Journal of Multimedia Information System VOL. 12, NO. 3, September 2025 (pp. 95-110): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2025.12.3.95

Data Augmentation Using a Convolutional Autoencoder for Age Estimation from Facial Images

Beom Kwon[1*]

Abstract

Accurate age estimation from facial images plays a pivotal role in various computer vision applications, such as biometric authentication, surveillance, and human-computer interaction. However, the inherent data imbalance across different age ranges particularly the scarcity of samples in older age groups poses a significant challenge for model training. In this paper, we propose a data augmentation strategy based on a convolutional autoencoder (CAE) to address this limitation. By generating synthetic facial images through convex interpolation in the latent space, the proposed method compensates for underrepresented age groups while preserving realistic facial features. To ensure age consistency in the augmented data, a convex combination of age labels is used in tandem with the latent representations. Extensive experiments are conducted on the FG-NET dataset using the leave-one-person-out (LOPO) cross-validation protocol. Evaluation results across six convolutional neural network (CNN) models demonstrate that our method consistently improves age estimation performance. Notably, MobileNetV2 with randomly initialized weights achieves a mean absolute error (MAE) of 2.77, outperforming existing state-of-the-art approaches on FG-NET.

Key Words: Convolutional Autoencoder, Data Augmentation, Deep Learning, Age Estimation.

I. INTRODUCTION

Accurate age estimation from facial images has attracted considerable attention in the fields of computer vision and biometrics. Age information derived from facial appearance is essential for various real-world applications, including age-based access control, personalized services, social media content filtering, and demographic analysis [1]. Compared to categorical tasks such as gender classification or facial expression recognition, age estimation is inherently more challenging due to significant variations in facial features caused by aging, genetics, lifestyle, and environmental factors [2].

In recent years, deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in visual recognition tasks, including age estimation. However, the performance of CNN-based age prediction models remains limited in practice, mainly due to data imbalance issues present in publicly available facial age datasets [3]. One notable example of a publicly available dataset for age estimation is the FG-NET aging database [4], which contains facial images of individuals ranging in age from 0 to 69 years. Despite its popularity in the field, the FG-NET dataset exhibits a significant age-dependent sampling bias, as illustrated in Fig. 1.

As shown in the figure, the number of facial images varies considerably across age groups. There is a high concentration of images for individuals in their infants, teens and

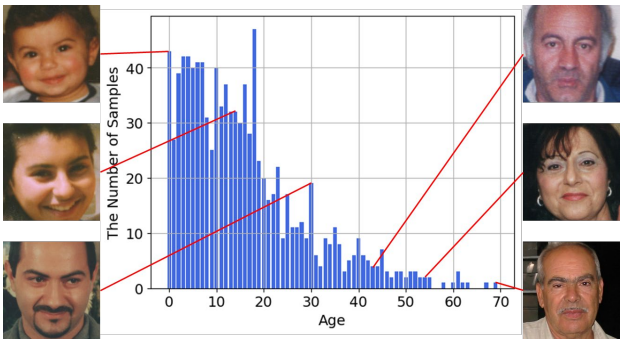

Fig. 1. Age distribution of facial images in the FG-NET aging database. The histogram shows the number of samples for each age, revealing significant sampling bias across age groups. Example facial images from various age ranges are displayed alongside the histogram.

Manuscript received July 31, 2025; Revised September 01, 2025; Accepted September 04, 2025. (ID No. JMIS-25M-07-023)
Corresponding Author (*): Beom Kwon, +82-2-940-4752, bkwon@dongduk.ac.kr
[1]Division of Interdisciplinary Studies in Cultural Intelligence, Dongduk Women's University, Seoul, Korea, bkwon@dongduk.ac.kr

twenties, while the number of samples for elderly individuals (over 60 years) is relatively sparse. This imbalance leads to biased learning and limits the generalization performance of CNN-based models. In particular, the model tends to perform well on age ranges with abundant training data, while prediction accuracy deteriorates for underrepresented age groups [5]. Although various data augmentation techniques have been proposed to mitigate such imbalance in classification tasks [6-8], limited research has explored effective augmentation strategies tailored for regression-based age estimation.

To address this limitation, we propose a targeted data augmentation approach designed to alleviate the effects of age-dependent sampling bias and data imbalance. The method generates synthetic facial images for age groups with limited training samples, thereby improving the performance and robustness of regression-based age estimation models.

The proposed augmentation strategy is based on a convolutional autoencoder (CAE) specifically tailored for facial age estimation. The available facial images are grouped into age intervals of uniform size to divide the full age range of the dataset. A separate CAE is trained for each group to learn latent representations of facial features. During the augmentation phase, new synthetic facial images are generated by feeding the decoder with a convex combination of the latent vectors of two randomly selected images from the same age group. The corresponding age label of each generated image is calculated as a weighted average of the original age labels.

Unlike previous approaches that rely on generative adversarial networks (GANs) or variational autoencoders (VAEs), which often require complex training objectives and are prone to instability, the proposed CAE framework offers a simpler yet effective alternative for generating realistic and structurally consistent synthetic face images within each age group. By learning low-dimensional latent representations from grayscale facial images, our method enables interpolation in the latent space to synthesize new age-preserving samples, thereby mitigating the issue of data imbalance.

In contrast to GAN- and VAE-based augmentation methods, which typically involve explicit modeling of the aging process or identity-to-age transformations, our CAE-based approach synthesizes age-representative samples through latent space interpolation between real faces within the same age group. This design helps preserve realistic facial structure and texture while addressing data imbalance in a task-specific manner.

By applying the proposed data augmentation technique to underrepresented age groups, we effectively alleviate the data imbalance issue and enhance the training process of CNN-based age estimation models. Experimental results demonstrate that our approach significantly improves age prediction accuracy, particularly for age ranges with limited training samples. The main contributions of this study can be summarized as follows:

- We propose a novel data augmentation strategy based on CAEs to address age-dependent data imbalance in facial age estimation tasks. The proposed method generates realistic synthetic facial images through latent space interpolation, enabling effective training of age estimation models in underrepresented age groups.
- We systematically evaluate the proposed augmentation method across six representative CNN architectures, including both randomly initialized and pre-trained models, under the leave-one-person-out (LOPO) cross-validation protocol using the FG-NET dataset. The results consistently demonstrate improved performance, particularly for models trained from scratch.

The remainder of this paper is organized as follows. Section II reviews related work in facial age estimation, including traditional methods, deep learning-based approaches, and data augmentation strategies. Section III introduces the proposed CAE-based data augmentation framework in detail. Section IV describes the experimental setup, dataset characteristics, and evaluation metrics, followed by a comprehensive analysis of the results. Finally, Section V concludes the paper and outlines future research directions.

II. RELATED WORK

Age estimation from facial images has been an active research topic in computer vision, with significant progress made through the application of machine learning and deep learning techniques. Existing methods can be broadly categorized into three groups: (1) handcrafted feature-based approaches, (2) deep learning-based approaches, and (3) data augmentation for age estimation, which have been explored to improve performance under diverse real-world conditions and data limitations.

2.1. Handcrafted Feature-Based Approaches

Early studies on facial age estimation primarily relied on handcrafted feature extraction techniques to describe facial appearance, texture, and geometry. Methods such as local binary patterns (LBP) [9], histogram of oriented gradients (HOG) [10], and local directional and moment patterns (LDMP) [11] were widely adopted to capture age-related facial cues. Although these methods are computationally efficient and interpretable, they often struggle to generalize under unconstrained imaging conditions, including variations in lighting, pose, and facial expressions [12].

