A Unified Framework for Age-Invariant Face Recognition Using GANs and Vision Transformers

Pratik Chakraborty
CSE (AI&ML)
MIT Manipal
Roll Number: 57
Reg No: 220962350
Phone Number: 9972848989
Email: pratikchakra18@gmail.com

Shaik Nurul Ameen
CSE (AI&ML)
MIT Manipal
Roll Number: 51
Reg No: 220962320
Phone Number: 9989944624
Email: ameenshaik8800@gmail.com

Sriram Sunderrajan
CSE (AI&ML)
MIT Manipal
Roll Number: 77
Reg No: : 220962444
Phone Number: 7200412107
Email: write2sriram7@gmail.com

***Abstract*—Age-invariant face recognition (AIFR) remains a persistent challenge in computer vision due to significant facial changes over time. In this work, we propose a multi-stage deep learning framework that integrates CycleGAN-based age transformation, contrastive and triplet loss learning, and multiple backbone architectures including ResNet, EfficientNet, and Vision Transformers (ViT). Our approach leverages synthetic aging to enhance data diversity and applies Siamese networks and MLP classifiers for identity similarity estimation. Experiments on UTKFace, MORPH, and FG-NET datasets demonstrate the effectiveness of our framework in learning robust, age-invariant facial representations. The results highlight the complementary strengths of transformer-based attention and CNN-based embeddings for generalization across age progression scenarios.**

***Index Terms*—Age-invariant face recognition, deep learning, CycleGAN, contrastive loss, triplet loss, Vision Transformer, Siamese networks, synthetic aging, facial identity recognition.**

I. Introduction

A key area of study in computer vision is age-invariant facial recognition (AIFR), which has grown in importance due to the increasing need for reliable biometric systems in applications like security, surveillance, and authentication. The difficulties presented by large age differences, which can change facial geometry and texture with time and make correct recognition challenging, are addressed by AIFR in contrast to standard face recognition. In order to address these obstacles, deep learning approaches have evolved, allowing systems to perform cross-age face recognition tasks with greater accuracy and dependability.

The use of Generative Adversarial Networks (GANs), which have demonstrated remarkable efficacy in producing photorealistic age-progressed or age-regressed photos while maintaining individual identity, is one of the major developments in AIFR. The resilience of feature embeddings has been further improved by methods such as contrastive learning, which provide age-invariant identity representations. Furthermore, the use of convolutional neural networks (CNNs) like VGG and ResNet in conjunction with cutting-edge techniques like transfer learning and self-attention mechanisms has greatly increased recognition accuracy across a range of age groups. This review of the literature offers a thorough analysis of the most recent approaches, stressing both their advantages and disadvantages in dealing with the complexity of AIFR.

As the field advances, the development of high-quality datasets and the design of innovative architectures remain crucial to bridging existing research gaps. The papers included in this survey demonstrate how cutting-edge methodologies are shaping the future of AIFR, offering insights into both the challenges and potential solutions for achieving age-invariant recognition in real-world scenarios.

II. Literature Survey

Age invariant face recognition is an essential and growing field in the field of Computer Vision. There have been significant advancements in this domain due to advancements in Deep Learning techniques. Some of the literature used state of the art techniques like Generative Adversarial Networks(GANs) to simulate realistic age and contrastive learning to create robust identity embeddings. Additionally,feature extraction techniques use CNNs like VGG and ResNet to improve the recognition accuracy across various ages. The 10 papers presented in this literature survey explore the latest developments to age invariant face recognition.

A. GAN-Based Approaches for Age Simulation

1) **Deep Face Age Progression: A Survey:** This survey paper gives a comprehensive review of the various methods used in face age progression, focusing on deep learning techniques. It mentions the usage of GANs for generating photorealistic age progressed faces while maintaining identity preservation. The paper also mentions the importance of having high quality datasets for effective evaluation and training [1].
2) **Learning Face Age Progression: A Pyramid Architecture of GANs:** This paper introduces a pyramid based GAN architecture. The model refines low resolution features to high resolution images, ensuring photorealism and consistent identity preservation. The model was able to capture fine grained details across age transformations. It surpassed traditional GANs in performance of generating realistic age progressed images [2].

