https://doi.org/10.2352/EI.2026.38.12.GENAI-182

When Synthetic Faces Fail: Exploring LoRA Fine-Tuning Limitations for Age-Specific Face Generation

Julian Götzinger, Raphael Antonius Frick; Fraunhofer SIT | ATHENE: Darmstadt, Germany

Abstract

Facial age estimation systems require diverse training data across all age groups, yet existing datasets exhibit significant demographic biases and pose privacy concerns. We evaluate whether Low-Rank Adaptation (LoRA) fine-tuned text-to-image diffusion models can generate age-specific synthetic faces suitable for training age estimation models. We train 199 age-gender specific LoRA models on a standardized compilation of five established datasets and generate a balanced synthetic dataset of 29,850 images. Across four prediction paradigms and seven test datasets, models trained on synthetic data produce substantially higher error than real-data baselines on all regression tasks. Per-age analysis on held-out data shows uniformly high MAE (21–27 years) even for age groups with abundant training data, indicating that data imbalance is not the primary cause. Relabeling the synthetic images with an external age estimator reduces MAE by roughly half, confirming that the generated faces are visually plausible but do not depict the intended target ages. These findings indicate that standard LoRA cannot reliably encode age as a semantic attribute in diffusion model outputs.

Introduction

Automatic age estimation (AAE) systems analyze facial images to predict an individual's age. These systems are deployed in age-gated access control, detection of child sexual exploitation material (CSEM), forensic investigations, and biometric identification [1]. Their reliability is essential in these security-sensitive contexts.

AAE systems based on deep learning require large-scale, labeled facial datasets for training. Existing datasets, however, suffer from several well-documented limitations [1, 2]: legal and ethical restrictions on collecting images of minors create coverage gaps; age labels are often approximate or self-reported; children and elderly individuals are systematically underrepresented; and many datasets are drawn from narrow domains such as celebrity images or criminal justice records, limiting generalization.

Recent advances in diffusion models [3] combined with parameter-efficient fine-tuning methods such as Low-Rank Adaptation (LoRA) [4] offer a potential path forward. LoRA enables adaptation of large pre-trained models using minimal data by training small low-rank update matrices while keeping the base model frozen. This makes it attractive for generating synthetic training data with specific demographic attributes, as it requires only a small number of real images per target group and avoids distributing the original data.

LoRA has been widely adopted for style transfer in diffusion models, where it effectively captures global visual patterns. However, facial aging is characterized by localized, non-linear changes [1] that differ qualitatively from artistic style. Whether standard LoRA can learn to control age as a semantic attribute in generated faces has not been empirically tested at scale. This paper provides that investigation.

We train 199 age-gender specific LoRA models on a standardized compilation of five established face datasets, generate a balanced synthetic dataset of 29,850 images, and evaluate downstream age estimation performance across four prediction paradigms and seven test datasets. Our contributions are:

1. A large-scale empirical evaluation of LoRA-based age control involving 199 separately trained models and $\sim$30k synthetic images, one of the most extensive assessments of parameter-efficient fine-tuning for demographic attribute control to date.

2. A per-age diagnostic analysis on held-out data demonstrating that synthetic failure is uniform across all age groups, including those with abundant training data, ruling out data imbalance as the primary explanation.

3. Relabeling experiments using an external age estimator (MiVOLO [5]) confirming that the generated faces are visually plausible but semantically inaccurate for age.

Approach

Our approach consists of three stages: compiling and standardizing a real face dataset, training age-gender specific LoRA models, and generating a balanced synthetic dataset. The preprocessing and generation workflows are illustrated in Figure 1 and Figure 2, respectively.

Dataset Compilation

We compile a combined dataset (AAFMU) from five established face age datasets: AgeDB [6] (16,488 images), APPA-REAL [7] (7,591), FG-NET [8] (1,002), MORPH-II [9] (55,134), and UTKFace [10] (24,106), totaling 104,321 images. Using multiple sources is intended to provide diversity in settings, demographics, and imaging conditions, reducing the risk that LoRA models learn domain-specific artifacts (e.g., the uniform backgrounds of MORPH-II mugshots) rather than age-related features.