In recent years, researchers have explored combining handcrafted features with modern machine learning models

to enhance age estimation performance, especially in scenarios with limited computational resources. For instance, Nagaraju and Reddy [13] proposed a hybrid model that integrates handcrafted features derived from local diagonal extreme patterns (LDEP) with deep features extracted using the Inception-v3 architecture. Their method demonstrated competitive performance on multiple age estimation datasets, highlighting the complementary benefits of combining traditional and deep feature representations.

Additionally, Khalifa and Sengul [14] investigated the fusion of LBP and HOG features with classical machine learning classifiers, including support vector machines (SVM) and k-nearest neighbors (KNN), for age group prediction. Their approach achieved high classification accuracy, reporting up to 99.87% on age estimation datasets, emphasizing the continued relevance of handcrafted features in efficient and interpretable age estimation systems.

These recent studies demonstrate that despite the dominance of deep learning, handcrafted feature-based approaches remain important, particularly for applications with limited hardware capabilities or where model transparency is prioritized.

2.2. Deep Learning-Based Approaches

The advent of deep learning, particularly CNNs, has significantly advanced age estimation tasks. CNN-based models automatically learn hierarchical feature representations from raw facial images, enabling improved performance compared to traditional methods. Several works have explored CNN architectures for age estimation, treating the problem as either a classification, regression, or hybrid task.

Classification-based methods formulate age estimation as a discrete categorization problem by dividing age into pre-defined groups [15]. Levi and Hassner [16] demonstrated the effectiveness of deep CNNs for age and gender classification tasks using unconstrained facial images, highlighting the superiority of deep learning over traditional handcrafted methods. Sheoran et al. [17] proposed an age and gender prediction model based on deep CNNs combined with transfer learning. Their approach utilized pretrained networks to improve performance on relatively small-scale datasets, demonstrating the benefit of leveraging large facial image repositories for knowledge transfer.

Benkaddour [18] presented a CNN-based feature extraction model for age estimation and gender classification tasks. Their work emphasized the role of deep feature representations in boosting classification performance, particularly under varying imaging conditions. Mustapha et al. [19] developed a CNN model tailored for classifying facial images into distinct age groups. Through systematic experimentation, they validated the model's robustness in real-world scenarios where age group boundaries are not always visually clear. More recently, Zhang et al. [20] introduced GroupFace, a method that addresses the severe class imbalance issue in age group classification. By incorporating a multi-hop attention graph convolutional network and group-aware margin optimization, their approach achieved improved accuracy, particularly for underrepresented age groups.

Regression-based approaches aim to predict continuous age values directly from facial images. Niu et al. [21] introduced ordinal regression with CNNs to model the age estimation task, highlighting the potential of multi-output frameworks for improving accuracy. Distance-based regression CNN models [22], advanced loss functions [23], and deep regression forests [24] have further refined continuous age prediction. Notably, Wang et al. [25] proposed an attention-based dynamic patch fusion approach to enhance face-based age estimation, where key facial regions are adaptively emphasized, leading to substantial improvements in regression accuracy.

Hybrid methods combine elements of classification and regression to exploit the strengths of both methodologies. Gao et al. [26] proposed a deep label distribution learning framework that accounts for label ambiguity, effectively blending classification and regression components. Duan et al. [27] introduced a hybrid CNN-ELM model for simultaneous age and gender prediction. Such hybrid approaches often yield better performance, particularly when dealing with the inherent uncertainty of age estimation tasks.

2.3. Data Augmentation for Age Estimation

Despite these advancements, the performance of CNN-based age estimation models remains constrained by the quality and quantity of available training data. Public datasets such as FG-NET, MORPH [28], and UTKFace [29] have been widely utilized for age estimation research. However, these datasets often suffer from severe age imbalance, with a disproportionately low number of images for certain age groups, particularly elderly individuals. This sampling bias leads to degraded model generalization and reduced accuracy for underrepresented age groups.

Data augmentation techniques have been extensively applied to address data scarcity and imbalance issues in computer vision tasks [30]. Traditional augmentation methods include geometric transformations, color perturbations, and image flipping. More recently, generative models such as GANs and autoencoders have been leveraged to synthesize realistic images for tasks like face generation [31], domain adaptation [32], and attribute manipulation [33]. In the context of age estimation, several studies have explored synthetic data generation to alleviate data imbalance.

For instance, Makhmudkhujaev et al. [34] introduced Re-Aging GAN (RAGAN), which achieves personalized face age transformation by compelling the input identity to guide the generation process, resulting in high-quality age-

progressed images. Another method utilizes VAEs for data augmentation. Chadebec and Allassonnière [35] proposed an efficient sampling technique from a VAE in low sample size settings, demonstrating significant improvements in classification tasks by generating synthetic data that enhances model training. Additionally, Alrubaye et al. [36] implemented advanced data augmentation and balancing strategies to improve human age detection using CNNs. By integrating datasets and applying novel augmentation techniques, they achieved a high F1 score, underscoring the importance of data diversity in model performance.

While these methods show promising results, they often require complex training procedures and may suffer from image quality limitations [37]. In this work, we propose a novel data augmentation method utilizing a CAE to generate synthetic facial images through latent space interpolation. Unlike existing GAN-based methods, our approach provides a simpler and more controllable framework for generating realistic images with continuous age labels, effectively addressing the age imbalance problem in facial age estimation.

III. PROPOSED METHOD

This section introduces the proposed data augmentation framework aimed at mitigating age-dependent sampling bias in facial image-based age estimation. The framework comprises two major components: (1) a robust data preprocessing pipeline for facial image normalization, and (2) the generation of synthetic facial images via a CAE. An overview of the proposed methodology is depicted in Fig. 2.

3.1. Data Preprocessing for Facial Image Normalization

In order to minimize the influence of irrelevant factors such as skin tone variations and illumination differences on age estimation, a comprehensive set of preprocessing steps is applied to the input facial images. First, all facial images are converted to grayscale, as using color images may cause the model to rely on skin tone information rather than focusing on structural features that are more directly correlated with age.

Next, to ensure consistent facial alignment across the dataset, face alignment is performed based on facial landmark detection and geometric transformation. The specific type and number of detected landmarks depend on the employed detection algorithm. In this study, we utilize a pre-trained facial landmark detection model that identifies 68 two-dimensional (2-D) facial landmarks per image [4].

Let i be the index of a landmark, where $i \in \{0, 1, \cdots, 66, 67\}$, and let the position of the i^{th} landmark be denoted as $p_i = (u_i, v_i)$. Among the detected landmarks, the left and right pupils correspond to indices 31 and 36, respectively. Their positions are represented as $p_{31} = (u_{31}, v_{31})$ and $p_{36} = (u_{36}, v_{36})$. To correct for head rotation and ensure horizontal alignment of the eyes, the angle θ between the straight line connecting the pupils and the horizontal axis is calculated as:

