2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Simultaneous Facial Age Transformation and Reenactment

Jie-Ying Zhang*
Nat'l Taiwan Univ. Sci. Tech.
m11151021@mail.ntust.edu.tw

Li-Syun Hsiung*
Nat'l Taiwan Univ. Sci. Tech.
m11103606@mail.ntust.edu.tw

Gee-Sern Jison Hsu
Nat'l Taiwan Univ. Sci. Tech.
jison@mail.ntust.edu.tw

***Abstract*—This paper explores concatenating pre-trained models for simultaneous facial age transformation and face reenactment, emphasizing image quality enhancement. We introduce an identity recognition loss function during age transformation model development to separate identity and age features, optimizing it with a finely-tuned age prediction model. Our research highlights the success of this concatenated training process, especially in remarkable image generation results.**

***Index Terms*—age estimation, identity recognition, facial reenactment, age transformation**

I. Introduction

In the realm of computer vision, challenges are prevalent in the domains of facial age transformation and face reenactment. Despite generative model advancements, methods [1] [2] [7] are often specialized, lacking versatility. Even advanced diffusion models struggle with control and identity preservation in facial transformation. Our goal is to create a framework overcoming these challenges, enabling multiple image editing tasks simultaneously. We integrate pre-trained models, coupled with the inclusion of loss functions capable of decorrelating identity and age features to address the challenges of identity preservation. Subsequently, we perform rigorous fine-tuning on the "FFHQ-aging Multi-Pose" dataset to achieve desired outcomes, unlocking new possibilities in this field.

The contributions of this paper are summarized as follows:

- Introduced a training pipeline concatenating pre-trained models for age transformation and face reenactment.
- We propose a data set that involves altering the pose of the original ECAF and SCAF datasets [3], aimed at fine-tuning the age estimator to enhance the quality of the "FFHQ-aging Multi-Pose" data set.

II. Related Work

In age transformation, Hsu et al. [1] balanced age alteration with identity preservation using generative adversarial networks, ensuring effective age transformation. Z. Huang et al. [3] introduced MTLFace, a framework addressing age-invariant face recognition and age synthesis while preserving facial identity.

For face reenactment, Hsu et al. [2] introduced a method that preserves the individual's identity while transferring facial expressions from a source to a target face using a dual-generator architecture.

Our approach combines the Dual-Generator Network [2] and AgeTransGAN [1], enhancing the "FFHQ-aging" dataset [7] with ControlNet [4]. Additionally, we integrate MTLFace to improve AgeTransGAN's capability to accurately preserve facial identity across various age groups.

III. Methodology

The proposed pipeline comprises three main components: an age estimator for data relabeling, Attention-based Identity Feature Decomposition (AIFD), and the inference stage. Detailed descriptions are provided below.

A. Age Estimator for data relabeling

In the initial phase of our proposal, we employ an Age Estimator to reannotate the Augmented FFHQ-aging dataset. Our Age Estimator implementation closely follows the training process and configuration outlined in [1], demonstrating remarkable performance across various datasets. We initialize the model with pre-trained weights from ImageNet and perform additional pre-training on the IMDB-WIKI dataset, followed by fine-tuning using the alternate SCAF dataset.

B. Attention-based Identity Feature Decomposition (AIFD)

In cross-age face recognition, age disparities lead to variations in facial features, giving rise to distinctive identification challenges. Z. Huang et al. [3] introduced an approach involving the decomposition of combined feature maps within a higher-level semantic domain using an attention mechanism, referred to as "Attention-Based Feature Decomposition" (AFD). The complexity of manipulating feature vectors compared to feature maps underscores the rationale behind this method. Aging or rejuvenation effects, such as the presence of beards and wrinkles, manifest within the semantic feature space but are less evident in one-dimensional features. [3] utilized a ResNet backbone as an encoder to extract mixed feature maps X from an input image I, $i.e., X = E(I)$.

$$X = \underbrace{X \circ \sigma(X)}_{X_{age}} + \underbrace{X \circ (1 - \sigma(X))}_{X_{id}} \tag{1}$$

where $\circ$ indicates element-wise multiplication, and σ is the attention module. The attention module focuses on age-related information in feature maps, guided by an age estimation task, while the residual segment, containing identity related details, is supervised via a face recognition task.

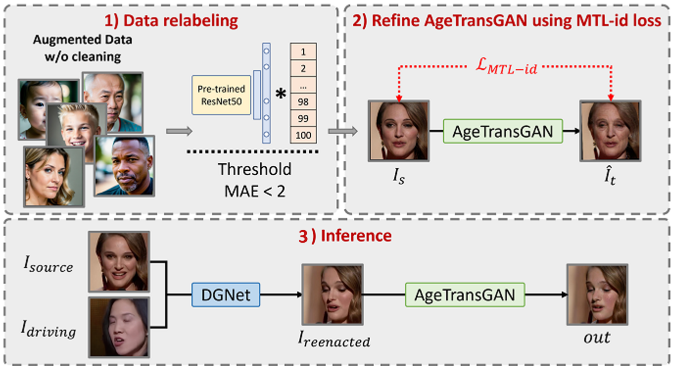

Fig. 1. Outline of the proposed method. The first step covers data preprocessing, the second involves fine-tuning the AgeTransGAN [1] model with MI loss on FFHQ-aging Multi-Pose data set, and the third demonstrates simultaneous facial reenactment and age transformation.

