality elements

a) Compliant image **b) Low contrast**

source: ISO/IEC 39794-5:2019, Annex D
https://www.iso.org/standard/72156.html

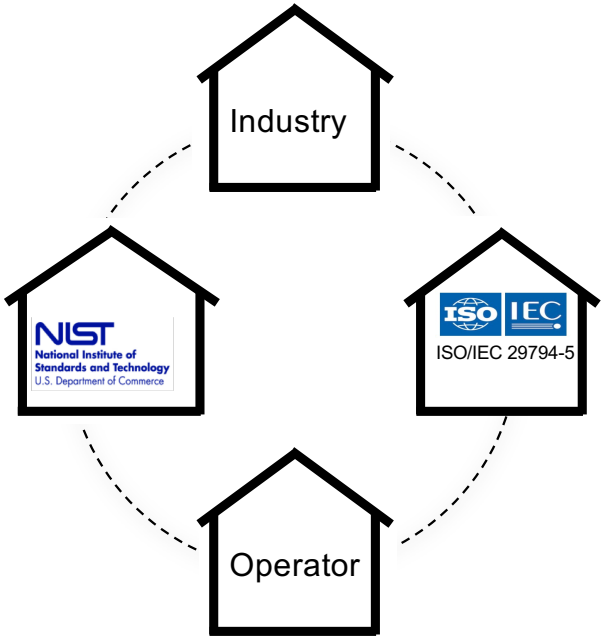
source: ISO/IEC 19794-5:2011



source: ISO/IEC 39794-5

1. Motivation

- Biometric Sample Quality
  - Standardisation process for ISO/IEC 29794-5
    https://www.iso.org/standard/81005.html
- Quality algorithm performance
  - NIST FRVT
    https://pages.nist.gov/frvt/html/frvt_quality.html
    Extension for quality elements coming soon
  - Workshop on face quality assessment
    https://eab.org/events/program/261

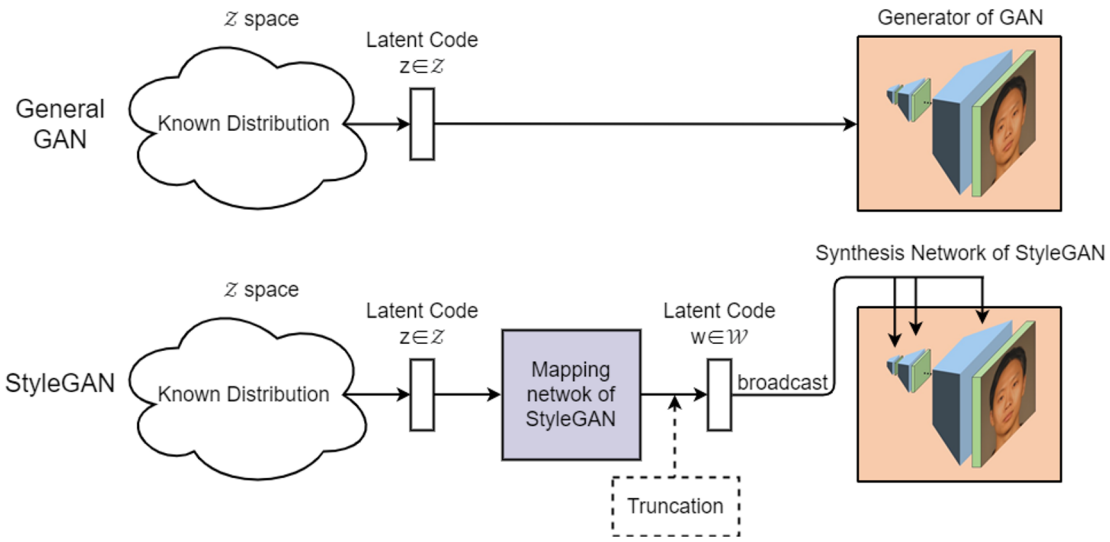


Overview

- 1. Motivation
- **2. Methodology & Database**
- 3. Experiments & Results
  - 3.1 Evaluating non-mated samples
  - 3.2 Evaluating mated samples
- 4. Conclusions
- 5. Future Works

2. Methodology & Database

- Synthetic Data Generation: Models
  - StyleGAN [Karras2019]
  - StyleGAN2 [Karras2020]



2. Methodology & Database

- Synthetic Data Generation: Truncation

$$w' = \bar{w} + \psi(w - \bar{w})$$

w': truncated latent vector
$\bar{w}$: center of the mass of the latent space
ψ: truncation factor
w: sampled latent vector

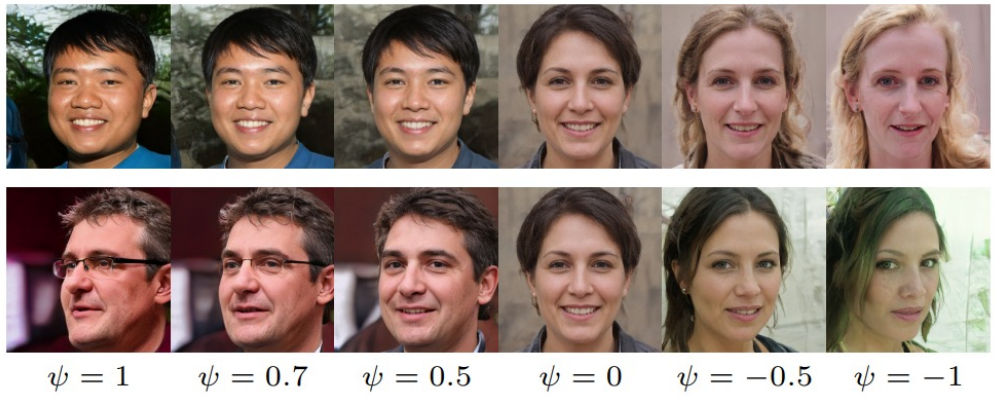


Image from [Karras2019]