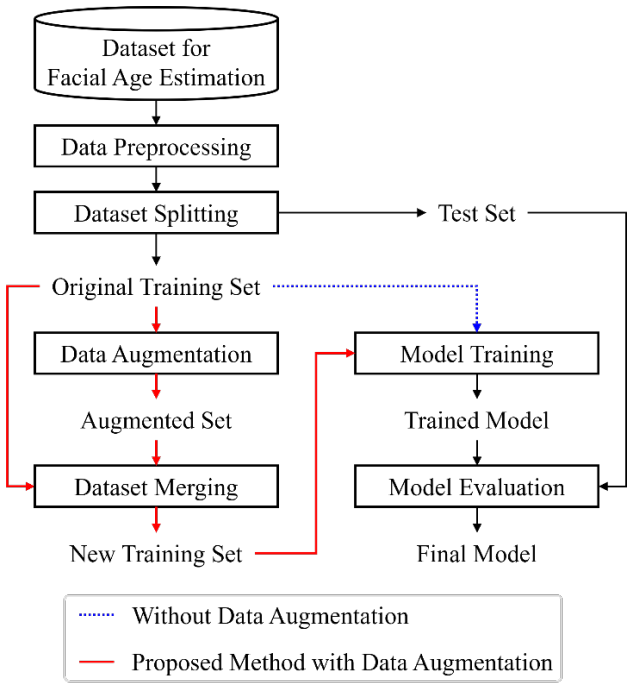

Fig. 2. Overall framework of the proposed age estimation system. The dotted blue arrows indicate the baseline process without data augmentation, while the solid red arrows represent the proposed method with synthetic data generation based on the CAE for mitigating data imbalance.

$$\theta = \arctan\left(\frac{v_{36} - v_{31}}{u_{36} - u_{31}}\right) \times \frac{180}{\pi}. \quad (1)$$

Regarding θ, it represents the degree of head tilt in the facial image. Using (1), an affine matrix $A(\theta)$ is constructed with the center of the image (c_x, c_y) as the rotation pivot:

$$A(\theta) = \begin{bmatrix} \alpha & \beta & (1-\alpha) \times c_x - \beta \times c_y \\ -\beta & \alpha & \beta \times c_x + (1-\alpha) \times c_y \end{bmatrix}, \quad (2)$$

where $\alpha = \cos(\theta)$ and $\beta = \cos(\theta)$.

To align the facial image based on the positions of the pupils, an affine transformation is applied using (2). Let (x, y) represent the coordinates of a pixel in the original image prior to alignment, and let (x', y') denote the corresponding coordinates after the alignment process. The relationship between these coordinates is given by the affine transformation shown in the following equation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A(\theta) \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (3)$$

This transformation, as described in (3), is applied to each pixel in the image to convert its original coordinate (x, y) to the aligned coordinate (x', y'), thereby ensuring consistent orientation across all facial images. After alignment, the facial landmarks are updated by applying the same transformation, ensuring their positions remain consistent with the rotated image:

$$\begin{bmatrix} u_i' \\ v_i' \end{bmatrix} = A(\theta) \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}, \forall i \in \{0, 1, \cdots, 66, 67\}. \quad (4)$$

To localize the facial region within the image, an axis-aligned minimum bounding box is computed by identifying the minimum and maximum u_i' and v_i' coordinates among all detected facial landmarks. This bounding box serves as the basis for cropping the image, enabling the extraction of a tightly aligned facial region for further processing. Through this alignment and cropping process, variations in head tilt, rotation, and image framing are normalized, resulting in standardized facial images suitable for subsequent age estimation tasks.

To mitigate the effects of varying lighting conditions across images, contrast limited adaptive histogram equalization (CLAHE) [38] is applied to the cropped grayscale facial images. CLAHE enhances local contrast while limiting noise amplification, improving the model's robustness to illumination differences. To enforce a square aspect ratio suitable for CNN input, the bounding box is adjusted based on the relative dimensions of height and width. Specifically, if the height exceeds the width, the horizontal boundaries are symmetrically extended; conversely, if the width is greater, the vertical boundaries are expanded. This adjustment ensures that the cropped facial region forms a square while preserving the aspect ratio of the face. As the final preprocessing step, all facial images are resized to a fixed dimension of 48×48 pixels to ensure consistent input size for subsequent CNN-based age estimation.

The entire preprocessing pipeline, including face alignment, contrast enhancement, and resizing, is visualized in Fig. 3 to enhance reproducibility and understanding of each transformation step applied to the facial images.

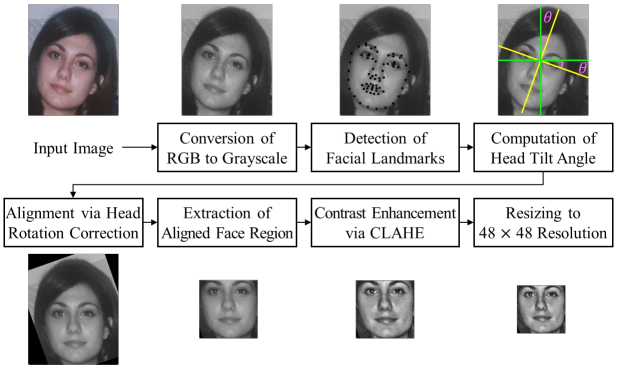

Fig. 3. Step-by-step visualization of the facial image preprocessing pipeline. Starting from the original RGB image, the figure illustrates (1) grayscale conversion, (2) facial landmark detection, (3) head rotation correction based on pupil alignment, (4) face region cropping using the updated landmarks, (5) contrast enhancement using CLAHE, and (6) final resizing to 48×48 pixels. These preprocessing steps ensure structural consistency and mitigate illumination variation prior to age estimation.

3.2. CAE-based Data Augmentation

To mitigate the issue of data imbalance in facial age estimation, particularly for underrepresented age groups, we employ a data augmentation strategy based on a CAE. The CAE consists of an encoder and a decoder. The encoder transforms a preprocessed grayscale facial image $\mathbf{I} \in \mathbb{R}^{48\times48}$ into a compact latent representation $\mathbf{z} \in \mathbb{R}^d$ through a series of convolutional layers and non-linear activations, where d denotes the dimensionality of the latent space ($d = 32$ in our implementation). This encoding process is formally defined as:

$$\mathbf{z} = f_{\text{encoder}}(\mathbf{I}), \quad (5)$$

where $f_{\text{encoder}}(\cdot)$ denotes the encoder network. The decoder reconstructs the original image from the latent code using transposed convolutional layers:

$$\hat{\mathbf{I}} = f_{\text{decoder}}(\mathbf{z}), \quad (6)$$

where $f_{\text{decoder}}(\cdot)$ is the decoder network and $\hat{\mathbf{I}}$ is the reconstructed output.

Unlike GANs, which rely on adversarial training between a generator and a discriminator, or VAEs, which impose a probabilistic prior over the latent space and introduce sampling variability, our CAE is optimized solely using a reconstruction loss. This approach ensures stable training dynamics and better preservation of high-frequency details crucial for capturing age-specific facial features. To this end, the CAE is trained to minimize the pixel-wise reconstruction error between the original and reconstructed facial images, encouraging the encoder–decoder architecture to learn compact and meaningful latent representations that retain age-relevant information.

Let $\{\mathbf{I}^{(j)}\}_{j=1}^{N}$ denote a set of N preprocessed grayscale facial images, and let $\{\hat{\mathbf{I}}^{(j)}\}_{j=1}^{N}$ represent the corresponding reconstructions generated by the CAE. The reconstruction loss is defined as the mean squared error (MSE) between the original and reconstructed images:

$$\mathcal{L}_{\text{recon}} = \frac{1}{N}\sum_{j=1}^{N} \left\| \mathbf{I}^{(j)} - \hat{\mathbf{I}}^{(j)} \right\|_2^2, \quad (7)$$

where $\|\cdot\|_2^2$ denotes the squared Euclidean distance. This loss function ensures that the reconstructed images closely resemble the input images in pixel space, thereby promoting the retention of fine-grained visual characteristics important for age estimation. To reflect the diversity of facial features across age ranges, we train a separate CAE for each five-year age interval, covering the full age range present in typical facial age estimation datasets.

