Evaluation and Improvement of eu-LISA Synthetic Biometric Datasets

Marcel Grimmer, Christoph Busch
Norwegian Biometrics Laboratory
{marceg, christoph.busch}@ntnu.no

Outline

- Motivation & Project Overview
- Face Age Modification
- Experimental Setup
- Experimental Results
  - Biometric Quality
  - Comparison Score Analysis
- Conclusions

Outline

- **Motivation & Project Overview**
- Face Age Modification
- Experimental Setup
- Experimental Results
  - Biometric Quality
  - Comparison Score Analysis
- Conclusions

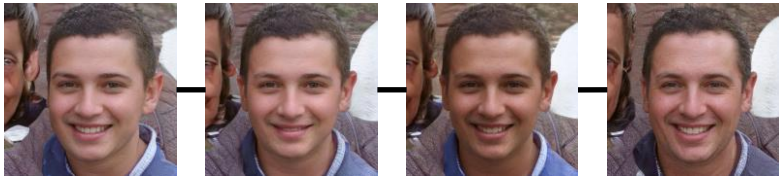
Motivation

- Need:
  - To train and 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 synthetically generated (face) samples provide similar characteristic to the bona-fide samples.
  - Evaluate quality and comparison score distributions.

Project Overview

- Evaluate synthetic non-mated face images [Zhang2021]

- Generate synthetic mated samples by editing facial attributes
  - Head Pose [Grimmer2021]
  - Facial expression [Grimmer2021]
  - Illumination [Grimmer2021]
  - **Age** [Alaluf2021]



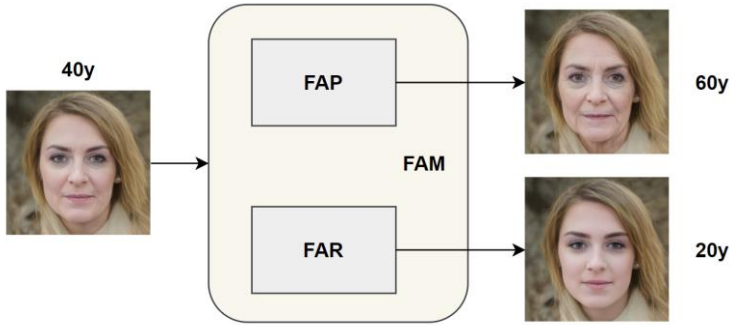
- Evaluate synthetic mated face images and compare to real data

Outline

- Motivation & Project Overview
- **Face Age Modification**
- Experimental Setup
- Experimental Results
  - Biometric Quality
  - Comparison Score Analysis
- Conclusions