2. Methodology & Database

- Mated Synthetic Data Generation:
  - Semantic editing

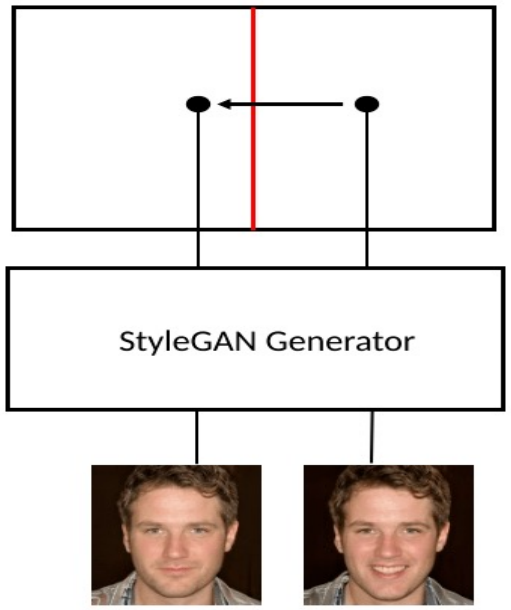
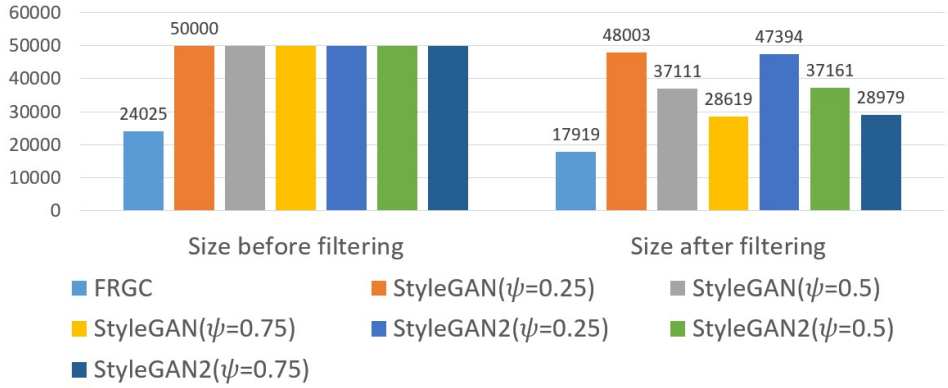


Figure: Simplified illustration of InterFaceGAN [Shen2020]: The red boundary splits the latent space into two subspaces. Latent vectors sampled on the right side of the boundary are reconstructed as smiling individuals, while those on the left side have a neutral expression. By shifting the latent vector (black dot) beyond the boundary, the same identity with neutral expression is obtained.