Once the CAE is trained, its encoder and decoder can be leveraged to generate synthetic facial images for data augmentation. Specifically, to create an augmented sample, two images $\mathbf{I}_a$ and $\mathbf{I}_b$ are randomly selected from the same age group. Their corresponding latent representations, $\mathbf{z}_a$ and $\mathbf{z}_b$, are obtained via the encoder:

$$\mathbf{z}_a = f_{\text{encoder}}(\mathbf{I}_a), \;\; \mathbf{z}_b = f_{\text{encoder}}(\mathbf{I}_b). \tag{8}$$

A new latent vector $\mathbf{z}_{\text{aug}}$ is then generated using a convex combination of the two latent vectors:

$$\mathbf{z}_{\text{aug}} = \gamma \times \mathbf{z}_a + (1-\gamma) \times \mathbf{z}_b, \tag{9}$$

where γ denotes the interpolation weight. To maintain sufficient variation while ensuring balanced contribution from both inputs, γ is sampled from a continuous uniform distribution $U(0,1)$ for each augmented sample. The decoder then transforms $\mathbf{z}_{\text{aug}}$ into a new facial image $\hat{\mathbf{I}}_{\text{aug}}$, which resembles a plausible face that lies in-between the two original samples in the latent space:

$$\hat{\mathbf{I}}_{\text{aug}} = f_{\text{decoder}}(\mathbf{z}_{\text{aug}}). \tag{10}$$

To assign an appropriate label to the synthesized image, the corresponding age value is computed as a convex combination of the original age labels:

$$t_{\text{aug}} = \gamma \times t_a + (1-\gamma) \times t_b, \tag{11}$$

where t_a and t_b denote the ground-truth ages associated with the two selected original images used for interpolation. This interpolation strategy not only generates plausible intermediate facial appearances but also produces corresponding age labels that lie within the convex hull of the original values, preserving label consistency for regression training.

This augmentation procedure is repeated until a sufficient number of samples is generated for each underrepresented age group. The generated images preserve realistic facial attributes and offer continuous age labels, enhancing the training diversity of the CNN-based age estimation model. By leveraging latent space interpolation, the proposed CAE-based method provides a controllable and effective solution for data augmentation in regression-based facial age prediction.

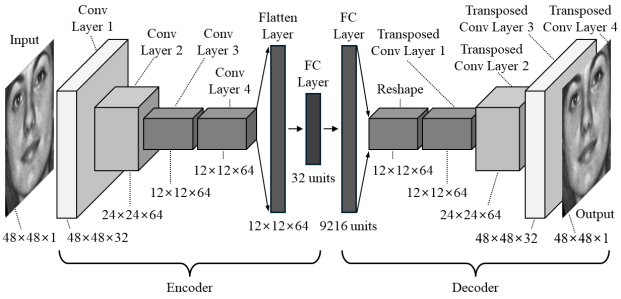

Fig. 4. The structure of the CAE adopted in this study, comprising an encoder, a latent space, and a decoder, designed for generating synthetic age-specific facial images through convex combinations of latent representations.

Fig. 4 shows the architecture of the CAE employed in this study. This CAE architecture is adopted from our previous work on facial emotion recognition, where it was demonstrated to be effective for learning compact and expressive facial representations through reconstruction objectives [39]. As shown in the figure, the CAE architecture is designed to process 48×48 grayscale facial images and encode them into a 32-dimensional latent vector $\mathbf{z} \in \mathbb{R}^{32}$. The encoder comprises four convolutional layers:

- Conv Layer 1: 32 filters, kernel size 3×3, stride 1, padding "same", followed by rectified linear unit (ReLU) activation.
- Conv Layer 2: 64 filters, kernel size 3×3, stride 2, padding "same", followed by ReLU activation.
- Conv Layer 3: 64 filters, kernel size 3×3, stride 2, padding "same", followed by ReLU activation.
- Conv Layer 4: 64 filters, kernel size 3×3, stride 1, padding "same", followed by ReLU activation.

The output is flattened and passed through a fully connected (FC) layer with 32 units, forming the latent representation $\mathbf{z}$.

The decoder mirrors this structure in reverse using transposed convolution layers:

- FC Layer: Input $\mathbf{z} \in \mathbb{R}^{32}$ is mapped to 9,216 units, reshaped into a 12×12×64 tensor.
- Transposed Conv Layer 1: 64 filters, kernel size 3×3, stride 1, padding "same", with ReLU.
- Transposed Conv Layer 2: 64 filters, kernel size 3×3, stride 2, padding "same", with ReLU.
- Transposed Conv Layer 3: 32 filters, kernel size 3×3, stride 2, padding "same", with ReLU.
- Transposed Conv Layer 4: 1 filter, kernel size 3×3, stride 1, padding "same", with ReLU, producing the reconstructed image $\hat{\mathbf{I}}$.

The network is trained to minimize the MSE defined in (7). Optimization is performed using the Adam optimizer

with a default learning rate of 0.001. To accommodate the varying number of training samples across different age groups, the batch size is not fixed but adaptively chosen from {2, 4, 8, 16, 32} based on the size of each group. Smaller batch sizes are used for age groups with limited data, while larger batch sizes are applied to groups with more abundant samples. Training is performed for a fixed number of epochs, and early stopping is applied based on validation loss to prevent overfitting.

The complete process of the proposed CAE-based data augmentation strategy is illustrated in Fig. 5. As shown in the figure, the training phase involves learning a latent representation of facial features using a CAE trained to minimize the reconstruction loss. During the augmentation phase, two images from the same age group are selected, their latent vectors are interpolated, and the decoder generates a new synthetic image that reflects intermediate characteristics. The corresponding continuous age label is assigned using the same convex combination of the original labels. This approach ensures that the generated data remain semantically valid while enhancing the diversity and density of training samples in underrepresented age intervals.

IV. EXPERIMENTAL RESULTS

4.1. Dataset and Experimental Setup

To evaluate the performance of the proposed method, experiments were conducted using the FG-NET aging database [4], a widely adopted benchmark dataset for facial age estimation tasks. The FG-NET dataset consists of 1,002 face images from 82 subjects, with ages ranging from 0 to 69 years. Each subject is represented by multiple images captured at different ages, with an average of approximately 12 age-separated images per person, as shown in Fig. 6. The images in FG-NET exhibit considerable diversity in terms of resolution, lighting conditions, facial expressions, and occlusions such as eyeglasses and facial hair, reflecting the variability typically encountered in real-world scenarios.

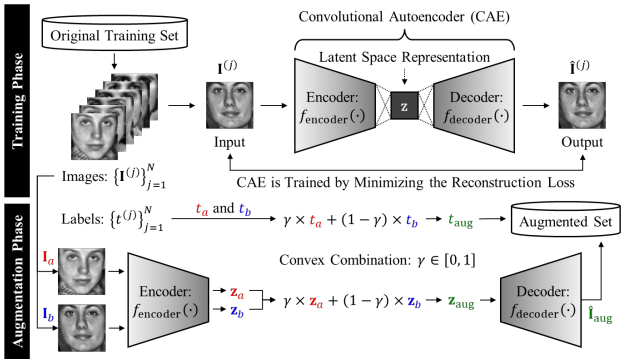

Fig. 5. Overview of the proposed CAE-based data augmentation strategy for facial age estimation, consisting of training and augmentation phases.

