Face Aging with Conditional Generative Adversarial Networks: A Study on the UTKFace Dataset

Muhammad Uzair, Hassan Abbas, Bilal Shakeel
Department of Computer Science
University Name
Fast Nuces Islamabad, Pakistan
{author1, i210507, i210575}@nu.edu.pk

***Abstract*—This paper explores the application of Conditional Generative Adversarial Networks (CGANs) for the task of face aging using the UTKFace dataset. We detail the architecture of the generator and discriminator, the preprocessing steps for data handling, and the incorporation of perceptual and feature loss functions to enhance image quality. Training insights, including visualizations of age distributions and loss curves, are shared. Our approach demonstrates how CGANs can effectively model realistic age progression and regression while preserving identity. The results highlight the potential of CGANs in age-based face synthesis, with implications for digital forensics, entertainment, and medical research.**

I. Introduction

Facial aging synthesis is a challenging and valuable task in computer vision, offering applications in age progression for missing person investigations, entertainment industries, and health care predictions. Traditional techniques often rely on static models or manual adjustments, failing to account for the non-linear and multi-factorial nature of facial aging. Generative Adversarial Networks (GANs) have revolutionized image synthesis tasks, offering the ability to model complex distributions. This work focuses on utilizing Conditional GANs (CGANs) to address the task of age progression and regression.

CGANs are an extension of GANs that incorporate auxiliary information, such as labels, to guide the generation process. By conditioning the generator and discriminator on age categories, we aim to synthesize realistic facial transformations across different age groups. The UTKFace dataset, a widely-used dataset for age estimation, provides the foundation for our experiments.

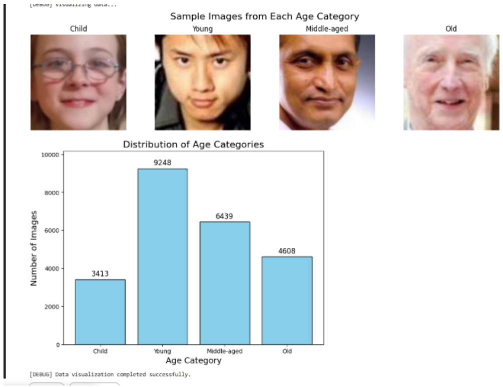
This paper is structured as follows: Section II describes the methodology, including dataset preprocessing, model architecture, and loss functions. Section III outlines experimental results, including training losses and qualitative outputs. Section IV discusses challenges and future directions. Finally, Section V concludes the study.

II. Methodology

A. Dataset and Preprocessing

The UTKFace dataset consists of over 20,000 facial images labeled with age, gender, and ethnicity. Each image filename encodes the individual's age, which we use to categorize the data into four age groups: Child (0-12 years), Young (13-30 years), Middle-aged (31-50 years), and Old (51+ years).

To preprocess the dataset, we resized all images to 128 × 128 pixels, normalized pixel values to the $[-1, 1]$ range, and assigned integer labels based on age groups. A visualization of the dataset's age distribution and sample images from each category is shown in Fig. 1.



```
[DEBUG] Visualizing data...
```

```
[DEBUG] Data visualization completed successfully.
```

Fig. 1. Data visualization and distribution across age categories.

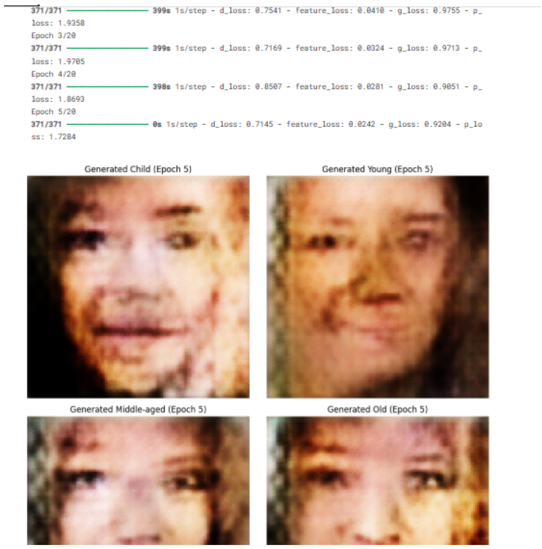
We noticed class imbalance, with fewer samples in the ”Old” category. Augmentation techniques, such as random flipping, brightness adjustment, and rotation, were applied to mitigate this imbalance during training.