3) **Lifespan Age Transformation Synthesis:** This paper introduces a lifespan age transformation method. This method combines neural rendering with 3D face modeling to synthesize facial images across a wide age range. This approach was able to effectively capture age related changes in both facial geometry and texture. The paper emphasized the importance of balancing photorealism and identity preservation in age transformation tasks [3].

B. Disentangled Representation Learning

1) **Look Across Elapse: Disentangled Representation Learning and Photorealistic Cross-Age Face Synthesis for Age-Invariant Face Recognition:** This paper introduces a disentangled representation learning framework and a photorealistic face synthesis module. The disentangled framework is used to separate age specific features from identity specific features to ensure robust cross age recognition. The photorealistic face synthesis module is used to generate consistent images across age groups. The model had state of the art performances on benchmark datasets for age invariant face recognition [4].
2) **Decorrelated Adversarial Learning for Age-Invariant Face Recognition:** This paper uses a decorrelated adversarial learning framework to minimize the influence of age on identity recognition. The model uses adversarial training to learn embeddings that are invariant to age while retaining discriminative identity features. This approach demonstrated state of the art performance on multiple benchmark datasets [10].

C. Contrastive Learning for Robust Embeddings

1) **Cross-Age Contrastive Learning for Age-Invariant Face Recognition:** This paper introduces a contrastive learning framework particularly for cross age face recognition. The model is trained to learn embeddings by maximizing the similarity between faces of the same identity at different ages while minimizing the similarity with different identities [5].
2) **Age-Invariant Face Recognition: A Survey on Facial Aging Databases, Techniques, and Effect of Aging:** This survey paper reviews facial aging databases, existing techniques, and the impact of aging on face recognition performance. It mentions the importance of contrastive learning in generating robust embeddings capable of handling age related variations. The paper also examines the challenges posed by aging on face recognition systems [7].

D. Transfer Learning and Dataset-Based Approaches

1) **Transfer Learning Anchored Cross-Age Face Recognition:** The paper leverages transfer learning to address the challenges of cross age face recognition. Pre-trained models were fine-tuned on age specific datasets, enabling the model to generalize well across diverse age groups. This approach helps in mitigating dataset biases and improving performance in unseen data [6].
2) **Age-Invariant Face Recognition by Multi-Feature Fusion and Decomposition with Self-Attention:** This paper introduces a self-attention-based framework that combines global and local features for age invariant face recognition. The use of self-attention enhances feature alignment across age groups, resulting in improved recognition accuracy [8].
3) **Age-Invariant Face Recognition: A Comprehensive Study of Methods, Datasets, and New Algorithm for Immigration Control:** This paper focuses on age invariant face recognition tailored for immigration control applications. It introduces a new method that combines age simulation with feature disentanglement to handle large age gaps. The method demonstrates significant improvements in recognition accuracy under practical constraints, making it highly suitable for real-world applications [9].

This literature survey highlights significant advancements in age-invariant face recognition, showcasing diverse approaches such as GAN-based age simulation, contrastive learning for robust embedding generation, and self-attention mechanisms for feature alignment. Techniques like disentangled representation learning and multi-scale GAN architectures have successfully preserved identity consistency while modeling realistic age transformations. Additionally, the use of pre-trained models and transfer learning has enhanced the scalability of age-invariant recognition systems. These papers collectively demonstrate the evolving capabilities of deep learning to address the challenges of cross-age recognition and improve the robustness of biometric systems in real world scenarios.