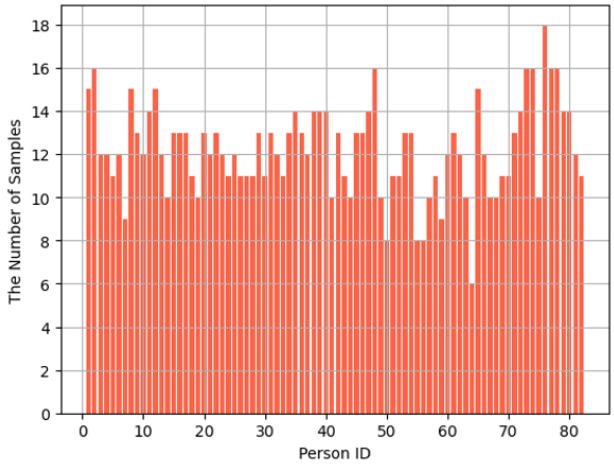

Fig. 6. Distribution of the number of facial images per subject in the FG-NET dataset.

Each image in the dataset is annotated with 68 facial landmark points, enabling detailed shape analysis and alignment operations. In our work, these landmarks are used to preprocess the face images prior to training and augmentation.

To ensure fair and standardized evaluation, we adopt the LOPO cross-validation protocol, which is the de facto standard for experiments conducted on the facial age estimation task. In LOPO, for each of the 82 subjects, the model is trained on the images of the remaining 81 subjects and tested on the held-out subject. This process is repeated until each subject has been used as the test subject once, and the final performance is reported as the average across all trials. This person-independent evaluation strategy effectively assesses the model's generalizability across unseen individuals, which is critical for real-world applicability.

During each fold of the LOPO cross-validation, the proposed CAE-based data augmentation strategy is applied to the facial images of the 81 subjects to generate synthetic samples for age groups with relatively limited data. To address limitations in data availability and facilitate robust training of separate CAE models per age group, the FG-NET dataset is partitioned into 14 age groups, each spanning five years: 0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and 65–69.

This grouping strategy is motivated by two main factors. First, the FG-NET dataset contains several age values (e.g., 56, 57, 59, 64, 65, 66, and 68) for which no facial images are available, making it infeasible to apply age-specific augmentation or model training. Second, the number of available facial images decreases with increasing age, particularly in the older age ranges, as illustrated in Fig. 1. Grouping ages into five-year intervals alleviates the sparsity of training data within individual ages and facilitates more stable and effective learning of the CAE models. The number of facial images contained in each age group is

summarized in Table 1.

As illustrated in Fig. 1 and Table 1, the FG-NET dataset exhibits a significant imbalance in the distribution of samples across age groups. While there is an abundance of samples for younger individuals particularly those under age 20 the number of samples diminishes considerably in older age ranges. To mitigate this imbalance, we employ the proposed CAE-based data augmentation strategy. During each fold of the LOPO cross-validation, we first identify the age group with the largest number of original training samples. For every other age group with fewer samples, synthetic images are generated using the CAE until their sample count matches that of the majority group. Each synthetic image is created by interpolating the latent vectors of two randomly selected training samples from the same age group, with the interpolation weight γ sampled from a continuous uniform distribution $U(0,1)$. Since this procedure is repeated per fold based on the training data composition, the number of generated samples varies dynamically across cross-validation splits. For this reason, we do not provide fixed counts of augmented samples in Table 1.

Fig. 7 presents representative examples of the synthetic facial images generated using the proposed CAE-based augmentation strategy. For each five-year age group, the CAE is trained separately and used to produce synthetic samples by interpolating the latent vectors of two original training images from the same group. As shown in the figure, the generated images retain realistic facial features and smooth age transitions, validating the capability of the proposed method to generate high-quality and age-consistent samples even in age ranges with sparse data. These synthesized images contribute to balancing the age distribution of training data, which is particularly critical for the older age groups.

Table 1. Number of facial images in each five-year age group within the FG-NET dataset. The dataset is partitioned into 14 age groups to mitigate data imbalance and facilitate age-specific model training. The table shows a significant decrease in sample count as age increases, particularly for groups over 40 years.

Age group	Age intervals	No. of samples
1	0–4	193
2	5–9	178
3	10–14	174
4	15–19	165
5	20–24	84
6	25–29	60
7	30–34	46
8	35–39	33
9	40–44	28
10	45–49	18
11	50–54	12
12	55–59	3
13	60–64	6
14	65–69	2

Fig. 7. Examples of synthetic facial images generated by the proposed CAE-based augmentation strategy. For each five-year age group, a separate CAE is trained and used to interpolate between latent vectors of two training images. The resulting synthetic images exhibit realistic age-consistent features and help mitigate age imbalance in the FG-NET dataset.

To quantify the accuracy of age estimation, we adopt the mean absolute error (MAE), a widely used regression metric. Let M be the total number of test samples. For each test image $k \in \{1, 2, \cdots, M\}$, let $t^{(k)}$ denote the ground-truth age and $\tilde{t}^{(k)}$ denote the predicted age. The MAE is then defined as

$$\mathrm{MAE} = \frac{1}{M}\sum_{k=1}^{M}\left|t^{(k)} - \tilde{t}^{(k)}\right|. \tag{12}$$

A lower MAE indicates more accurate age predictions, with zero representing a perfect prediction. By adhering to the LOPO protocol and utilizing established evaluation metric, our experimental setup provides a rigorous framework for assessing the effectiveness of the proposed CAE-based data augmentation strategy in facial age estimation.

4.2. Results

To evaluate the effectiveness of the proposed CAE-based data augmentation strategy, we conducted experiments using a variety of widely used CNN architectures: LeNet-5 [40], AlexNet [41], VGG16, VGG19 [42], ResNet50V2 [43], and MobileNetV2 [44]. These models have been extensively applied in computer vision tasks and offer varying levels of depth and complexity, making them suitable benchmarks for assessing generalizability. All networks were trained using the MSE loss function defined in (7), with a batch size of 32. The Adam optimizer with a learning rate of 0.001 was employed to update network parameters, and early stopping was applied to prevent overfitting by monitoring the validation loss.

For each CNN model, we performed LOPO cross-validation using the FG-NET dataset. The aggregated performance results across all folds are presented in Table 2. These results illustrate the performance improvements achieved through the integration of the proposed data augmentation method across different network architectures.

Table 2. Age estimation performance of six CNN models with and without the proposed CAE-based data augmentation. MAE is reported under the LOPO cross-validation protocol on the FG-NET dataset. "w/o DA" and "w/ DA" indicate training without and with the proposed data augmentation, respectively.

Model	Weights	MAE	
		w/o DA	w/ DA
LeNet-5	Random Initialization	7.40	6.93
AlexNet	Random Initialization	6.56	5.99
VGG16	Random Initialization	3.64	3.21
VGG16	Pre-training on ImageNet	3.89	3.34
VGG19	Random Initialization	3.66	3.17
VGG19	Pre-training on ImageNet	3.79	3.29
ResNet50V2	Random Initialization	3.11	2.99
ResNet50V2	Pre-training on ImageNet	3.37	3.13
MobileNetV2	Random Initialization	2.92	2.77
MobileNetV2	Pre-training on ImageNet	3.03	2.89

As summarized in the table, we evaluated the effectiveness of the proposed CAE-based data augmentation strategy across six representative CNN architectures: LeNet-5, AlexNet, VGG16, VGG19, ResNet50V2, and MobileNetV2. These models were chosen to cover a range of architectural complexities, from shallow to deep, and lightweight designs. To investigate the impact of weight initialization on age estimation performance, we considered two training settings: random initialization and transfer learning via pre-training on the ImageNet dataset.

For LeNet-5 and AlexNet, all network parameters were initialized randomly, and the models were trained from scratch using the training data generated for each fold in the LOPO cross-validation. These relatively shallow networks served as baselines for evaluating the general benefit of data augmentation in low-capacity models. For the deeper architectures VGG16, VGG19, ResNet50V2, and MobileNetV2 we evaluated two variants: (1) random initialization and (2) fine-tuning from pre-trained ImageNet weights. In the fine-tuning setting, all convolutional layers were initialized with weights pre-trained on ImageNet, and only the FC layers were modified. Specifically, the final classification layer was replaced with a regression head consisting of a single neuron to predict continuous age values. The entire network was then fine-tuned end-to-end using our training protocol under the LOPO scheme.