Preprocessing Pipeline

We apply a four-stage preprocessing pipeline to reduce variability unrelated to age:

Face Retargeting. We use LivePortrait [11] to normalize head pose to a frontal orientation. Images where no face is detected are excluded.

Background Removal. FaceParsing [12], based on BiSeNet [13], segments the facial region and replaces the back-

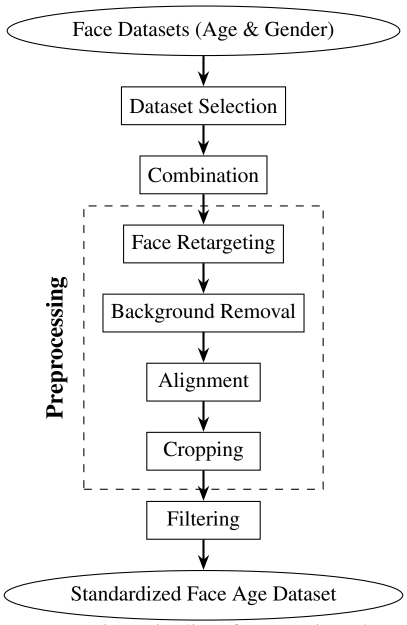


Figure 1: Preprocessing pipeline for creating the standardized face age dataset.

ground with uniform white, removing potential distractors such as clothing or scenery.

Alignment and Cropping. FFHQ-Alignment [14] standardizes facial landmark positions across all images. Images are cropped and resized to 512×512 pixels, producing consistent eye, nose, and mouth coordinates.

Quality Filtering. Multi-stage filtering using the Cleanvision toolkit [15] removes duplicates, blurry, overexposed, and grayscale images. Additional custom filtering removes images with zero or multiple detected faces.

After preprocessing, 30,995 images remain. The resulting age and gender distributions exhibit a strong skew toward males and toward the 20–40 age range (Figure 3), reflecting the composition of the source datasets.

LoRA Fine-Tuning

The preprocessed dataset is partitioned by age (0–100) and binary gender. After removing empty partitions, 199 non-empty age-gender groups remain. For each group, we fine-tune Stable Diffusion XL 1.0 [17] using LoRA (rank 16) via the Kohya-sd framework [18]. All models are trained for 60 epochs with AdamW (lr=5×10^{-4}), batch size 5, FP16 precision, and seed 42.

Each training image is captioned with a trigger word (`xlfusmifa`) followed by a gender class label (girl/boy for ages $\leq$18, woman/man for ages >18). This minimal captioning strategy is intended to encourage the model to associate the trigger word with the visual appearance of the corresponding age group, rather than relying on explicit age descriptions in the prompt.

Synthetic Dataset Generation

Using the 199 trained models, we generate 150 images per age-gender group (5 prompts × 30 images each) with the AUTOMATIC1111 interface [19]. Generation uses the DPM++ 2M Karras sampler, CFG scale 7.0, 30 sampling steps, LoRA weight 0.9, and output resolution 512×512. All prompts describe photoreal-