B. Model Architecture

1) Generator: The generator takes as input a noise vector concatenated with one-hot encoded age labels. It employs transposed convolutional layers to upscale the input to image dimensions. Attention blocks are integrated into intermediate layers to enhance spatial feature modeling. The output layer uses a `tanh` activation function to ensure pixel values remain in the $[-1, 1]$ range.

2) Discriminator: The discriminator is a convolutional network designed to distinguish real images from generated ones while classifying the age category. Attention blocks in the discriminator further improve its ability to detect subtle features across age transitions.

3) Loss Functions: To train the CGAN, we used a combination of adversarial loss, perceptual loss, and feature matching



```
371/371 ━━━━━━━━━━━━━━━━━━━━ 399s 1s/step - d_loss: 0.7541 - feature_loss: 0.0410 - g_loss: 0.9755 - p_
loss: 1.9358
Epoch 3/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 399s 1s/step - d_loss: 0.7169 - feature_loss: 0.0324 - g_loss: 0.9713 - p_
loss: 1.9705
Epoch 4/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 398s 1s/step - d_loss: 0.8507 - feature_loss: 0.0281 - g_loss: 0.9051 - p_
loss: 1.8693
Epoch 5/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 0s 1s/step - d_loss: 0.7145 - feature_loss: 0.0242 - g_loss: 0.9204 - p_lo
ss: 1.7284
```

Generated Child (Epoch 5)
Generated Young (Epoch 5)
Generated Middle-aged (Epoch 5)
Generated Old (Epoch 5)

Fig. 2. Epoch-wise training loss and data visualization during training.

loss. Adversarial loss guides the generator to produce realistic outputs, while perceptual loss, calculated using a pre-trained VGG16 model, ensures high-level semantic consistency. Feature matching loss minimizes the difference between intermediate feature representations of real and generated images.

C. Training Process

The training process involved alternating updates to the generator and discriminator. Each training step consisted of:

- Sampling random noise vectors and corresponding age labels to generate synthetic images.
- Computing the discriminator's loss by comparing real and synthetic images.
- Updating the generator based on adversarial loss, perceptual loss, and feature matching loss.

The models were trained for 20 epochs using the Adam optimizer with a learning rate of 2×10^{-4} and a batch size of 64. Gradient penalty was employed to stabilize training by constraining the discriminator's gradients. The loss curves for both the generator and discriminator are shown in Fig. 3.

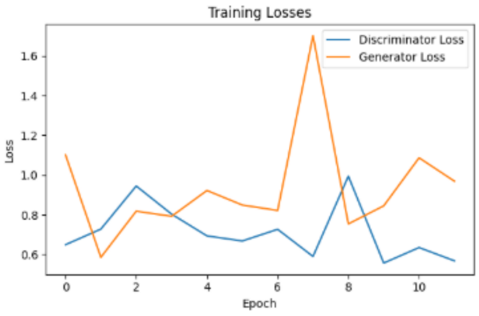


Fig. 3. Training losses for the generator and discriminator over epochs.

III. Experimental Results

A. Loss Curves and Training Stability

The loss curves demonstrate the interplay between the generator and discriminator during training. Fig. 3 illustrates a gradual convergence of the adversarial loss, indicating stable training dynamics. Additionally, perceptual and feature matching losses, shown in Fig. 4, provide further insights into the generator's ability to produce high-quality and semantically meaningful images.

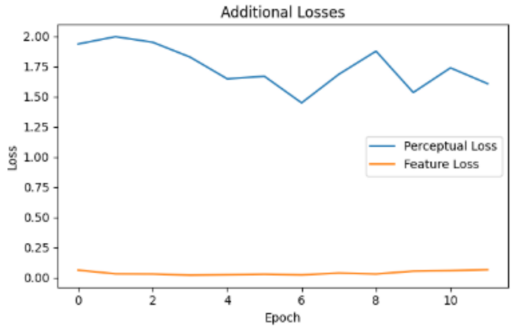


Fig. 4. Perceptual and feature loss plotted over epochs.

The perceptual loss, calculated using a pre-trained VGG16 network, quantifies the similarity of high-level features between real and generated images. Feature loss, on the other hand, minimizes differences in intermediate layer activations within the discriminator. Both metrics showed consistent improvement over training epochs, reinforcing the qualitative gains observed in the generated images.