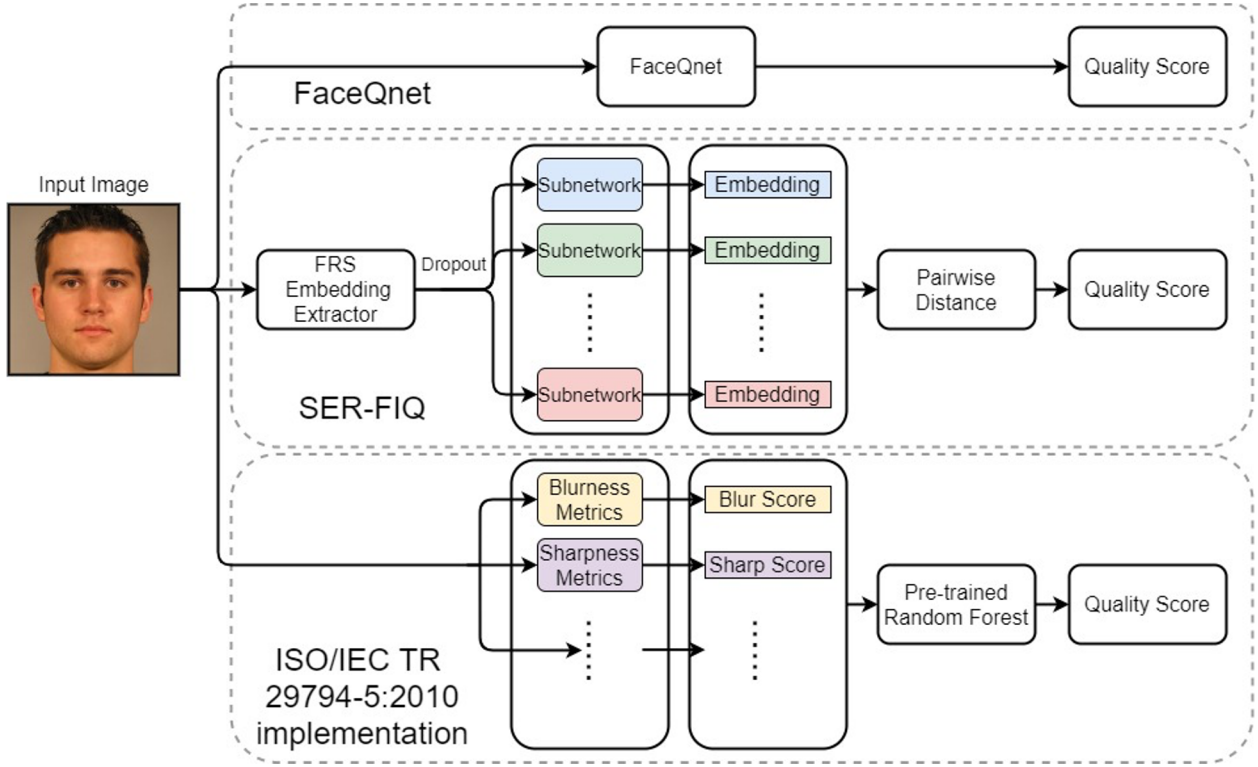
2. Methodology & Database

- Database
  - Synthetic data
    - StyleGAN
    - StyleGAN2
  - Representative bona fide data
    - Face Recognition Grand Challenge (FRGC) [Phillips2005]
  - Non-mated samples
  - Mated samples



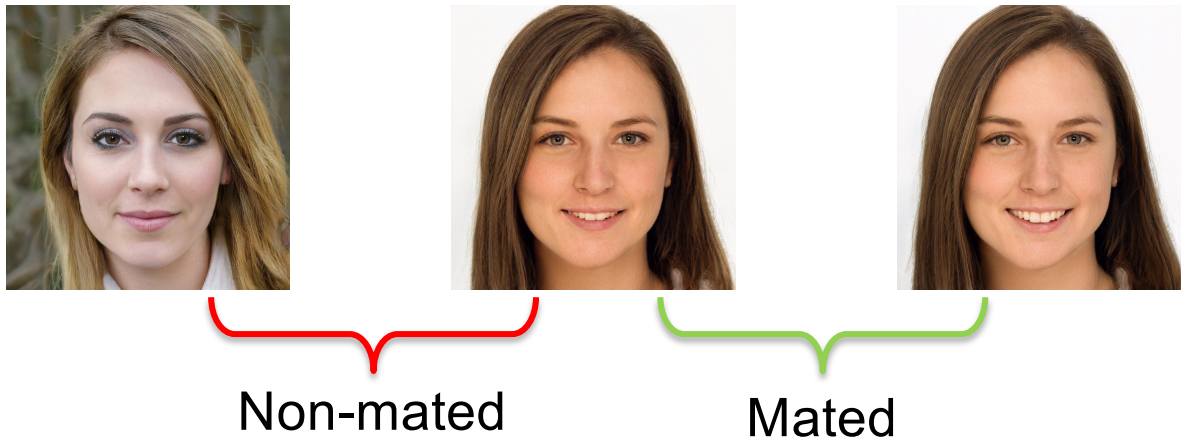
2. Methodology & Database

- Face Quality Assessment
  - FaceQnet v1 [HernándezOrtega2020]
  - SER-FIQ [Terhorst2020]
  - Implementation based on ISO/IEC TR 29794-5:2010 [ISO29794-5TR] [Wasnik2017]



2. Methodology & Database

- Evaluation Methodology
  - Distribution of quality scores
  - Distribution of comparison scores (based on ArcFace [Deng2019])

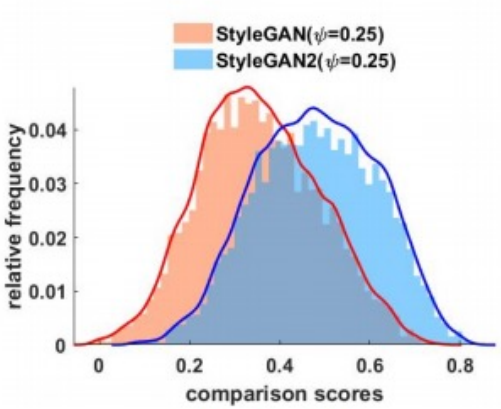


Overview

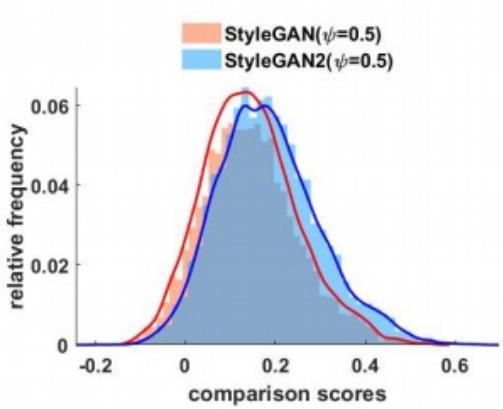
- 1. Motivation
- 2. Methodology & Database
- **3. Experiments & Results**
  - **3.1 Evaluating non-mated samples**
  - 3.2 Evaluating mated samples
- 4. Conclusions
- 5. Future Works