The results in Table 2 demonstrate the consistent benefit of the proposed CAE-based augmentation across all configurations. In the case of LeNet-5 and AlexNet, which relied solely on randomly initialized weights, the MAE decreased from 7.40 to 6.93 and from 6.56 to 5.99, respectively, showing that even shallow models benefit significantly from the enriched data distribution. When examining the deeper networks under random initialization, VGG16, VGG19, ResNet50V2, and MobileNetV2 also showed noticeable improvements (e.g., ResNet50V2 improved from 3.11 to 2.99). While transfer learning from large-scale datasets such as ImageNet often yields performance gains in general visual tasks, its effectiveness in facial age estimation can be limited due to the domain mismatch. ImageNet contains few or no facial images and lacks the fine-grained age-related facial variations required for accurate regression. Consequently, in our experiments, CNNs fine-tuned from ImageNet-pretrained weights (e.g., VGG16, VGG19, ResNet50V2, MobileNetV2) did not outperform their counterparts trained from scratch. For instance, MobileNetV2 achieved an MAE of 2.77 with random initialization, which was slightly better than its ImageNet-pretrained version (2.89). A similar trend was observed for ResNet50V2, where the randomly initialized model outperformed the fine-tuned version.

These findings suggest that, for age estimation, pre-training on a task-specific or domain-relevant dataset would be more beneficial than general-purpose pre-training. Moreover, regardless of the initialization strategy, the proposed CAE-based data augmentation consistently improved model performance by enriching the training data for underrepresented age groups, thereby reducing data imbalance and enhancing generalization in LOPO evaluation.

While transfer learning from large-scale datasets such as ImageNet often yields performance gains in general visual tasks, its effectiveness in facial age estimation can be limited due to two primary factors. First, there exists a significant domain mismatch: ImageNet predominantly comprises object-centric and scene-level images, with few or no human facial images. Consequently, the features learned during pre-training are not well-suited for capturing the subtle age-specific variations in facial appearance required for regression-based age estimation. Second, the relatively small size of the FG-NET dataset further constrains the potential of transfer learning. With only a few hundred training samples per fold in LOPO evaluation, the fine-tuning process may not sufficiently adapt the high-capacity pre-trained networks to the facial domain, leading to overfitting or suboptimal convergence. These observations align with our experimental findings, where randomly initialized networks consistently outperformed their ImageNet-pretrained counterparts, particularly when paired with our task-specific data augmentation.

Table 3 presents a comparative evaluation of facial age estimation models on the FG-NET dataset under the widely adopted LOPO cross-validation protocol. The table aggregates MAE values reported in prior studies that specifically adopted the LOPO setting, thereby ensuring a fair and consistent performance comparison. As shown, most existing methods achieve MAE values ranging from approximately

Table 3. Comparison of facial age estimation performance on the FG-NET dataset using the LOPO cross-validation protocol. Reported MAE values are extracted from existing studies that explicitly disclose LOPO-based evaluation on FG-NET. The final row presents the result of the proposed method using MobileNetV2 with randomly initialized weights and the CAE-based data augmentation strategy, which achieves the lowest MAE among all listed approaches.

Method	MAE
Aging Pattern Subspace (AGES) Algorithm [45]	6.22
Relevance Vector Machine (RVM) [46]	6.2
Regression with Uncertain Nonnegative Labels [47]	5.78
Improved Iterative Scaling (ISS) Algorithm [48]	5.77
Ranking with Uncertain Labels [49]	5.33
Synchronized Submanifold Embedding (SSE) [50]	5.21
LBP Kernel Density Estimate [51]	5.09
Locally Adjusted Robust Regressor (LARR) [52]	5.07
Probabilistic Fusion Approach (PFA) [53]	4.97
Ranking-KNN [54]	4.97
Rank-based Age Value Estimation [55]	4.89
Ordinal Discriminative Features (PLO) [56]	4.82
Biologically Inspired Features (BIF) [57]	4.77
Deep Expectation (DEX) [58]	4.63
Component and Holistic BIF [59]	4.6
Ordinal Hyperplane Ranking (OHRank) [60]	4.48
Biologically Inspired Active Appearance Model [61]	4.18
Mean-Variance Loss [62]	4.1
Deep Regression Forests (DRFs) [24]	3.85
Deep Hybrid-Aligned Architecture (DHAA) [63]	3.72
Adaptive Mean-Residue Loss [64]	3.61
Extended BIF (EBIF) [65]	3.17
Deep Random Forests [66]	3.05
MobileNetV2 (Random Init.)+Proposed CAE-DA	2.77

3.05 to 6.22 years, with performance generally improving in more recent studies that incorporate deep learning architectures or domain adaptation techniques.

Among these, the method by [66] achieves one of the strongest performances with an MAE of 3.05. However, it is noteworthy that even the most competitive prior approaches fail to surpass the 3.0 MAE. The proposed approach, leveraging the CAE-based data augmentation strategy and MobileNetV2 with randomly initialized weights, yields a new state-of-the-art result of 2.77 MAE. This performance not only surpasses all listed methods but also demonstrates the robustness and effectiveness of the proposed augmentation strategy in alleviating data imbalance and improving generalization to unseen individuals a critical requirement under the LOPO protocol. Furthermore, the superior performance is achieved without relying on pretraining with large-scale datasets like ImageNet, suggesting that the model benefits more from task-specific data augmentation than from generic transfer learning. This reinforces the notion that targeted augmentation especially in underrepresented age groups can be more beneficial than domain transfer in age estimation tasks.

V. DISCUSSION

While the proposed CAE-based data augmentation method demonstrates substantial improvements in facial age estimation performance, it is important to acknowledge several limitations inherent to the current design.

First, the effectiveness of the CAE model in generating realistic facial images is highly dependent on the consistency of pose and alignment in the training data. When trained on well-aligned frontal facial images, the CAE reliably reconstructs visually coherent outputs. However, if images with varying poses and orientations are included without proper alignment, the decoder may generate unrealistic faces or structural artifacts due to the entangled representation of pose and identity in the latent space. To mitigate this, we apply a robust face alignment preprocessing step based on facial landmark detection and rotation normalization, as illustrated in Fig. 3. This step helps standardize facial orientation and significantly improves the quality of the generated images.

Second, the latent space interpolation used for synthetic image generation assumes a linear manifold between two encoded representations. While this assumption works reasonably well in practice, it may not capture more complex nonlinear transformations between facial expressions, identity traits, or age progression paths. As a result, the diversity of synthesized samples may be limited, and subtle semantic shifts could be underrepresented.

Finally, since separate CAE models are trained per age group, the quality and stability of each model can be affected by the amount of available data within that group. In extremely underrepresented age intervals, the learned reconstructions may lack sufficient richness or generalizability, despite the use of adaptive batch sizes and regularization.

We consider these limitations to be promising avenues for future research, such as incorporating pose-invariant representation learning, nonlinear interpolation strategies (e.g., spherical or geodesic interpolation), or more expressive generative backbones [67-100].

VI. CONCLUSION

This paper presented a CAE-based data augmentation framework for facial age estimation, targeting the issue of data imbalance inherent in real-world age datasets. By interpolating between latent representations of facial images,

our method generates age-consistent synthetic samples that enrich underrepresented age groups, particularly in the higher age ranges. To maintain semantic fidelity in augmented data, a convex combination of both the latent vectors and corresponding age labels was employed.