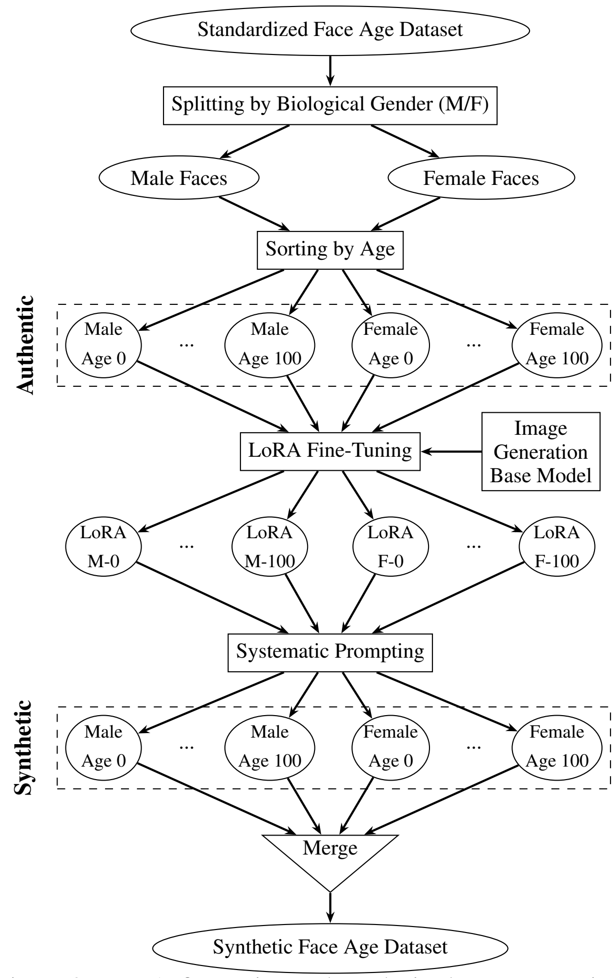


Figure 2: LoRA fine-tuning and synthetic dataset generation workflow.

istic portrait imagery. The resulting dataset contains 29,850 images with a uniform age distribution by design (Figure 5).

Experimental Setup
Age Estimation Architecture

All age estimation experiments use EfficientNet-B5 [20] from the timm library [21] as the backbone. We use the same architecture and hyperparameters for all training sets to ensure that performance differences reflect the training data rather than modeling choices. Models are trained with Adam (lr=1×10^{-4}, weight decay=1×10^{-5}), batch size 64, input resolution 256×256, and early stopping with patience 20. Data augmentation includes random horizontal flipping, rotation (±15°), and color jitter.

Prediction Paradigms

To ensure our conclusions are not specific to a single formulation, we evaluate four prediction paradigms:

1. **Exact Age Classification**: 101-class classification (ages 0–100) with cross-entropy loss. Evaluated by MAE.
2. **Categorical Classification**: 3-class (child $\leq$13, adolescent 14–17, adult $\geq$18) with cross-entropy loss. Evaluated by accuracy.

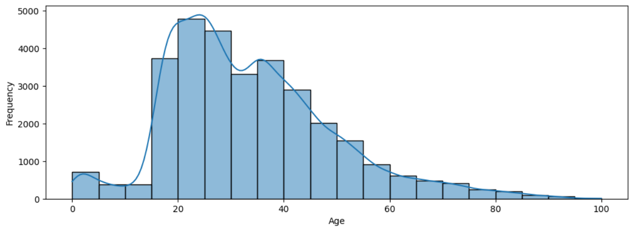
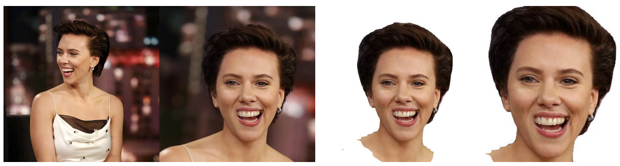


Figure 3: Age distribution of the standardized dataset used for LoRA fine-tuning (30,995 images). The distribution is heavily skewed toward the 20–40 range, with very few samples at extreme ages.



(a) Original (b) Retargeted (c) No backgr. (d) Align & crop

Figure 4: Effect of each preprocessing stage on an example image [16].

3. **Ordinal Regression**: 101 ordered binary classifiers with BCE loss [22]. Evaluated by MAE.
4. **Direct Regression**: Continuous age prediction with L1 loss. Evaluated by MAE.

Datasets and Baselines

We evaluate on seven test datasets in two categories:

Component datasets (subsets used in AAFMU compilation): AgeDB, APPA-REAL, FG-NET, MORPH-II, and UTKFace. We acknowledge potential indirect influence when evaluating on these sets, as the LoRA training data derives from them.

Held-out datasets (not used in AAFMU): CACD [23] and IMDB-WIKI [24]. These provide the most reliable generalization assessment and are the primary focus of our analysis.

We compare against three baselines: MORPH-II and UTKFace individually (the two largest component datasets) and the preprocessed AAFMU (AAFMU-std).

Results

Overall Performance

Table 1 presents performance on the held-out test sets CACD and IMDB-WIKI across all four paradigms. We focus on these two datasets as they were not used in any stage of training or LoRA fine-tuning.