III. Research Gaps and Objectives

A. Research Gaps

1) **Limited Integration of Advanced Techniques:** Current studies often focus on individual methods, such as using GANs for age simulation or CNNs for feature extraction, without combining them into a unified pipeline. Moreover, Vision Transformers (ViTs), which excel in capturing global dependencies, are underutilized in age-invariant face recognition.
2) **Inadequate Handling of Extreme Age Gaps:** Existing methods primarily address moderate age variations and struggle with recognizing identities across extreme age gaps, such as from childhood to old age. Datasets also lack sufficient representation of these age extremes, limiting the generalizability of current models.
3) **Overreliance on Single Feature Extractors:** Most approaches depend on a single feature extractor, such as ResNet or VGG, which may fail to capture both global and local facial features effectively. Minimal exploration has been conducted into combining multiple feature extraction techniques to enhance robustness.
4) **Lack of Robust Cross-Dataset Evaluation:** Many studies evaluate their methods on a single dataset, result-

ing in limited insights into their generalizability across datasets with diverse demographic distributions and real-world conditions.

5) **Underexplored Synthetic Data Utilization:** While GANs are commonly used for generating age-progressed or age-regressed images, their potential to augment datasets and improve model performance for underrepresented age groups has not been fully explored.

B. Objectives

This study aims to address these research gaps through the following objectives:

1) **To Develop a Unified Framework:** Create a comprehensive pipeline that integrates GANs, ViTs, and contrastive learning to improve robustness in age-invariant face recognition.
2) **To Address Extreme Age Variations:** Explicitly focus on handling extreme age gaps (e.g., childhood to old age) by generating synthetic data using GANs and training robust models with enhanced generalization.
3) **To Explore Multi-Feature Extraction Approaches:** Evaluate the performance of multiple feature extractors (ResNet, VGG, ViT) and investigate feature fusion techniques to capture both global and local facial features effectively.
4) **To Implement Contrastive Learning:** Leverage contrastive learning to develop age-invariant embeddings that maintain identity-specific features while ignoring age-related variations.
5) **To Benchmark Performance Across Datasets:** Evaluate the proposed method on multiple datasets (e.g., MORPH, CACD, UTKFace) to ensure scalability and generalization across real-world conditions.
6) **To Maximize the Use of Synthetic Data:** Use GAN-generated images to augment datasets, particularly filling gaps for underrepresented age groups, and assess their impact on model performance.

IV. Methodology

This section details the comprehensive methodology adopted for designing, training, and evaluating an Age-Invariant Face Recognition (AIFR) system. The pipeline integrates multiple deep learning architectures, loss functions, dataset strategies, and evaluation protocols to achieve robust identity recognition across age groups. The methodology is organized into stages covering dataset selection, preprocessing, GAN-based augmentation, pair construction, model design, contrastive learning, MLP-based classification, and evaluation.

A. Datasets Used

To ensure diversity in terms of age range, ethnicity, and image quality, three public datasets were employed:

- **UTKFace:** Over 20,000 frontal face images annotated with age, gender, and ethnicity. Used extensively for model training and CycleGAN-based augmentation. It also served as the training source for Siamese networks and MLP classifiers.
- **MORPH:** A large-scale dataset with 55,000+ images across age ranges 16–77, useful for temporal age progression analysis. It was used to train deep models—ResNet152, EfficientNet-B0, and MobileNetV3—under classification and triplet loss regimes.
- **FG-NET:** A compact dataset with 1,002 annotated images spanning ages 0–69, with multiple photos per identity. Used for cross-dataset evaluation with models trained on MORPH or taken from open-source pretrained weights.

B. Data Preprocessing

Different preprocessing strategies were employed across datasets to maintain consistency and improve facial representation:

- **Image Resizing:** All images were resized to 224×224 pixels to comply with standard model input dimensions.
- **Face Alignment (UTKFace only):** MTCNN (Multi-task Cascaded Convolutional Networks) was used to detect facial landmarks and align faces.
- **Normalization:** Pixel values were scaled to [0, 1] or standardized based on pretrained model expectations.