3.1 Evaluating non-mated samples

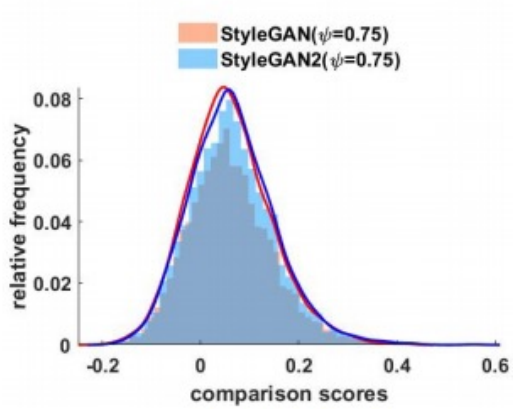
- Comparison of non-mated Distribution between StyleGAN and StyleGAN2



(a) KL-D: 0.702



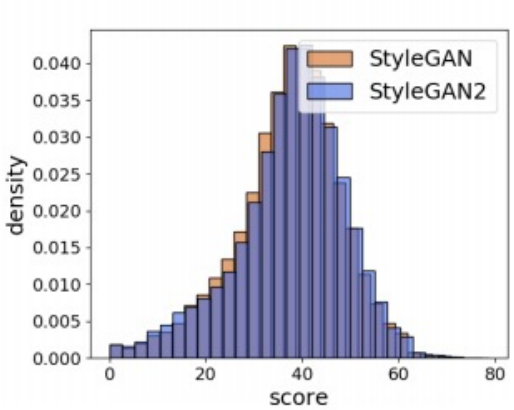
(b) KL-D: 0.107



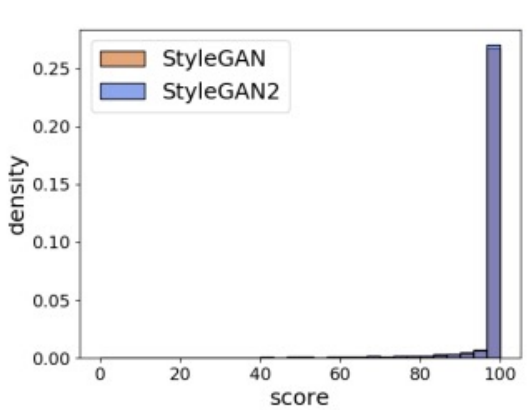
(c) KL-D: 0.007

3.1 Evaluating non-mated samples

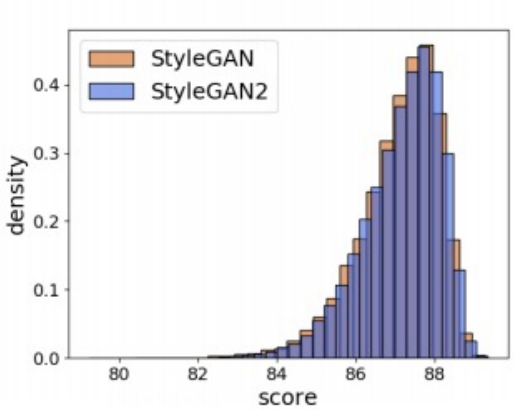
- Comparison of quality score distribution between StyleGAN and StyleGAN2