Across all three regression paradigms, models trained on synthetic data produce substantially higher MAE than all real-data baselines, including single-dataset baselines (MORPH, UTKFace). For direct regression on IMDB-WIKI, the synthetic-trained model reaches 19.12 years MAE compared to 8.22 years for AAFMU-std. Categorical classification accuracy appears comparable (0.93 vs. 0.95), but with only three broad classes this metric is insensitive to the large per-year errors observed in regression.

The underperformance is consistent across paradigms and test sets, indicating a systematic issue rather than a paradigm-specific artifact.

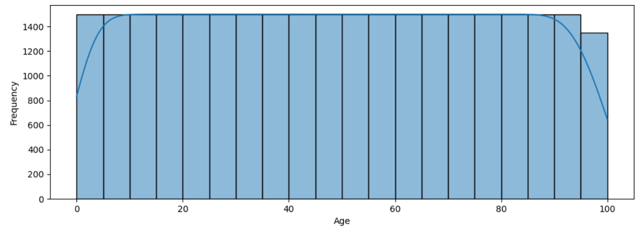


Figure 5: Age distribution of the synthetic dataset (29,850 images). The distribution is intentionally balanced at 150 images per age-gender group.

Per-Age MAE Analysis

A natural hypothesis for the synthetic data's poor performance is the severe imbalance in the LoRA training data: some age-gender groups have hundreds of training images while others have fewer than ten. If data scarcity were the primary cause, we would expect synthetic performance to be reasonable for well-represented ages (roughly 20–39) and degrade for underrepresented extremes.

To test this, we compute per-age-group MAE on the CACD test set for the direct regression paradigm (Figure 6), comparing models trained on synthetic data, MORPH-II, and APPA-REAL.

The synthetic-trained model shows uniformly high per-group MAE, ranging from 21.5 years (ages 20–29) to 26.7 years (ages 60–62). In the 20–29 range, where LoRA training data is most abundant, synthetic MAE is 21.5 years, comparable to the 26.7 years observed for the data-scarce 60–62 group. The expected correlation between training data availability and down-

Table 1: Age estimation performance on held-out test sets. MAE in years (lower is better); Acc as proportion (higher is better).

Training Data	Test Dataset			
	CACD		IMDB-WIKI	
	MAE	Acc	MAE	Acc
Exact Age Classification				
Synthetic	17.08	–	19.45	–
MORPH	12.70	–	12.35	–
UTKFace	14.58	–	13.76	–
AAFMU-std	**10.04**	–	**10.90**	–
Categorical Classification				
Synthetic	–	0.93	–	0.93
MORPH	–	**0.97**	–	**0.96**
UTKFace	–	0.79	–	0.87
AAFMU-std	–	0.95	–	0.94
Ordinal Regression				
Synthetic	15.16	–	19.98	–
MORPH	10.76	–	10.88	–
UTKFace	14.22	–	14.22	–
AAFMU-std	**8.41**	–	**9.23**	–
Direct Regression				
Synthetic	15.40	–	19.12	–
MORPH	11.27	–	11.11	–
UTKFace	13.92	–	12.90	–
AAFMU-std	**8.15**	–	**8.22**	–

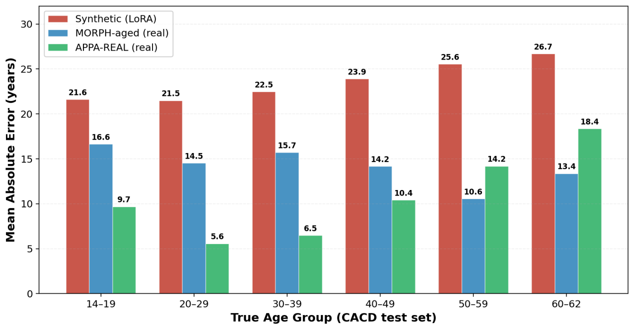


Figure 6: Per-age-group MAE on CACD (direct regression). Synthetic MAE is uniformly high across all groups, including the data-rich 20–39 range. Real-data baselines show the expected correlation between training data availability and performance.

stream performance is absent for the synthetic condition.