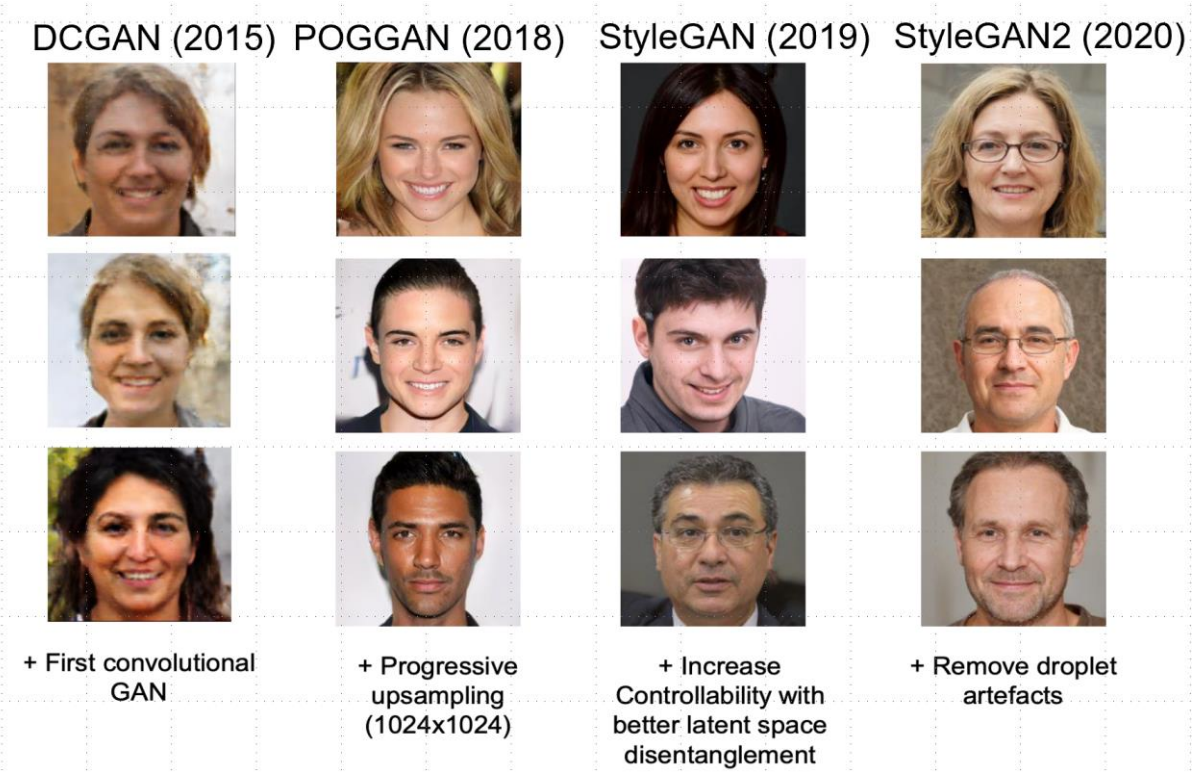
Face Age Modification

- FAM: Use generative models to predict future (FAP) or past appearance (FAR) of individuals



- Photorealism through concept of adversarial learning
  - Generator: Learns to generate realistic face images
  - Discriminator: Learns to distinguish generated face images from real face images

Example of Synthetic Data



Application in Biometrics

- Face recognition systems must be …
  - Sensitive to inter-identity variation: Twins, doppelgängers, etc..
  - Robust against intra-identity variation:
    expression, head pose, illumination, and age

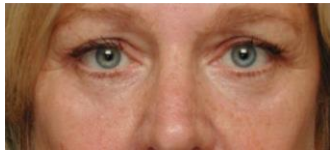
1996 2001 2006 2007 2010 2015

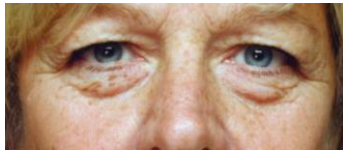
- Lack of available training samples captured over long time spans (>5y)
  - Recognition performance suffers

Application in Biometrics

- Idea: Simulate re-occuring anatomical changes in physiological characteristics with FAM algorithms

Real ageing





Synthetic ageing

- Possible Applications:
  - Increase robustness of face recognition systems through cross-age fine-tuning
  - Compensate age gap between probe and reference sample to reduce false negative identification rate (FNIR)
  - Evaluate impact of face ageing on FRSs → focus of this talk

Outline

- Motivation & Project Overview
- Face Age Modification
- **Experimental Setup**
- Experimental Results
  - Biometric Quality
  - Comparison Score Analysis
- Conclusions

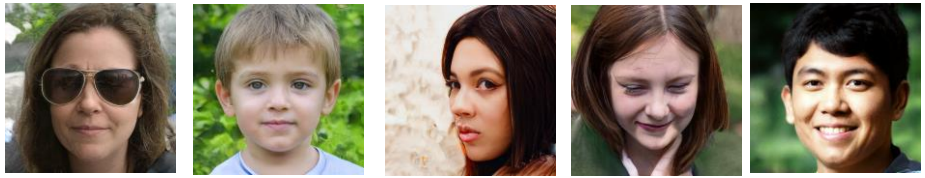
Experimental Setup

- Use StyleGAN [Karras2019] [Karras2020]
  to generate random non-mated face images (Base)
- Use SAM [Alaluf2021]
  for face ageing framework for age progression and regression (Mated Samples)