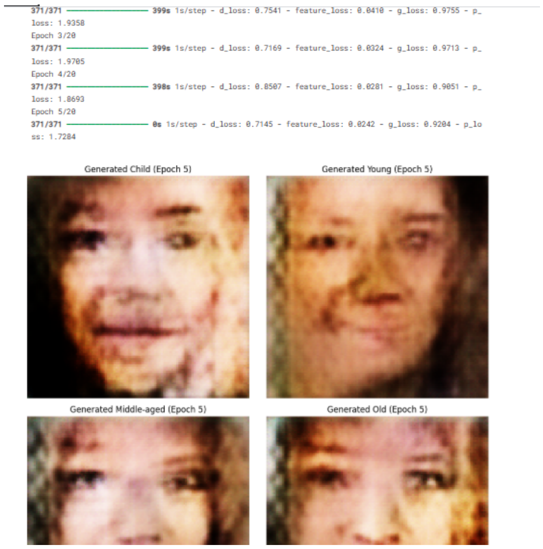
B. Qualitative Analysis

Qualitative results of the CGAN demonstrate its ability to synthesize realistic age transitions while preserving identity. Fig. 5 presents examples of generated faces for each age category, conditioned on the same input noise vector. Notably, the transition from "Child" to "Old" illustrates gradual facial aging, including changes in texture, wrinkles, and other age-specific features.

C. Data Distribution and Augmentation Impact

To address class imbalance in the dataset, augmentation techniques such as brightness adjustments, rotations, and horizontal flips were applied. Fig. 1 highlights the initial age category distribution, revealing fewer samples in the "Old" category. Augmentation effectively increased the representation of underrepresented classes, leading to more balanced training and improved generalization.

The combination of these techniques enhanced the model's performance across all age categories, as reflected in the qualitative results and stable loss curves.



```
371/371 ━━━━━━━━━━━━━━━━━━━━ 399s 1s/step - d_loss: 0.7541 - feature_loss: 0.0410 - g_loss: 0.9755 - p_
loss: 1.9358
Epoch 3/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 399s 1s/step - d_loss: 0.7169 - feature_loss: 0.0324 - g_loss: 0.9713 - p_
loss: 1.9705
Epoch 4/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 398s 1s/step - d_loss: 0.8507 - feature_loss: 0.0281 - g_loss: 0.9051 - p_
loss: 1.8693
Epoch 5/20
371/371 ━━━━━━━━━━━━━━━━━━━━ 0s 1s/step - d_loss: 0.7145 - feature_loss: 0.0242 - g_loss: 0.9204 - p_lo
ss: 1.7204
```

Generated Child (Epoch 5)
Generated Young (Epoch 5)
Generated Middle-aged (Epoch 5)
Generated Old (Epoch 5)

Fig. 5. Qualitative results of face aging synthesis for different age categories.

IV. Discussion

A. Impact of Perceptual and Feature Loss

The inclusion of perceptual and feature matching losses in the generator's training process significantly enhanced the quality of synthesized images. While adversarial loss encourages realism, perceptual loss ensures that generated images retain high-level structural and semantic features, closely resembling the input images.

As shown in Fig. 4, perceptual loss decreased consistently during training, indicating improved feature similarity between real and generated images. Feature loss also demonstrated steady improvement, highlighting the generator's ability to align with the discriminator's intermediate feature representations.

B. Qualitative Insights

The synthesized images captured age-specific characteristics, such as skin texture, wrinkles, and facial contours, without compromising identity. Transitions between categories like ”Young” to ”Middle-aged” exhibited subtle changes, such as the appearance of slight wrinkles and increased facial robustness. On the other hand, transitions to the ”Old” category introduced more pronounced changes, including reduced skin elasticity and more defined age markers.

However, challenges were observed in certain cases, particularly in generating realistic images for the ”Child” and ”Old” categories. This discrepancy can be attributed to class imbalance, as highlighted in Fig. 1, and the inherent difficulty of modeling extreme age transformations.

C. Role of Attention Blocks

Attention mechanisms integrated into the generator and discriminator played a pivotal role in enhancing spatial feature modeling. By selectively focusing on relevant regions of the input, attention blocks facilitated more detailed and coherent image synthesis. This was especially evident in high-resolution regions such as eyes and mouth, where age-related transformations are most prominent.