The real-data baselines behave differently. APPA-REAL achieves 5.6 years MAE in the 20–29 range, where its training distribution peaks, and degrades to 18.4 years for the 60–62 group. The MORPH-trained model shows a similar pattern. This contrast indicates that the synthetic data's failure is not primarily attributable to data imbalance, but to a more fundamental issue in how LoRA encodes age.

Generalization Gap

The per-age analysis also reveals a discrepancy between per-group and overall performance: the synthetic-trained model produces MAE above 21 years for every age group in Figure 6, while the overall MAE in Table 1 is 15.40 years. This difference arises because the CACD test set is concentrated in the 20–50 age range, and the overall MAE is weighted by the number of test samples per group. Additionally, the synthetic-trained model achieves substantially lower error when evaluated on a held-out split of the synthetic data itself (approximately 7 years) compared to its performance on real data. This gap suggests that the model learns features specific to the synthetic domain rather than transferable age representations.

Relabeling with an External Age Estimator

To disentangle the quality of the generated images from the accuracy of their age labels, we relabel the synthetic dataset using MiVOLO [5], an external age estimator not involved in any training. Table 2 shows the effect on the exact age classification paradigm.

Relabeling reduces MAE by roughly half. This suggests that the synthetic images depict faces with recognizable age characteristics, as MiVOLO can assign consistent ages to them, but the ages depicted do not correspond to the LoRA target ages. A caveat applies: the relabeled ages reflect MiVOLO's own biases and error characteristics rather than ground-truth ages. The improvement therefore indicates consistency between two learned models rather than absolute label correctness. It does, however, confirm that the LoRA generation process does not reliably control which age is depicted.

Table 2: Effect of relabeling synthetic images with MiVOLO on exact age classification MAE (years).

Label source	CACD	IMDB-WIKI
LoRA target age	17.08	19.45
MiVOLO estimate	**8.34**	**10.13**

Discussion

Failure Analysis

The results presented above point to a structural mismatch between LoRA's adaptation mechanism and the requirements of age control. We identify three properties of facial aging that are poorly served by standard low-rank adaptation:

Non-linear progression. Facial aging proceeds through qualitatively different phases: craniofacial growth in childhood, secondary sexual development in adolescence, and skin texture changes in later adulthood [1]. Each phase involves distinct biological processes and affects different facial structures. A single low-rank update, trained on one age group, must capture the relevant phase without interference from others. Our per-age analysis suggests this does not occur: performance is equally poor for all phases.

Spatial localization. Aging cues are concentrated in specific facial regions, including periorbital wrinkles, nasolabial folds, and changes in skin texture [1]. LoRA, as applied to diffusion model attention layers, modifies feature representations without explicit spatial selectivity. This contrasts with artistic style transfer, where the target pattern (e.g., brushstroke texture, color palette) tends to be spatially distributed across the image. Approaches that introduce spatial structure into the adaptation, such as applying different LoRA weights to different network blocks, or using conditioning mechanisms like ControlNet [25] or IP-Adapter [26] for explicit spatial guidance, could address this limitation, though they remain untested for age control.

High intra-class variance. People of the same chronological age can appear very different due to genetics, lifestyle, and environmental factors. Representing this variance within a single low-rank update of rank 16 is a demanding task, particularly for age groups with few training samples. While the per-age analysis shows that data scarcity is not the primary failure mode, it remains a compounding factor: real-data baselines also degrade at extreme ages where training data is sparse.

Visual Realism and Semantic Accuracy

As illustrated in Figure 7, the synthetic faces appear visually realistic across all target ages, yet many age groups produce faces with similar apparent age. The relabeling experiment confirms this quantitatively: the synthetic faces are realistic enough for an external model to assign consistent ages to them, but the assigned ages do not match the intended LoRA targets. This decoupling of visual quality from attribute accuracy has practical implications: standard generative quality metrics such as FID [27] and IS [28] would likely indicate acceptable image quality while failing to flag the age-label mismatch. Task-level evaluation, as performed in this work, is necessary when synthetic data is intended for training attribute-specific classifiers.