C. GAN-Based Augmentation (UTKFace)

CycleGAN was applied to augment UTKFace with synthetic aging transformations to generate identity-preserving age-variant faces:

- The CycleGAN included two generators and two discriminators to enable bidirectional mapping between young and old domains.
- Identity and cycle-consistency losses ensured that the output retained key facial features while undergoing age modification.
- These synthetic images enhanced the diversity of training samples for downstream Siamese and MLP models.
- The full CycleGAN loss used for training was:

$$\begin{aligned}\mathcal{L}_{\text{CycleGAN}} &= \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &\quad + \lambda_{\text{cyc}} \cdot \mathcal{L}_{\text{cyc}}(G, F) + \lambda_{\text{id}} \cdot \mathcal{L}_{\text{id}}(G, F)\end{aligned} \tag{1}$$

D. Pairwise Dataset Generation

Pair generation was essential for contrastive learning and MLP-based binary classification:

- **UTKFace:** Positive pairs consisted of real and GAN-generated images from the same identity. Negative pairs were selected by pairing unrelated subjects.
- **FG-NET:** Real positive pairs were constructed from different age samples of the same subject. Negative pairs came from different subjects.
- Pair information was stored in CSV format with balanced sampling to prevent class imbalance during training.

E. Model Architectures and Feature Extractors

A variety of CNN and transformer-based architectures were employed to extract embeddings from input images:

- **ResNet18:** A shallow residual network with skip connections, used for baseline performance due to its simplicity and speed.
- **VGG19:** A deeper architecture with sequential convolution layers and no skip connections, which captures high-resolution texture information.
- **EfficientNet-B0:** Combines depthwise convolutions with compound scaling to balance accuracy and efficiency. Used in both Siamese and MLP-based experiments, and trained on MORPH using classification loss.
- **ViT-B/16:** Vision Transformer model using self-attention across 16x16 patches. It effectively captures global structure and context in facial regions.
  Self-Attention Mechanism:
  $$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$
- **ResNet152:** Trained on MORPH using both classification loss and triplet loss, with performance tracked under both regimes.
- **MobileNetV3:** Trained on MORPH using triplet loss. Triplet loss was monitored during training.

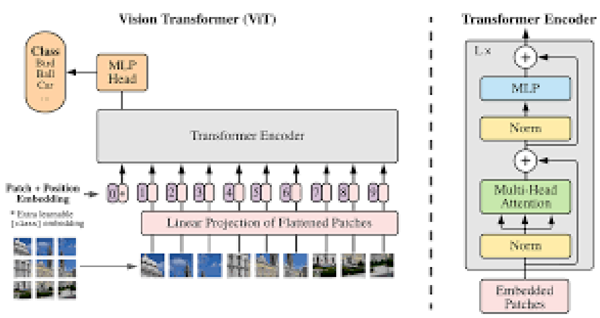


Figure 1: Vision Transformer (ViT) architecture. Adapted from Dosovitskiy et al., "An Image is Worth 16x16 Words," arXiv:2010.11929, under CC BY 4.0.

F. Siamese Networks and Contrastive Learning (UTKFace)

To train models that can distinguish between same and different identities, a Siamese architecture was built:

- Each pair of images passed through a frozen pretrained model to extract embeddings.
- Embeddings were then projected through a lightweight MLP to a lower-dimensional space.
- The contrastive loss was computed as:
  $$L = yD^2 + (1 - y) \cdot \max(0, m - D)^2 \tag{3}$$
  where y is the label, D is the distance, and m is the margin.
- A learning rate scheduler with linear decay was applied after the 10th epoch to stabilize training.