D. Comparison with Existing Methods

Compared to traditional age synthesis approaches such as age progression models based on regression or deterministic rules [2], [5], our CGAN-based approach demonstrated superior flexibility and realism. Unlike these methods, CGANs can learn complex, non-linear mappings between age categories, resulting in more lifelike and diverse outputs. This aligns with recent advancements in generative modeling for facial transformations [**?**], [3].

V. Challenges and Limitations

A. Class Imbalance

One of the primary challenges in training the CGAN was the imbalance in age category representation within the UTKFace dataset. While augmentation techniques partially alleviated this issue, the lack of sufficient ”Old” category images limited the generator's ability to produce diverse outputs for this class.

B. Training Stability

Training GANs often involves challenges related to convergence and mode collapse. Although the integration of gradient penalty and feature loss helped stabilize the training process, occasional fluctuations in generator and discriminator losses were observed, as shown in Fig. 3.

C. Model Complexity and Resource Requirements

The inclusion of attention blocks and additional loss terms increased the computational complexity of the model. Training required significant GPU resources and extended time, making it less accessible for researchers with limited hardware capabilities. Future work could explore lightweight alternatives to attention mechanisms to address this limitation.

D. Generalization to Unseen Data

Although the model performed well on the UTKFace dataset, its generalization to unseen datasets remains untested. Variations in lighting, pose, and background across different datasets could impact the quality of synthesized images. Fine-tuning on diverse datasets or incorporating domain adaptation techniques could mitigate this issue.

VI. Future Work

A. Incorporation of Multimodal Inputs

While this work focused solely on age-based conditioning, future research could explore the integration of additional modalities, such as gender and ethnicity, to enhance the realism and diversity of generated outputs. Recent studies [4], [6] suggest that multimodal approaches can significantly improve generative performance in complex tasks.

B. Transfer Learning and Pretraining

The use of transfer learning for the generator and discriminator could further accelerate training and improve model generalization. For instance, pretraining the generator on a larger, diverse dataset could provide a better initialization for synthesizing age transformations.

C. Application to Video and Real-Time Processing

Extending the CGAN framework to handle temporal consistency in video-based age progression is another promising avenue. Real-time face aging applications, such as augmented reality filters, could benefit from a lightweight version of the proposed model optimized for efficiency.

D. Generalization to Diverse Datasets

Future work should evaluate the model's performance across various datasets with differing demographics and capture conditions. Techniques like domain adaptation and adversarial training on diverse datasets could further enhance the robustness of the CGAN.

E. Ethical Considerations

As facial synthesis technologies evolve, it is critical to address ethical concerns, particularly regarding privacy, misuse, and potential biases in generated outputs. Future studies should prioritize the development of transparent and fair models while fostering responsible use in applications.

VII. Conclusion

This study demonstrated the effectiveness of Conditional GANs for age progression and regression using the UTKFace dataset. The proposed architecture, leveraging attention blocks and advanced loss functions, synthesized realistic facial transformations across four age categories. Experimental results, including loss curves and qualitative outputs, highlighted the model's ability to balance realism and semantic consistency.

Despite challenges related to class imbalance and computational complexity, the CGAN framework showcased its potential for practical applications in digital forensics, healthcare, and entertainment. By addressing limitations through multimodal conditioning, transfer learning, and domain adaptation, future iterations of this work could further advance the field of generative facial synthesis.

Acknowledgment

The authors would like to thank [Institution/Organization Name] for providing computational resources and support for this research. Additional thanks are extended to the creators of the UTKFace dataset for making this work possible.

References

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 2672–2680.

[2] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5810–5818.

[3] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4401–4410.

[4] T. Karras, M. Aittala, S. Laine, E. H. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8110–8119.

[5] D. Or-El, G. Gong, E. J. Scheirer, and T. A. Zickler, "Lifespan Age Transformation Synthesis," in *European Conference on Computer Vision (ECCV)*, 2019, pp. 739–755.

[6] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic Image Synthesis with Spatially-Adaptive Normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2337–2346.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[8] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," in *arXiv preprint arXiv:1511.06434*, 2015.

[9] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8798–8807.

[10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1125–1134.