(a) KL-D: 0.008



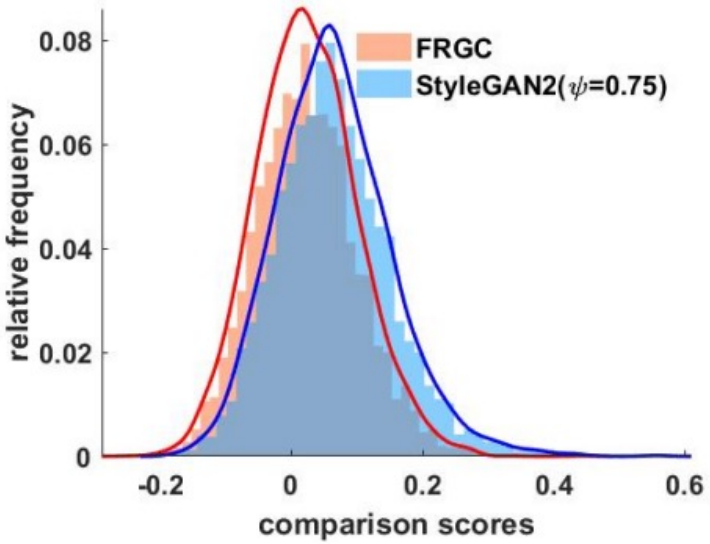
(b) KL-D: 0.003



(c) KL-D: 0.116

3.1 Evaluating non-mated samples

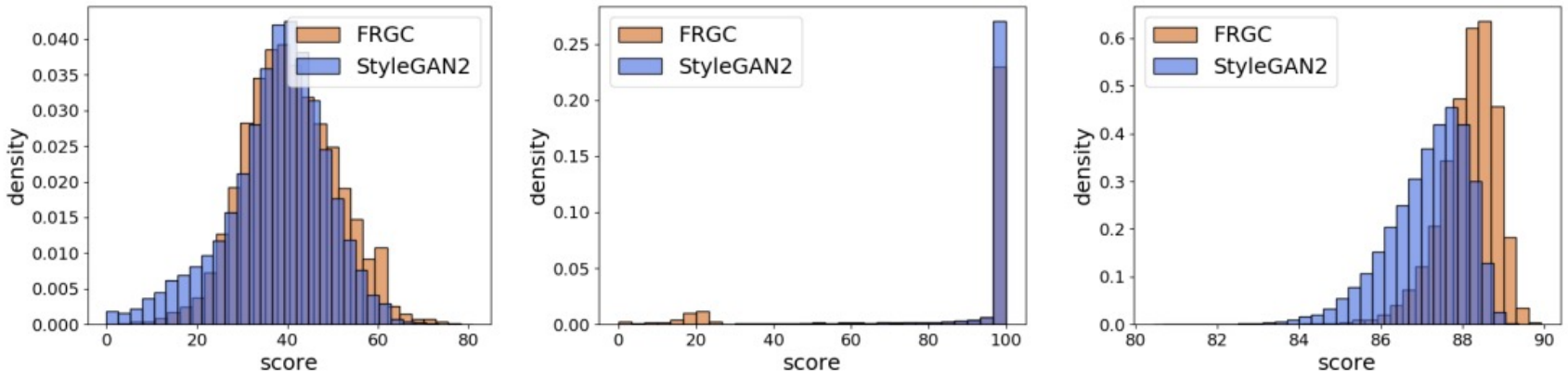
- Comparison of non-mated distribution between FRGC and StyleGAN2



KL-D: 0.184

3.1 Evaluating non-mated samples

- Comparison of quality score distribution between FRGC and StyleGAN2



(a) KL-D: 0.111 (b) KL-D: 0.444 (c) KL-D: 1.172

Fig. 6. Comparing the quality score distributions from various face quality algorithms between FRGC and StyleGAN2. (a) FaceQnet v1 (b) Random Forest Regressor (ISO/IEC TR 29794-5) (c) SER-FIQ

3.1 Evaluating non-mated samples

- Comparison of quality score distribution between FRGC and StyleGAN2 : single quality features

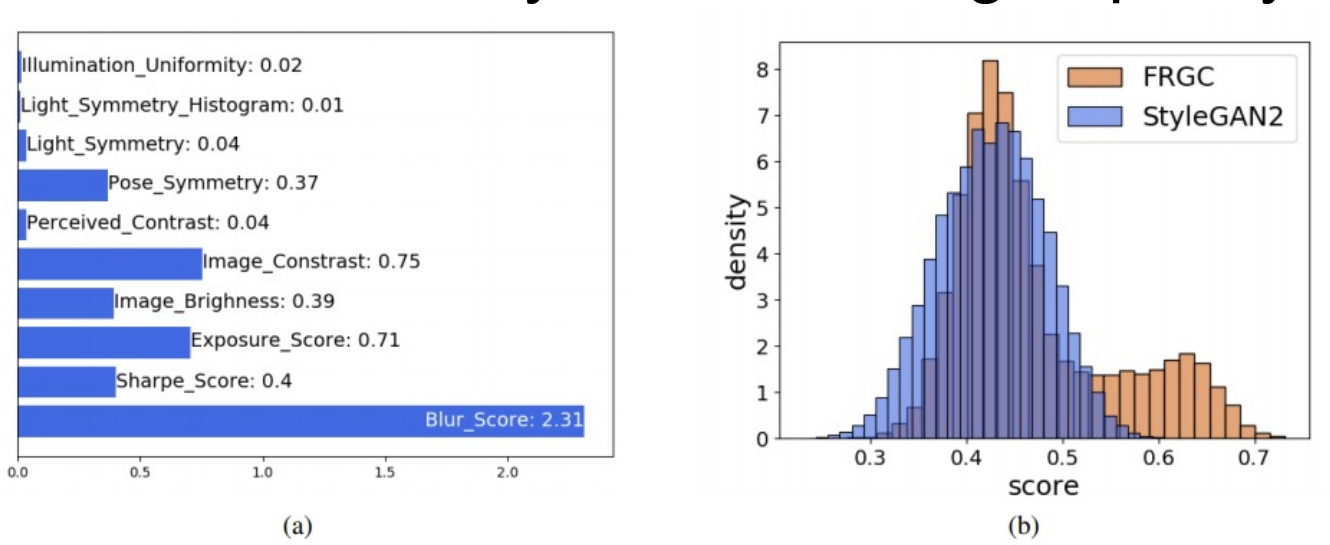


Fig. 7. Comparing the quality features from ISO/IEC TR 29794-5:2010 implementation between FRGC and StyleGAN2 ($\psi = 0.75$). (a) Kullback-Leibler Divergences between each quality features (b) Comparison of bluriness score distributions