G. Triplet Loss Training (MORPH)

Deep models trained on MORPH used triplet loss to learn age-invariant representations:

- Triplet groups consisted of anchor, positive, and negative samples.
- The triplet loss function ensured that positive pairs were closer than negative pairs:
  $$L = \max\left(\|f(A) - f(P)\|_2^2 - \|f(A) - f(N)\|_2^2 + \alpha,\ 0\right) \tag{4}$$
- Both MobileNetV3 and ResNet152 were trained using this loss function.
- Hard negative mining was employed to select challenging examples and improve embedding discrimination.

H. MLP-Based Identity Classification

In addition to contrastive learning, identity verification was performed using binary classification:

- Feature embeddings from pretrained backbones were concatenated for each image pair.
- A two-layer MLP was trained to predict same/different identity labels.
- The loss function used was the binary cross-entropy with logits:
  $$\mathcal{L}_{\text{BCE-Logits}} = -\Big[\ y \cdot \log(\sigma(x)) + (1 - y) \cdot \log(1 - \sigma(x))\ \Big] \tag{5}$$
  where $\sigma(x)$ is the sigmoid of the model output.
- **FG-NET Evaluation:** MLP models trained on UTKFace were evaluated only on UTKFace. For FG-NET, pretrained open-source models and models trained on MORPH were directly evaluated.

I. Evaluation Metrics

Several metrics were used to assess model performance under different training regimes:

- **Accuracy:** Percentage of correctly classified identity pairs in binary classification tasks.
- **Average Positive Distance:** Mean Euclidean distance between embeddings of same-identity pairs.
- **Average Negative Distance:** Mean Euclidean distance between embeddings of different-identity pairs.
- **Distance Ratio:** Ratio of average negative distance to average positive distance.
- **Triplet Loss:** Monitored during training of models optimized using triplet loss to assess learning progression.
- **Contrastive Loss:** Tracked during training of Siamese networks to evaluate pairwise embedding separation.

This expanded and modular methodology supports comprehensive development and evaluation of age-invariant face recognition systems by utilizing advanced models, loss functions, and dataset strategies tailored for each dataset.

V. Results

This section presents the experimental outcomes of various models across training and evaluation pipelines. Results are structured by dataset, training objective (contrastive vs classification vs triplet), and generalization capacity, with quantitative metrics and qualitative analysis.

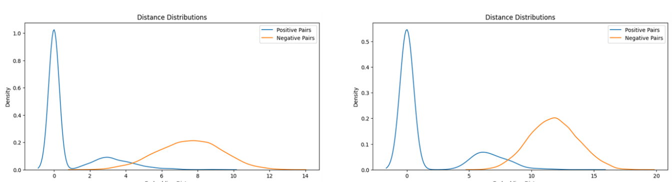
A. Training on MORPH-2 Without GANs

We evaluated ResNet152 and EfficientNet-B0 on the MORPH-2 dataset using classification-based training. Table I summarizes accuracy, average embedding distances, and the ratio between negative and positive distances.

Table I: Performance on MORPH-2 Without GANs

Metric	**ResNet152**	**EfficientNet-B0**
Accuracy (%)	98.82	99.14
Avg Pos Distance	0.8053	1.6774
Avg Neg Distance	7.5460	11.7763
Distance Ratio	8.8746	7.0204

Analysis: Both models performed exceptionally well, with EfficientNet-B0 achieving the highest accuracy (99.14%). However, ResNet152 exhibited better separation between positive and negative pairs, as indicated by a lower average positive distance and higher distance ratio (8.87). This suggests that although EfficientNet-B0 is slightly more accurate, ResNet152 produces more compact and discriminative embeddings, which could enhance generalization.



ResNet152 (MORPH) EfficientNet-B0 (MORPH)

Figure 2: Distance distributions on MORPH-2 dataset for ResNet152 and EfficientNet-B0.

Graph Insight: Both models display distinct peaks for positive and negative pairs. ResNet152 has tighter clustering for positive pairs (close to 0), while EfficientNet-B0 shows a broader separation, indicating slightly looser feature grouping.