Experimental Setup

- Filter out images with
  - Inter-eye-distance < 90px
  - Age < 12 years
  - Extreme Yaw and Pitch angles
  - Poor illumination

IED Age Yaw Pitch Illumination



Dataset	# Images before filtering	# Images after filtering
FRGC v2 [Phillips2005]	24,025	17,919
StyleGAN Base	50,000	25,918
SAM (target age = 10)	25,918	18,290
SAM (target age = 20)		22,671
SAM (target age = 30)		23,253
SAM (target age = 40)		23,513
SAM (target age = 50)		22,671
SAM (target age = 60)		17,174
SAM (target age = 70)		10,028

Outline

- Motivation & Project Overview
- Face Age Modification
- Experimental Setup
- **Experimental Results**
  - Biometric Quality
  - Comparison Score Analysis
- Conclusions

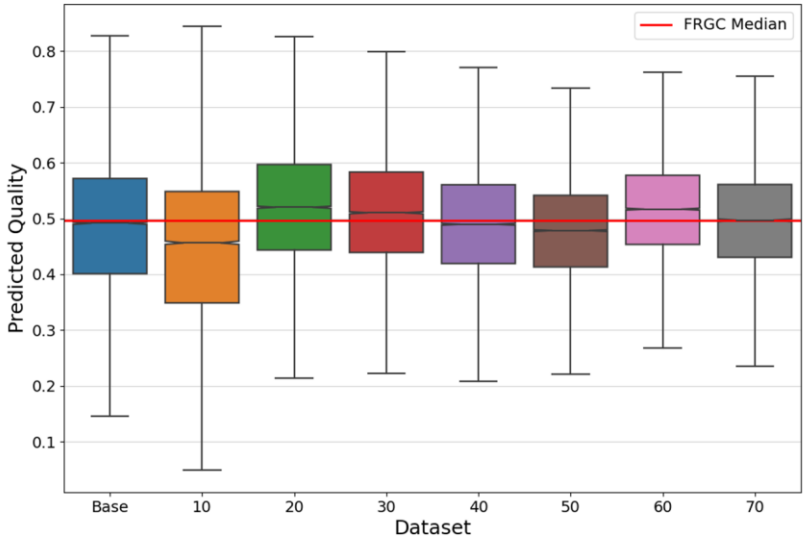
Evaluation

- Part A: Biometric Quality
  - Use Face image quality assessment algorithms (FIQAAs)
  - Predict biometric quality in numeric range [0,1]
    - 1: Perfect face recognition utility
    - 0: Worst face recognition utility
  - FIQAA: FaceQnet v1 [HernándezOrtega2020]

- Part B: Comparison Score Analysis
  - Face recognition: ArcFace [Deng2019]
  - Synthetic versus Synthetic
  - Synthetic versus Real (Short-term ageing)

Biometric Quality

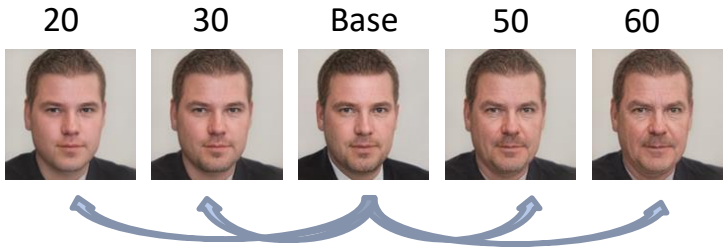
- Face Image Quality Assessment Algorithm: FaceQnet v1
- No statistical significant differences between synthetic and real datasets (95% confidence)