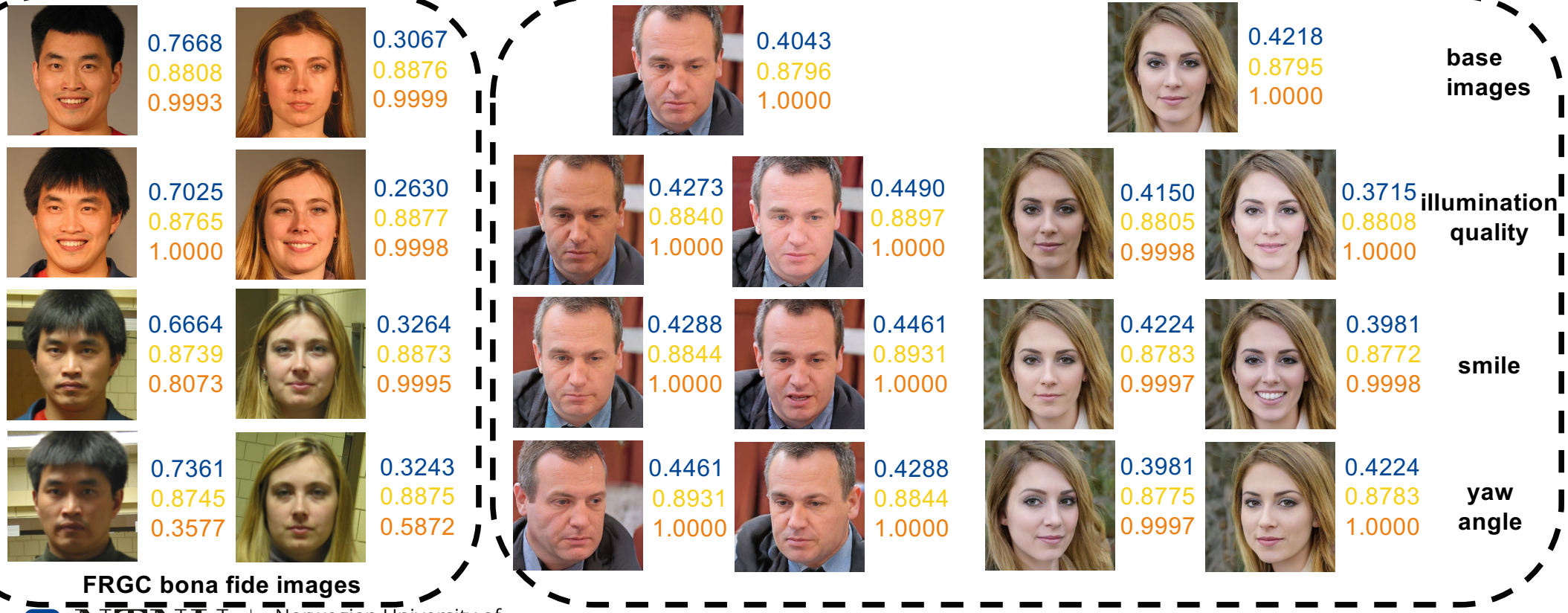
Overview

- 1. Motivation
- 2. Methodology & Database
- **3. Experiments & Results**
  - 3.1 Evaluating non-mated samples
  - **3.2 Evaluating mated samples**
- 4. Conclusions
- 5. Future Works

3.2 Evaluating mated samples

FaceQnet v1 quality score
SER-FIQ quality score
ISO/IEC TR 29794-5: 2010
implementation quality score

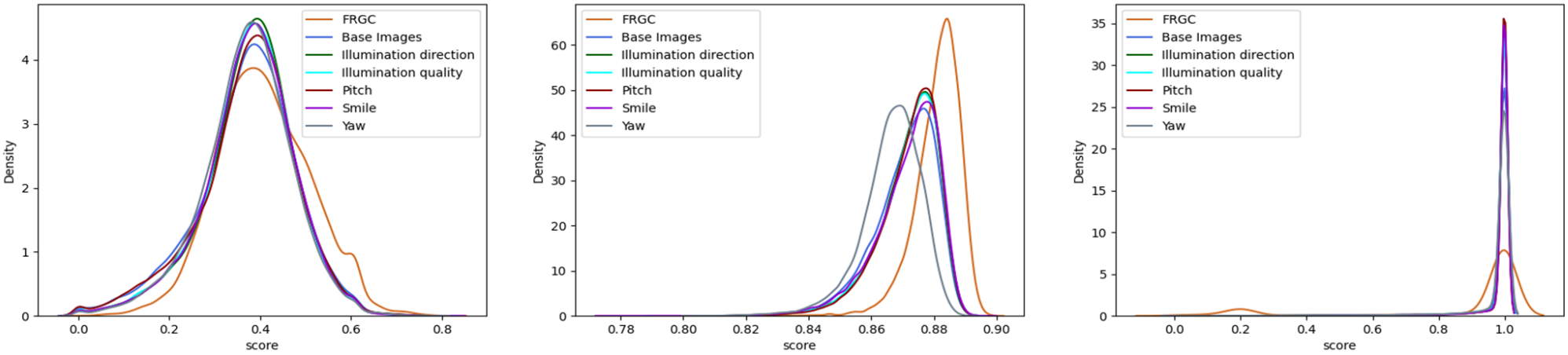
- Example of mated samples with quality scores



3.2 Evaluating mated samples

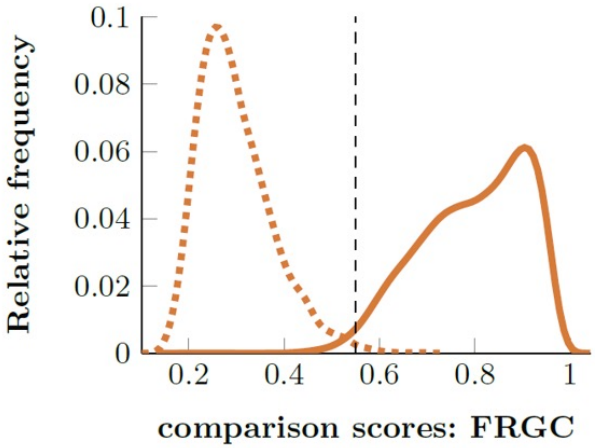
- Comparison of quality score distribution between mated samples generated