In our approach, we integrate the residual part with identity-related features, passing them through convolutional layers for Age-invariant Face Recognition (AIFR). We introduce an AIFD Encoder, denoted as E_A, for AIFR. Subsequently, E_A encodes input images, creating the multi-task learning id loss (MI loss) function. This loss function integrates into the AgeTransGAN model, enhancing identity feature preservation during fine-tuning on the "FFHQ-aging Multi-Pose" dataset.

$$\mathcal{L}_{MI} = 1 - \cos(E_A(I_s), E_A(\hat{I}_t)) \tag{2}$$

where E_A is the attention-based feature decomposition (AIFD) Encoder. And I_s stands for the input Image of AgeTransGAN, $\hat{I}_t$ stands for the output.

IV. Experiment

We employed the "FFHQ-aging Multi-Pose" dataset, previously introduced in prior research [6], containing a total of 10,400 images. Among these images, 90% were designated for the training set, with the remaining 10% reserved for testing.

Given the current scarcity of age datasets featuring multi-angle facial images and the absence of age estimators suitable for evaluating such images, we addressed this gap by augmenting the SCAF [3] and ECAF [3] datasets using the method proposed in [2]. Our augmentation process was designed to incorporate multi-angle facial images into these datasets. Consequently, we designated SCAF as the training set and ECAF as the testing set.

Subsequently, we conducted fine-tuning on the age estimator introduced by [5]. This fine-tuning process resulted in a reduction in the mean age error from 8.54 to 7.22, which enhanced the accuracy of age estimation.

To further improve AgeTransGAN's ability to preserve identity across various age groups, we introduced the MI Loss into the framework. The performance results, including the impact of the MI Loss, are presented in "TABLE I."

In our inference process, as depicted in "Fig.1", we expanded upon previous experiments, successfully achieving simultaneous facial reenactment and age transformation. During inference, we observed a correlation between smaller face angles and younger ages, with the correlation gradually increasing as the angles became more acute. These findings are showcased in the high-quality results depicted in "Fig. 2."

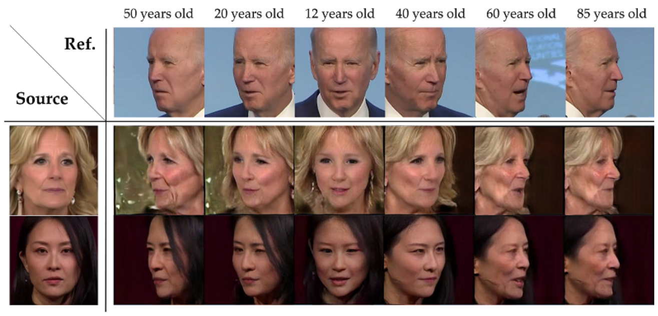

Fig. 2. Inference results for simultaneous face representation and age transformation

TABLE I
COMPARING THE EFFECT OF ADDING MI LOSS TO AGETRANSGAN

Age group	*0-2*	*3-6*	*7-9*	*10-14*	***15-19***	*30-39*	*40-49*	*50-69*	*70+*
Face Verification Rate (%)									
w/o MI Loss	*83.2*	*95.4*	*98.5*	***99.3***	*99.3*	***99.7***	*99.3*	*99.1*	*93.6*
with MI Loss	***85.3***	***96.4***	***99.1***	*99.2*	***99.7***	*99.6*	***99.7***	***99.6***	***94.2***
EAM (Estimated Age Mean)									
Real	*1.5*	*4.9*	*8.6*	*12.8*	*18.9*	*31.9*	*43.9*	*57.2*	*68.9*
w/o MI Loss	***2.1***	*5.2*	*8.3*	*13.8*	*17.6*	***33.4***	***41.8***	***54.6***	***65.7***
with MI Loss	***2.1***	***5.0***	***8.6***	***13.4***	***18.5***	*34.7*	*41.2*	*54.1*	*65.3*

V. Conclution

In conclusion, facial manipulation generative models have often been domain-specific, limiting cross-domain transformations. Our approach focuses on concatenating existing models for concurrent facial age transformation and face reenactment. Utilizing the 'FFHQ-aging Multi-Pose' augmented dataset, we optimize it with a finely tuned age transition model and introduce an identity recognition loss function to separate identity from age features. Our research validates the effectiveness of this concatenated training approach, yielding impressive image generation results. In summary, our study advances multi-domain image generation through concatenated models and innovative methodologies with broad applicability.

References

[1] Hsu, Gee-Sern and Xie, Rui-Cang and Chen, Zhi-Ting and Lin, Yu-Hong. "AgeTransGAN for Facial Age Transformation with Rectified Performance Metrics." In: ECCV (2022)

[2] Hsu, Gee-Sern, Chun-Hung Tsai, and Hung-Yi Wu. "Dual-generator face reenactment." In: CVPR (2022)

[3] Huang, Zhizhong, Junping Zhang, and Hongming Shan. "When age-invariant face recognition meets face age synthesis: A multi-task learning framework." In: CVPR (2021)

[4] Zhang, Lvmin, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." arXiv:2302.05543 (2023).

[5] Paplham, Jakub, and Vojtech Franc. "Unraveling the Age Estimation Puzzle: Comparative Analysis of Deep Learning Approaches for Facial Age Estimation." In: arXiv:2307.04570 (2023).

[6] Zhang, Jie-Ying, Hsiung, Li-Syun, and Hsu, Gee-Sern. "Unleashing Stable Diffusion for Multi-Modal Facial Image Augmentation.", In: Aris (2023)

[7] Or-El, R., Sengupta, S., Fried, O., Shechtman, E., Kemelmacher-Shlizerman, I. (2020). "Lifespan age transformation synthesis." In: ECCV (2020)