Evaluation

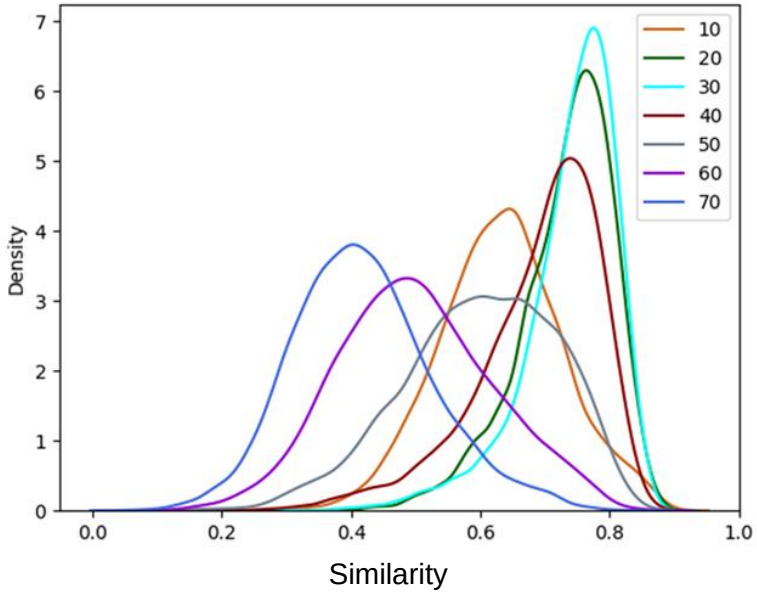
- Part A: Biometric Quality
  - Use Face image quality assessment algorithms (FIQAAs)
  - Predict biometric quality in numeric range [0,1]
    - 1: Perfect face recognition utility
    - 0: Worst face recognition utility
  - FIQAA: FaceQnet v1

- Part B: Comparison Score (CS) Analysis
  - Face recognition: ArcFace
  - Synthetic versus Synthetic
  - Synthetic versus Real (Short-term ageing)