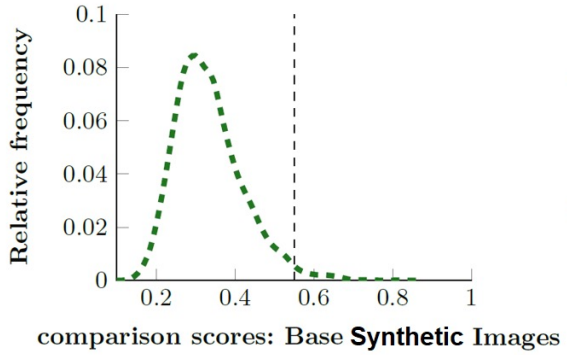
FaceQnet v1 | SER-FIQ | ISO/IEC TR 29794-5:2010

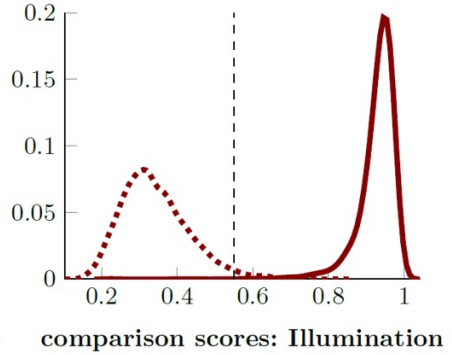


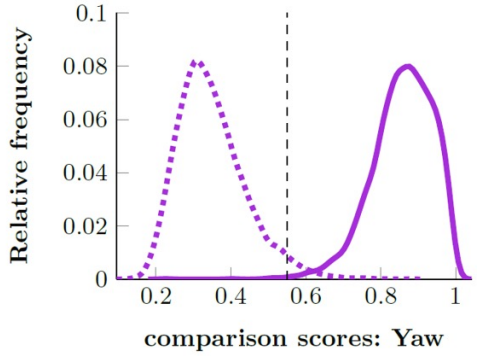
3.2 Evaluating mated samples

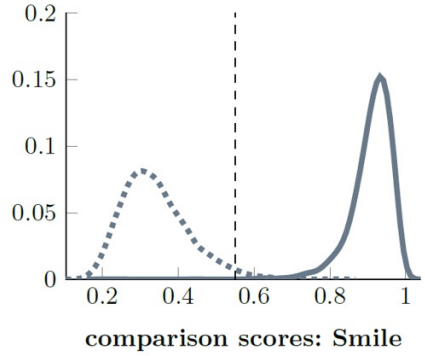
- Comparison of quality score distribution between mated samples generated