We validated the effectiveness of our approach through extensive experiments on the FG-NET dataset using the LOPO cross-validation protocol. The proposed augmentation strategy consistently improved performance across six different CNN architectures. Among them, MobileNetV2 with randomly initialized weights achieved the lowest MAE of 2.77, surpassing previously reported methods and demonstrating that transfer learning from unrelated domains like ImageNet is not always beneficial for fine-grained facial analysis tasks such as age estimation.

Our findings underscore the critical role of tailored data augmentation strategies in alleviating data imbalance and enhancing model generalization in facial analysis tasks. Beyond its performance on FG-NET, the proposed framework is inherently scalable and can be extended to larger, more diverse datasets. Because the method does not depend on domain-specific priors and requires only paired facial images and age labels, it can be applied to datasets with broader demographic distributions, varied poses, or higher resolutions.

Future research will explore the integration of this framework into other facial attribute estimation tasks (e.g., gender or ethnicity) and evaluate its performance on more complex benchmarks. In addition, incorporating adversarial regularization or contrastive objectives into the training process may further improve the realism and diversity of the synthesized faces.

ACKNOWLEDGEMENT

This study was supported by the Dongduk Women's University grant in 2024.

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, no. 42, pp. 1-35, Jun. 2018.

[2] R. R. Atallah, A. Kamsin, M. A. Ismail, S. A. Abdelrahman, and S. Zerdoumi, "Face recognition and age estimation implications of changes in facial features: A critical review study," *IEEE Access*, vol. 6, pp. 28290-28304, May 2018.

[3] P. Punyani, R. Gupta, and A. Kumar, "Neural networks for facial age estimation: A survey on recent advances," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3299-3347, Sep. 2020.

[4] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455, Apr. 2002.

[5] G. Panis and A. Lanitis, "An overview of research activities in facial age estimation using the FG-NET aging database," in *Proceedings of the 13th European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 737-750.

[6] A. Psaroudakis and D. Kollias, "MixAugment & Mixup: Augmentation methods for facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, Jun. 2022, pp. 2367-2375.

[7] J. Yu, Y. Liu, R. Fan, and G. Sun, "Mixcut: A data augmentation method for facial expression recognition," *arXiv Prep. arXiv: 2405.10489*, 2024.

[8] E. Randellini, L. Rigutini, and C. Saccà, "Data augmentation and transfer learning approaches applied to facial expressions recognition," *arXiv Prep. arXiv: 2402. 09982*, 2024.

[9] J. Lu, V. E. Liong, and J. Zhou, "Cost-sensitive local binary feature learning for facial age estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5356-5368, Dec. 2015.

[10] H. Moghadam-Fard, S. Khanmohammadi, S. Ghaemi, and F. Samadi, "Human age-group estimation based on ANFIS using the HOG and LBP features," *Electrical and Electronics Engineering: An International Journal*, vol. 2, no. 1, pp. 21-29, Feb. 2013.

[11] M. Sawant, S. Addepalli, and K. Bhurchandi, "Age estimation using local direction and moment pattern (LDMP) features," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30419-30441, Apr. 2019.

[12] K. ELKarazle, V. Raman, and P. Then, "Facial age estimation using machine learning techniques: An overview," *Big Data and Cognitive Computing*, vol. 6, no. 4, pp. 1-22, Oct. 2022.

[13] K. Nagaraju and M. B. Reddy, "Automated handcrafted features with deep learning based age group estimation model using facial profiles," *Multimedia Tools and Applications*, vol. 83, pp. 42149-42164, Apr. 2024.

[14] T. Khalifa and G. Sengul, "The integrated usage of LBP and HOG transformations and machine learning algorithms for age range prediction from facial images," *Tehnički Vjesnik*, vol. 25, no. 5, pp. 1356-1362, Oct. 2018.

[15] O. Agbo-Ajala and S. Viriri, "Deep learning approach for facial age classification: A survey of the state-of-the-art," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 179-213, Jan. 2021.

[16] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, Jun. 2015, pp. 34-42.

[17] V. Sheoran, S. Joshi, and T. R. Bhayani, "Age and gender prediction using deep CNNs and transfer learning," in *Proceedings of the International Conference on Computer Vision and Image Processing (CVIP)*, Prayagraj, India, Dec. 2020, pp. 293-304.

[18] M. K. Benkaddour, "CNN based features extraction for age estimation and gender classification," *Informatica*, vol. 45, no. 5, pp. 697-703, 2021.

[19] M. F. Mustapha, N. M. Mohamad, G. Osman, and S. H. Ab Hamid, "Age group classification using convolutional neural network (CNN)," in *Proceedings of the International Conference on Mathematics, Statistics and Computing Technology (ICMSCT)*, Bangkok, Thailand, Oct. 2021, pp. 1-11.

[20] Y. Zhang, Y. Shou, W. Ai, T. Meng, and K. Li, "GroupFace: Imbalanced age estimation based on multi-hop attention graph convolutional network and group-aware margin optimization," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 605-619, Dec. 2024.

[21] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, Jun. 2016, pp. 4920-4928.

[22] R. Jumbadkar, V. Kamble, and M. Parate, "Development of facial age estimation using modified distance-based regressed CNN model," *Traitement du Signal*, vol. 42, no. 2, pp. 1041-1056, Apr. 2025.

[23] A. I. Mohammed, S. H. Ali, O. M. S. Hassan, and S. O. Salih, "A deep learning model with a new loss function for age estimation," *Journal of Duhok University*, vol. 26, no. 2, pp. 367-380, 2023.

[24] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. L. Yuille, "Deep regression forests for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, Jun. 2018, pp. 2304-2313.

[25] H. Wang, V. Sanchez, and C. T. Li, "Improving face-based age estimation with attention-based dynamic patch fusion," *IEEE Transactions on Image Processing*, vol. 31, pp. 1084-1096, Jan. 2022.

[26] B. B. Gao, C. Xing, C. W. Xie, J. Wu, and X. Geng, "Deep label distribution learning with label ambiguity," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2825-2838, Jun. 2017.

[27] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448-461, Jan. 2018.

[28] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR)*, Southampton, UK, Apr. 2006, pp. 341-345.

[29] 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)*, Honolulu, HI, Jul. 2017, pp. 5810-5818.

[30] T. Kumar, R. Brennan, A. Mileo, and M. Bendechache, "Image data augmentation approaches: A comprehensive survey and future directions," *IEEE Access*, vol. 12, pp. 187536-187571, Sep. 2024.

[31] A. Melnik, M. Miasayedzenkau, D. Makaravets, D. Pirshtuk, E. Akbulut, and D. Holzmann, et al., "Face generation and editing with StyleGAN: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 3557-3576, May 2024.

[32] Q. Wang, F. Meng, and T. P. Breckon, "Data augmentation with norm-AE and selective pseudo-labelling for unsupervised domain adaptation," *Neural Networks*, vol. 161, pp. 614-625, Apr. 2023.

[33] Y. Liu, Q. Li, Q. Deng, Z. Sun, and M. H. Yang, "GAN-based facial attribute manipulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14590-14610, Dec. 2023.

[34] F. Makhmudkhujaev, S. Hong, and I. K. Park, "Re-aging GAN: Toward personalized face age transformation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Oct. 2021, pp. 3908-3917.

[35] C. Chadebec and S. Allassonnière, "Data augmentation with variational autoencoders and manifold sampling," in *Proceedings of the MICCAI Workshop on Deep Generative Models*, Strasbourg, France, Oct. 2021, pp. 184-192.