Related Work

Age estimation. Deep learning approaches have substantially improved age estimation accuracy over earlier methods based on hand-crafted features [29]. Current methods use classification [24, 30], ordinal regression [22], or direct regression

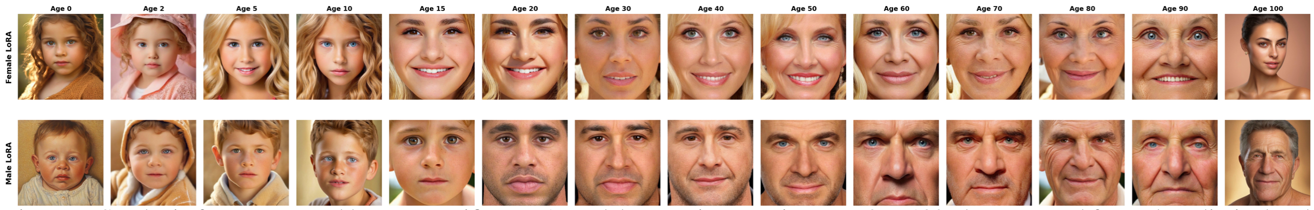


Figure 7: Synthetic faces generated by age-specific LoRA models. Despite targeting ages 0 to 100, the generated faces show limited and inconsistent age variation across target groups. Visual realism is high, but age-specific features do not reliably correspond to the intended target age.

formulations, often with specialized architectures. Demographic bias remains a persistent challenge [1].

Synthetic data for face analysis. GAN-based approaches have been used for data augmentation in face analysis [31] and for mitigating demographic imbalance [32]. Most such work targets identity-related tasks (face recognition, verification) rather than fine-grained attribute control. Age progression and regression methods [33] transform specific individuals across ages, which differs from our goal of generating diverse training samples at a target age.

Diffusion models and LoRA. Diffusion models [3, 34] achieve state-of-the-art image synthesis quality. LoRA [4] enables parameter-efficient adaptation and has been widely adopted for style transfer and subject-driven generation. Our work evaluates its applicability to a different task, demographic attribute control, and finds it insufficient in its standard configuration.

Conclusion

We evaluated whether standard LoRA fine-tuning of diffusion models can produce synthetic faces with reliable age attributes for training age estimation systems. Across 199 age-gender specific models, four prediction paradigms, and seven test datasets, the answer is negative: models trained on synthetic data consistently underperform real-data baselines by a wide margin on all regression tasks. Per-age analysis shows this underperformance is uniform across age groups, including those well-represented in the training data, pointing to a structural limitation of the adaptation mechanism rather than data scarcity. Relabeling experiments confirm that the generated faces are visually plausible but do not depict the intended target ages.

These results establish empirical bounds on the applicability of standard LoRA for demographic attribute control in generative models and highlight the gap between visual realism and semantic attribute accuracy in current synthetic data pipelines.

Limitations

We used a single base model (SDXL 1.0) and a single adaptation method (LoRA, rank 16) without extensive hyperparameter tuning. Different configurations (higher ranks, alternative learning rates, or other diffusion architectures) may produce different outcomes. Our demographic analysis is limited to binary gender; ethnicity and other variables were not considered. The evaluation architecture (EfficientNet-B5) is not state-of-the-art for age estimation, though our analysis depends on relative rather than absolute performance.

Future Work

The failure analysis suggests several directions. Adaptation strategies with spatial structure, such as applying LoRA at different ranks or to different blocks depending on the layer's role in generating local versus global features, could better target age-related regions. Conditioning approaches like ControlNet [25] or IP-Adapter [26] could provide explicit spatial guidance. Hybrid protocols combining real and synthetic data may be more practical than fully synthetic pipelines. Systematic hyperparameter search over LoRA rank, learning rate, and training duration is also warranted.

Acknowledgments

This research work has been funded by BMBF and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Age estimation via face images: A survey," *EURASIP Journal on Image and Video Processing*, vol. 2018, p. 42, Dec. 2018.

[2] K. ELKarazle, V. Raman, and P. Then, "Facial Age Estimation Using Machine Learning Techniques: An Overview," *Big Data and Cognitive Computing*, vol. 6, p. 128, Oct. 2022.

[3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022.

[4] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *CoRR*, vol. abs/2106.09685, 2021.