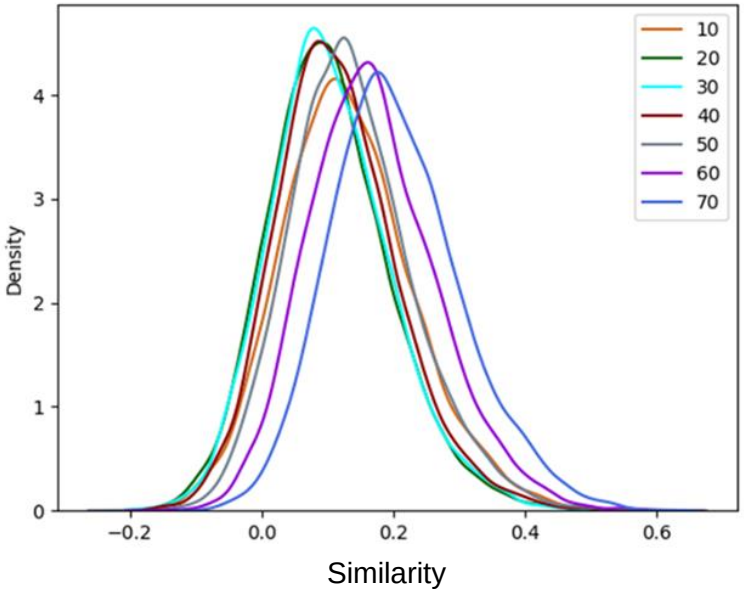


CS Analysis: Synthetic vs Synthetic

Mated Comparisons

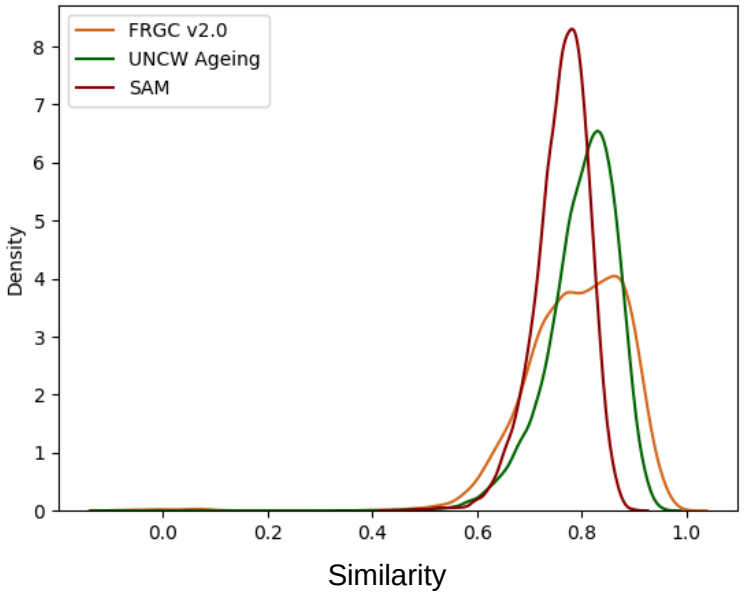


Non-Mated Comparisons

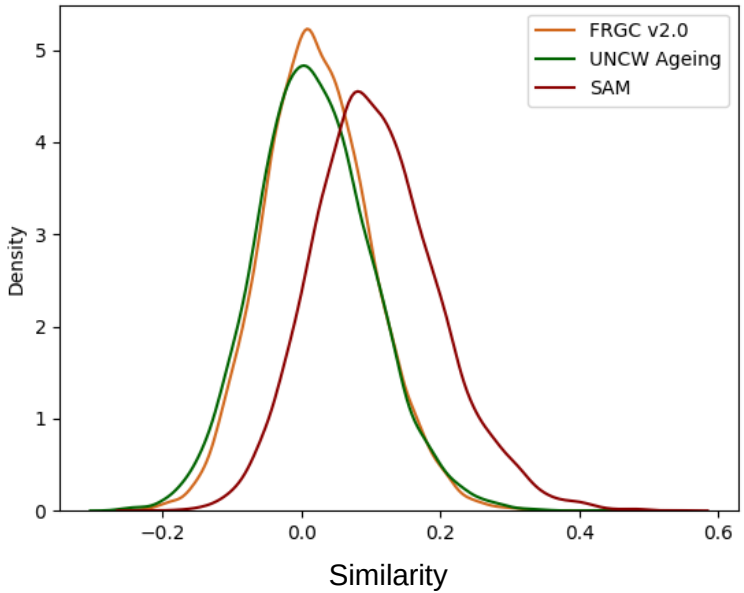


CS Analysis: Synthetic vs Real

Mated Comparison



Non-Mated Comparisons



UNCW Ageing Dataset: [Ricanek2006]

Outline

- Motivation & Project Overview
- Face Age Modification
- Experimental Setup
- Experimental Results
  - Biometric Quality
  - Comparison Score Analysis
- **Conclusions**

Conclusion

- FAM algorithms based on manipulations in the latent space
  - High visual quality with high resolution (1024x1024)
  - Accurate simulation of age progression and regression
- No statistically significant difference between synthetic and real data in terms of biometric quality (FaceQnet v1)
- Mated comparison score analysis confirms decreasing similarity with proceeding age
- Further experiments reommended to evaluate face recognition robustness to long-term age differences between reference and probe image

Conclusion

- 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)

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
- [Grimmer2021] M. Grimmer, H. Zhang, R. Ramachandra, K. Raja, and C. Busch “Generation of Non-Deterministic Synthetic Face Datasets Guided by Identity Priors”, arXiv preprint arXiv:2112.03632, 2021.
- [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.
- [Deng2019] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition", IEEE CVPR, 2019.
- [Alaluf2021] Y. Alaluf, O. Patashnik, and D. Cohen-Or, “Only a Matter of Style: Age Transformation Using a Style-Based Regression Model”, arXiv preprint arXiv:2102.02754, 2021.
- [Ricanek2006] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, pp. 341-345.

Thank you for your attention