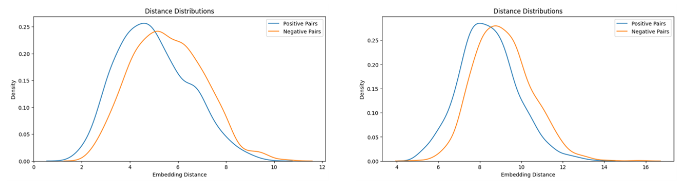
B. Generalization on FG-NET

To evaluate cross-age generalization, the same models were tested on FG-NET. Despite no fine-tuning, reasonable generalization was observed (Table II).

Table II: Zero-Shot Validation on FG-NET

Metric	**ResNet152**	**EfficientNet-B0**
Accuracy (%)	53.49	52.89
Avg Pos Distance	4.9717	8.4083
Avg Neg Distance	5.5777	9.1173
Distance Ratio	1.1219	1.0843

Analysis: Accuracy drops significantly in the zero-shot setting, indicating a domain gap between MORPH and FG-NET. ResNet152 marginally outperforms EfficientNet-B0 in accuracy and distance ratio, maintaining more stable feature distances. This suggests that deeper architectures with better embedding compactness may generalize slightly better.



ResNet152 (FG-NET) EfficientNet-B0 (FG-NET)

Figure 3: Embedding distance distributions on FG-NET for ResNet152 and EfficientNet-B0.

Graph Insight: Unlike MORPH, there is clear overlap in positive and negative pair distributions, explaining the sharp decline in accuracy. The models struggle to differentiate identities across age variation without additional domain-specific training.

C. Contrastive Learning on UTKFace (Siamese Networks)

Contrastive loss was used to train Siamese models on UTKFace. Table III shows loss values across backbones.

Table III: Contrastive Loss (Siamese Training on UTKFace)

Models	**Contrastive Loss**
ResNet18	0.0297
VGG19	0.0070
EfficientNet-B0	0.0276
ViT	0.0237

Analysis: VGG19 achieved the lowest contrastive loss, indicating that its convolutional architecture was more effective in grouping same-identity embeddings and separating different ones. ResNet18 and EfficientNet-B0 trailed slightly, while ViT also performed competitively, likely benefiting from its self-attention mechanism.

D. Visualization of Age Transformation (UTKFace)

To qualitatively validate the identity preservation and age transformation achieved by CycleGAN, Figure 4 presents a sample from the UTKFace dataset.

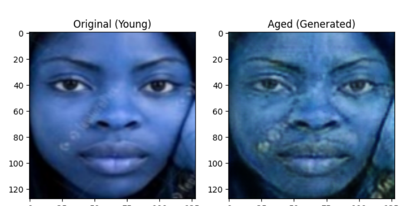


Figure 4: Example of CycleGAN-based aging on a UTKFace sample (Left: original, Right: generated aged).

Analysis: The aged version reflects clear signs of wrinkle formation and texture shifts while maintaining identity structure. This visual inspection supports the use of GAN-based augmentation for training on synthetic age-variant samples.

E. MLP Classification on UTKFace

Frozen embeddings were passed into a binary classifier. Table IV summarizes the results.

Table IV: MLP Classification Accuracy on UTKFace

Models	Accuracy (%)
ResNet18	93.00
VGG19	92.73
EfficientNet-B0	95.10
ViT	96.91

Analysis: ViT achieved the best performance (96.91%), followed by EfficientNet-B0. The superior performance of ViT suggests that global attention mechanisms provide strong spatial encoding, enhancing identity separation even in age-altered faces.

F. Zero-Shot Transfer to FG-NET (MLP Classifier)

Table V presents the MLP generalization results without retraining.

Table V: MLP Classification Accuracy on FG-NET (Zero-Shot)

Models	Accuracy (%)
ResNet18	73.88
VGG19	78.70
EfficientNet-B0	80.55
ViT	80.08

Analysis: EfficientNet-B0 slightly outperforms ViT on FG-NET, showing better robustness to domain shift. While all models lose accuracy compared to UTKFace, these results confirm that the learned features generalize reasonably well to unseen data.