[5] M. Kuprashevich and I. Tolstykh, "Mivolo: Multi-input transformer for age and gender estimation," in *International conference on analysis of images, social networks and texts*, pp. 212–226, Springer, 2023.

[6] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: The First Manually Collected, In-the-Wild Age Database," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Honolulu, HI, USA), pp. 1997–2005, IEEE, July 2017.

[7] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, and R. Rothe, "Apparent and real age estimation in still images with deep residual regressors on appa-real database," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 87–94, 2017.

[8] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 442–455, Apr. 2002.

[9] K. Ricanek and T. Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, (Southampton, UK), pp. 341–345, IEEE, 2006.

[10] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," Mar. 2017.

[11] J. Guo, D. Zhang, X. Liu, Z. Zhong, Y. Zhang, P. Wan, and D. Zhang, "LivePortrait: Efficient Portrait Animation with Stitching and Retargeting Control," July 2024.

[12] Y. Valikhujaev, "Face-parsing: Bisenet implementation for face segmentation." `https://github.com/yakhyo/face-parsing`, 2024.

[13] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation," 2018.

[14] Happy-Jihye, "FFHQ-alignment: Facial Alignment Tool for FFHQ Dataset." `https://github.com/happy-jihye/FFHQ-alignment`, 2023.

[15] Cleanlab Inc., "Cleanvision: Data-centric computer vision tools." GitHub, 2024.

[16] "Jimmy Kimmel Live!," 2018. Episode aired 2018, ABC Television. Still image used to demonstrate preprocessing pipeline.

[17] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," July 2023.

[18] Kohya, "Sd-scripts: Training scripts for stable diffusion." GitHub, 2024.

[19] AUTOMATIC1111, "Stable Diffusion WebUI." `https://github.com/AUTOMATIC1111/stable-diffusion-webui`, 2022.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sept. 2020.

[21] R. Wightman, "PyTorch Image Models (timm)," 2019.

[22] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal Regression with Multiple Output CNN for Age Estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4920–4928, June 2016.

[23] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), vol. 8694, pp. 768–783, Cham: Springer International Publishing, 2014.

[24] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 252–257, 2015.

[25] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," 2023.

[26] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang, "IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models," *arXiv preprint arXiv:2308.06721*, 2023.

[27] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, G. Klambauer, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a nash equilibrium," *CoRR*, vol. abs/1706.08500, 2017.

[28] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, eds.), vol. 29, Curran Associates, Inc., 2016.

[29] G. Guo, Guowang Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (Miami, FL), pp. 112–119, IEEE, June 2009.

[30] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42, 2015.

[31] A. Antoniou, A. Storkey, and H. Edwards, "Data Augmentation Generative Adversarial Networks," Mar. 2018.

[32] M. Georgopoulos, J. Oldfield, M. A. Nicolaou, Y. Panagakis, and M. Pantic, "Enhancing Facial Data Diversity with Style-based Face Aging," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Seattle, WA, USA), pp. 66–74, IEEE, June 2020.

[33] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning Face Age Progression: A Pyramid Architecture of GANs," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (Salt Lake City, UT), pp. 31–39, IEEE, June 2018.

[34] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *CoRR*, vol. abs/2006.11239, 2020.

Author Biography

Julian Götzinger is a research associate at Fraunhofer SIT / ATHENE Center in Darmstadt. He obtained his Master's degrees in Computer Science and IT Security from TU Darmstadt in 2024. In the "Media Security & IT Forensics" department, he researches deep learning-based face verification methods, with a particular focus on applications for the protection of minors. His broader research interests include fairness and robustness in biometric systems, explainability and interpretability of deep learning models, as well as synthetic data generation.

Raphael Antonius Frick is a researcher at the Fraunhofer Institute for Secure Information Technology, where he has been working for several years in the "Media Security & IT Forensics" department, and is additionally pursuing his PhD at TU Darmstadt. His focus lies on the development of new methods for detecting AI-generated and manipulated images and videos, particularly for uncovering digital fraud schemes such as CEO fraud or identity abuse. He also investigates how image, text, and contextual information can be combined to detect disinformation in social networks at an early stage and to better understand its spread.