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

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, resulting 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 outlines the proposed methodology for developing a robust Age-Invariant Face Recognition (AIFR) model. The methodology is structured into five key stages: dataset preparation, data preprocessing, model pipeline (divided into individual components), and evaluation metrics.

A. Datasets Used

Three datasets are selected to ensure effective model training and evaluation:

- **UTKFace:** A dataset containing around 20,000 images with age labels spanning from 0 to 116 years. It provides diverse ethnic backgrounds, making it suitable for model training. UTKFace contains frontal face images with variations in age, gender, and ethnicity, making it an ideal choice for training robust models that generalize across demographics.
- **MORPH:** A dataset with approximately 55,000 images covering age ranges between 16 to 77 years. This dataset enhances model robustness by offering improved diversity across racial groups. MORPH is particularly effective in addressing age progression and regression modeling due to its detailed labeling of identity and age progression timelines.
- **FG-NET:** A dataset comprising roughly 1,000 images spanning ages from 0 to 69 years. It is used exclusively for model evaluation to assess generalization capabilities. Despite its smaller size, FG-NET is valuable for verifying model accuracy across extreme age ranges and unseen subjects.

Each dataset is resized, normalized, and aligned to ensure consistency across different data sources.

B. Data Preprocessing

Data preprocessing involves several crucial steps to improve data quality and enhance model performance:

- **Image Resizing:** All images are resized to a fixed resolution of 224×224 pixels to ensure uniform input dimensions across datasets.
- **Normalization:** Pixel values are normalized to the range [0, 1] to accelerate model convergence during training.
- **Face Alignment:** Facial landmarks are detected using techniques such as MTCNN or Dlib to align faces for better feature extraction.
- **Augmentation:** Additional augmentation techniques such as rotation, flipping, and brightness adjustments are applied to improve model generalization and prevent overfitting.

These preprocessing steps ensure consistent image quality and improve model robustness.

C. GAN-Based Augmentation

To address the limited presence of extreme age groups, we implement a Generative Adversarial Network (GAN) based on the Wasserstein GAN with Gradient Penalty (WGAN-GP) architecture.

Loss Function:

$$L = \mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] - \mathbb{E}_{x\sim P_r}[D(x)] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \tag{1}$$

The generator synthesizes realistic aged face images while the discriminator distinguishes between real and generated samples. The gradient penalty term ensures stable convergence and improved image quality.

To enhance the effectiveness of GAN-based augmentation, we focus on generating synthetic faces for underrepresented age groups such as infants and elderly individuals. These synthetic images are added to the UTKFace and MORPH datasets to ensure a balanced age distribution, improving model robustness across diverse age groups.

D. Feature Extraction

A combination of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) is used for robust feature extraction:

- **CNNs:** Efficiently capture local facial features through convolutional layers with residual connections. CNNs are chosen for their ability to capture fine-grained texture information and facial landmarks crucial for face recognition tasks.
- **ViTs:** Extract global facial patterns via self-attention mechanisms to enhance identity preservation across varying age groups. ViTs are particularly effective in learning contextual dependencies between facial features, improving model performance on aging variations.

Self-Attention Mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Where Q, K, and V represent the query, key, and value matrices, and d_k is the dimensionality of the key vector.

E. Contrastive Learning with Triplet Loss

To improve identity consistency across age groups, contrastive learning with the Triplet Loss function is applied. The model learns to minimize the distance between anchor-positive pairs while maximizing the distance between anchor-negative pairs.

Triplet Loss Function:

$$L = \sum_{i=1}^{m} \max\left(\|f(A^{(i)}) - f(P^{(i)})\|_2^2 - \|f(A^{(i)}) - f(N^{(i)})\|_2^2 + \alpha, 0 \right) \tag{3}$$

Where A is the anchor image, P is the positive image (same identity), N is the negative image (different identity), and α is the margin parameter ensuring separation between positive and negative pairs.

To improve performance, triplet sampling strategies such as hardest positive/negative mining are incorporated to focus on challenging identity distinctions.

F. Identity Classification Using MLP

The extracted feature embeddings are passed through a Multi-Layer Perceptron (MLP) classifier to predict the corresponding identities. The MLP structure includes:

- Fully connected layers with ReLU activation for non-linearity.
- Dropout layers for regularization to mitigate overfitting.
- A softmax output layer for final identity prediction.

Cross-Entropy Loss Function:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{4}$$

Where y_i is the true class label and $\hat{y}_i$ is the predicted probability for class i.

G. Evaluation Metrics

The following metrics are employed to evaluate model performance:

- **Accuracy:** Measures the proportion of correctly identified identities.
- **Equal Error Rate (EER):** Determines the point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR).
- **F1-Score:** Balances precision and recall to assess overall performance.
- **ROC Curve:** Visualizes the trade-off between sensitivity and specificity across thresholds.
- **Inference Time:** Evaluates the computational efficiency of the final model for real-time deployment scenarios.

This comprehensive evaluation strategy ensures that the model is assessed for both identity preservation and real-world applicability.

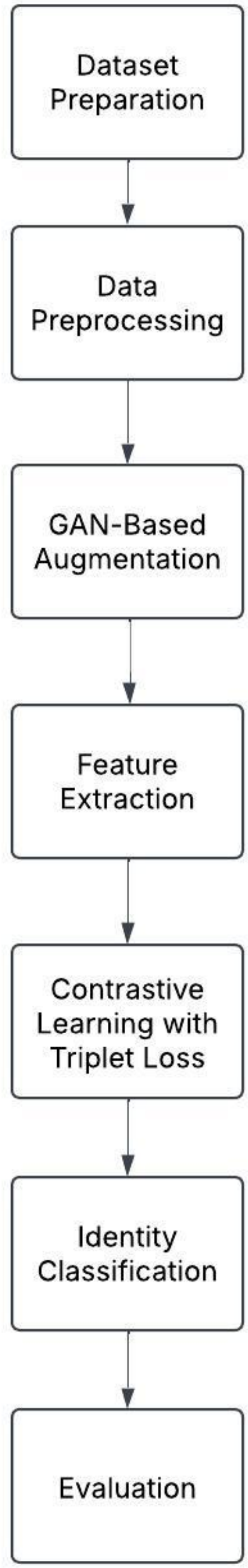

Figure 1: Proposed Methodology Flowchart for Age-Invariant Face Recognition.

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.