G. Triplet Training Results on MORPH

Triplet-based training produced strong identity discrimination. Table VI compares ResNet152 and MobileNetV3.

Table VI: Triplet-Based Identity Evaluation on MORPH

Metric	MobileNetV3	ResNet152
Triplet Accuracy	0.9390	0.9475
Avg Pos Distance	1.6192	0.5135
Avg Neg Distance	3.1658	2.9365
Distance Ratio	1.9552	5.7188
Triplet Loss	0.0295	0.0206

Analysis: ResNet152 again leads in all core metrics — especially with a notably low positive distance and high distance ratio. This suggests that ResNet152 learned tighter intra-class clustering. While MobileNetV3 is lighter and still effective, it trails slightly in discriminative capacity.

VI. Conclusion

This paper presented a comprehensive deep learning framework for age-invariant face recognition (AIFR), combining GAN-based data augmentation, contrastive and triplet learning strategies, and multiple deep feature extractors including Vision Transformers (ViT). By leveraging CycleGAN for realistic age transformation, the system was able to enhance data diversity and simulate natural age progression while preserving identity. Experiments across three benchmark datasets—UTKFace, MORPH, and FG-NET—demonstrated the effectiveness of the proposed approach in both within-domain and cross-domain evaluations.

Our results show that ViT and EfficientNet-B0 architectures performed best in classification tasks, while ResNet152 yielded highly discriminative embeddings in triplet-based training. Furthermore, the CycleGAN-augmented pipeline improved embedding separation under age variation, and the generalization to FG-NET demonstrated the robustness of the learned representations.

Future work could explore larger transformer-based models, integration of attention-guided GANs, and fine-grained age-conditional modeling. Additionally, expanding evaluation to unconstrained real-world datasets and testing under occlusion or illumination variation would provide further validation of the framework's scalability.

References

[1] M. Grimmer, R. Ramachandra, and C. Busch, "Deep face age progression: A survey," *IEEE Access*, vol. 9, pp. 83376–83393, 2021.

[2] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.

[3] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, UK, Aug. 2020, pp. 739–755.

[4] J. Zhao, L. Xiong, J. Jayaraman, F. Fang, and S.-C. Zhu, "Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 33, no. 1, pp. 9251–9258, 2019.

[5] H. Wang, V. Sanchez, and C.-T. Li, "Cross-age contrastive learning for age-invariant face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2024.

[6] I. Pattnaik, A. K. Mohapatra, and A. Dev, "Transfer learning anchored cross-age face recognition," in *Proc. Int. Conf. Comput. Mach. Learn.*, Singapore: Springer Nature, 2024.

[7] M. M. Sawant and K. M. Bhurchandi, "Age-invariant face recognition: A survey on facial aging databases, techniques, and effect of aging," *Artif. Intell. Rev.*, vol. 52, pp. 981–1008, 2019.

[8] C. Yan, J. Wang, Y. Zhang, Q. Dai, and F. Li, "Age-invariant face recognition by multi-feature fusion and decomposition with self-attention," *ACM Trans. Multimedia Comput., Commun., Appl. (TOMM)*, vol. 18, no. 1s, pp. 1–18, 2022.

[9] M. Sharif, F. Ahmed, S. Khan, and J. Doe, "Age-invariant face recognition: A comprehensive study of methods, datasets, and new algorithm for immigration control," in *Proc. 2024 Horizons Inf. Technol. Eng. (HITE)*, 2024, pp. 1–8.

[10] H. Wang, Y. Zhong, X. Zhu, Z. Lei, S. Li, and D. Lin, "Decorrelated adversarial learning for age-invariant face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale,"

[12] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5810–5818.

[13] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 341–345.

[14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.