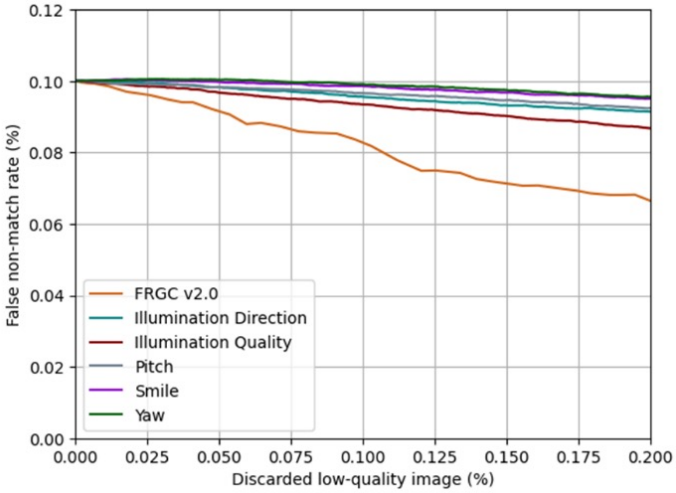




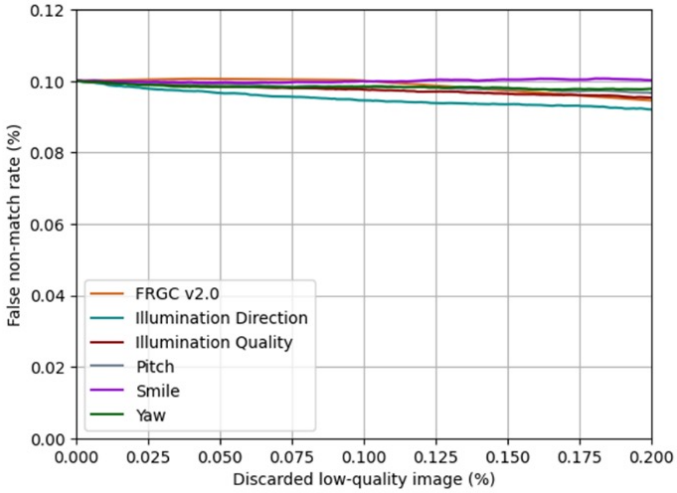
3.2 Evaluating mated samples

- Comparison of quality score distribution between mated samples generated

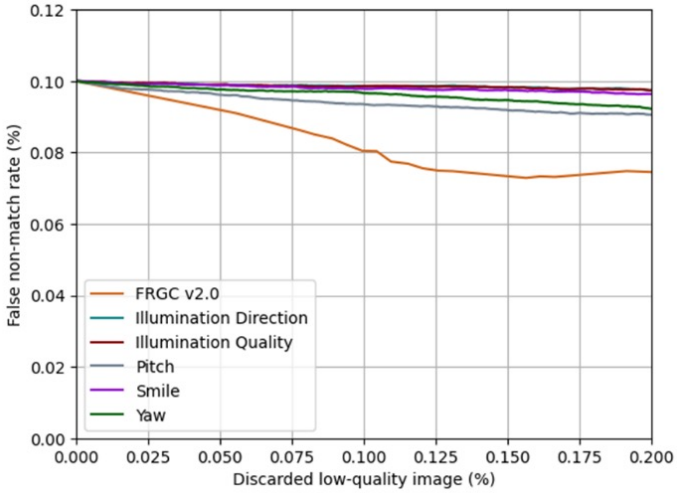
FaceQnet v1



ISO/IEC TR 29794-5:2010



SER-FIQ



Overview

- 1. Motivation
- 2. Methodology & Database
- 3. Experiments & Results
  - 3.1 Evaluating non-mated samples
  - 3.2 Evaluating mated samples
- **4. Conclusions**
  - **4.1 Conclusions for non-mated samples**
  - 4.2 Conclusions for mated samples
  - 4.3 Conclusions in general
- 5. Future Works

4.1 Conclusions for non-mated samples

- The applicability of synthetic data generated by StyleGAN and StyleGAN2 is similar.
- Only minor differences between synthetic and selected set of bona fide samples
  - Synthetic facial images are of high quality.
  - Minor differences in estimated biometric sample quality.
  - The variety of identity information is limited when the synthetic dataset is generated with a low truncation factor.

Overview

- 1. Motivation
- 2. Methodology & Database
- 3. Experiments & Results
  - 3.1 Evaluating non-mated samples
  - 3.2 Evaluating mated samples
- 4. Conclusions
  - 4.1 Conclusions for non-mated samples
  - **4.2 Conclusions for mated samples**
  - 4.3 Conclusions in general
- 5. Future Works

4.2 Conclusions for mated samples

- Analysis on quality score distributions
  - Similar for FaceQnet v1 and ISO/IEC TR 29794-5: 2010 implementation.
  - Differences in SER-FIQ due to yaw angle variation
- Analysis on the mated comparison scores
  - Mated samples can be generated without significant loss of identity information.
  - Higher intra-identity variation of the bona fide data can be observed
- Analysis on EDC curves
  - Biometric quality of bona fide samples are better predictable by FaceQnet v1 and SER-FIQ

4 Conclusions in general

- Considerable quality of synthetic data
- Remaining differences and challenges
- Encouraging as a starting stage
  - more future work and further testing remains necessary
- Not fit for purpose to **completely** assess operational systems
  - We can test workload (i.e. throughput) and workload reduction
  - For biometric performance testing we **shall** report results for synthetic data **and** non-synthetic data (ISO/IEC 19795-1:2021 Cl. 7.4.9)

5. Future Works

- Face quality assessment algorithms (FQAA)
  - Improve consistency and work on standardized FQAA algorithms
- Better approximation
  - of larger intra-identity variation of bona fide images
- Large scale tests needed
  - With large scale bona fide data sets of representative nature and large scale synthetic data sets
  - To be performed in the future to confirm these results

References

- [Zhang2021] H. Zhang, M. Grimmer, R. Raghavendra, K.Raja, C.Busch: "On the Applicability of Synthetic Data for Face Recognition", in Proceedings of 9th International Workshop on Biometrics and Forensics (IWBF), 2021
- [Donahue2016] J. Donahue, K. Philipp, and D. Trevor. "Adversarial feature learning." *arXiv preprint arXiv:1605.09782*, 2016.
- [Karras2019] T. Karras, S. Laine and T. Aila, “A style-based generator architecture for generative adversarial networks”, IEEE CVPR, 2019, pp. 4401-4410.
- [Karras2020] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. “Analyzing and improving the image quality of stylegan”, IEEE CVPR, 2020.
- [Phillips2005] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the face recognition grand challenge.", IEEE CVPR, 2005.
- [HernándezOrtega2020] J. Hernández-Ortega, J. Galbally, J. Fierrez and L. Beslay, "Biometric Quality: Review and Application to Face Recognition with FaceQnet." arXiv preprint arXiv:2006.03298, 2020.
- [Terhorst2020] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner and A. Kuijper. “SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness“, IEEE CVPR, 2020.
- [Wasnik2017] P. Wasnik, K. B. Raja, R. Ramachandra, C. Busch, “Assessing face image quality for smartphone based face recognition system”, IEEE IWBF, 2017, pp. 1-6.
- [ISO29794-5TR] ISO/IEC JTC 1/SC 37 Biometrics, ISO/IEC TR 29794-5:2010 Information technology — Biometric sample quality — Part 5: Face image data, International Organization for Standardization, 2010.
- [Deng2019] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition", IEEE CVPR, 2019.
- [Shen2020] Y. Shen, J. Gu, X. Tang and B. Zhou, “Interpreting the latent space of gans for semantic face editing”, IEEE CVPR, 2020, pp. 9243-9252

Thanks for your watching