[36] M. Alrubaye, A. A. Hameed, and A. Jamil, "Enhancing human age detection: The impact of data augmentation and balancing on CNN performance," in *Proceedings of the International Conference on Intelligent Systems, Blockchain, and Communication Technologies*, Sharm El-Sheikh, Egypt, Jul. 2024, pp. 803-818.

[37] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1-37, Dec. 2022.

[38] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, and T. Greer et al., "Adaptive histogram equalization and its variations," *Computer*

Vision, Graphics, and Image Processing, vol. 39, no. 3, pp. 355-368, Sep. 1987.

[39] B. Kwon, "Data augmentation using convolutional autoencoder for facial emotion recognition," in *Proceedings of the 24th International Conference on Electronics, Information, and Communication (ICEIC)*, Osaka, Japan, Jan. 2025, pp. 1-4.

[40] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, May 2017.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prep. arXiv: 1409.1556*, 2014.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, Sep. 2016, pp. 630-645.

[44] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, Jun. 2018, pp. 4510-4520.

[45] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

[46] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *Proceedings of the 37th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 1529-1532.

[47] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1-8.

[48] X. Geng, C. Yin, and Z. H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2401-2412, Oct. 2013.

[49] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, "Ranking with uncertain labels," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Beijing, China, Jul. 2007, pp. 96-99.

[50] S. Yan, H. Wang, Y. Fu, J. Yan, X. Tang, and T. S. Huang, "Synchronized submanifold embedding for person-independent pose estimation and beyond," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 202-210, Jan. 2008.

[51] J. Ylioinas, A. Hadid, X. Hong, and M. Pietikäinen, "Age estimation using local binary pattern kernel density estimate," in *Proceedings of the International Conference on Image Analysis and Processing (ICIAP)*, Naples, Italy, Sep. 2013, pp. 141-150.

[52] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178-1188, Jul. 2008.

[53] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "A probabilistic fusion approach to human age prediction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Anchorage, AK, Jun. 2008, pp. 1-6.

[54] Y. Liang, X. Wang, L. Zhang, and Z. Wang, "A hierarchical framework for facial age estimation," *Mathematical problems in Engineering*, vol. 2014, p. 242846, Apr. 2014.

[55] L. Zhang, X. Wang, Y. Liang, and L. Xie, "A new method for age estimation from facial images by hierarchical model," in *Proceedings of the 2nd International Conference on Innovative Computing and Cloud Computing (ICCC)*, Wuhan, China, Dec. 2013, pp. 88-91.

[56] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, Jun. 2012, pp. 2570-2577.

[57] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, Jun. 2009, pp. 112-119.

[58] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, no. 2, pp. 144-157, Aug. 2018.

[59] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *Proceedings of the International Conference on Biometrics (ICB)*, Madrid, Spain, Jun. 2013, pp. 1-8.

[60] K. Y. Chang, C. S. Chen, and Y. P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, Jun. 2011, pp. 585-592.

[61] L. Hong, D. Wen, C. Fang, and X. Ding, "A new biologically inspired active appearance model for face age estimation by using local ordinal ranking," in *Proceedings of the 5th International Conference on Internet Multimedia Computing and Service (ICIMCS)*, Huangshan, China, Aug. 2013, pp. 327-330.

[62] H. Pan, H. Han, S. Shan, and X. Chen, "Mean-variance loss for deep age estimation from a face," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, Jun. 2018, pp. 5285-5294.

[63] Z. Tan, Y. Yang, J. Wan, G. Guo, and S. Z. Li, "Deeply-learned hybrid representations for facial age estimation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, Macao, China, Aug. 2019, pp. 3548-3554.

[64] Z. Zhao, P. Qian, Y. Hou, and Z. Zeng, "Adaptive mean-residue loss for robust facial age estimation," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, Jul. 2022, pp. 1-6.

[65] M. Y. El Dib and M. El-Saban, "Human age estimation using enhanced bio-inspired features (EBIF)," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Hong Kong, China, Sep. 2010, pp. 1589-1592.

[66] O. Guehairia, F. Dornaika, A. Ouamane, and A. Taleb-Ahmed, "Facial age estimation using tensor based subspace learning and deep random forests," *Information Sciences*, vol. 609, pp. 1309-1317, Sep. 2022.

[67] B. Kwon, H. Lim, J. Park, and E. Noh, "Machine learning-based path loss prediction with novel diffraction and morphology features," *IEEE Antennas and Wireless Propagation Letters*, vol. 24, no. 7, pp. 2004-2008, Jul. 2025.

[68] H. Jo and B. Kwon, "Facial emotion recognition using canny edge detection operator and histogram of oriented gradients," *Journal of Multimedia Information System*, vol. 12, no. 1, pp. 1-12, Mar. 2025.

[69] H. Lee and B. Kwon, "Facial emotion recognition in children using convolutional neural network with data augmentation," *Journal of the Korea Society of Computer and Information*, vol. 30, no. 2, pp. 21-31, Feb. 2025.

[70] S. Chu and B. Kwon, "Facial emotion recognition using geometric and HOG features," *Journal of Korea Multimedia Society*, vol. 28, no. 1, pp. 112-125, Jan. 2025.

[71] B. Kwon, "Improving BMI classification accuracy with oversampling and 3-D gait analysis on imbalanced class data," *Journal of the Korea Society of Computer and Information*, vol. 29, no. 9, pp. 9-23, Sep. 2024.

[72] H. Song and B. Kwon, "Facial animation strategies for improved emotional expression in virtual reality," *Electronics*, vol. 13, no. 13, pp. 1-18, Jul. 2024.

[73] B. Kwon, "Gait-based gender classification using a correlation-based feature selection technique," *Journal of the Korea Society of Computer and Information*, vol. 29, no. 3, pp. 55-66, Mar. 2024.

[74] B. Kwon and E. Noh, "Path loss prediction using an ensemble learning approach," *Journal of the Korea Society of Computer and Information*, vol. 29, no. 2, pp. 1-12, Feb. 2024.

[75] B. Kwon and H. Son, "Accurate path loss prediction using a neural network ensemble method," *Sensors*, vol. 24, no. 1, pp. 1-20, Jan. 2024.

[76] B. Kwon, "An ensemble learning approach for emotion recognition from gait," in *Proceedings of the 10th Anniversary Korea-Japan Joint Workshop on Complex Communication Sciences (KJCCS)*, Beppu, Japan, Jan. 2024, pp. 1-4.

[77] B. Kwon and T. Oh, "Multi-time window feature extraction technique for anger detection in gait data," *Journal of the Korea Society of Computer and Information*, vol. 28, no. 4, pp. 41-51, Apr. 2023.

[78] B. Kwon and T. Kim, "Toward an online continual learning architecture for intrusion detection of video surveillance," *IEEE Access*, vol. 10, pp. 89732-89744, Aug. 2022.

[79] B. Kwon, J. Huh, K. Lee, and S. Lee, "Optimal camera point selection toward the most preferable view of 3-D human pose," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 533-553, Jan. 2022.

[80] B. Kwon and S. Lee, "Joint swing energy for skeleton-based gender classification," *IEEE Access*, vol. 9, pp. 28334-28348, Feb. 2021.

[81] B. Kwon and S. Lee, "Ensemble learning for skeleton-based body mass index classification," *Applied Sciences*, vol. 10, no. 21, pp. 1-23, Nov. 2020.

[82] B. Kwon and S. Lee, "Human skeleton data augmentation for person identification over deep neural network," *Applied Sciences*, vol. 10, no. 14, pp. 1-22, Jul. 2020.

[83] B. Kwon, M. Gong, and S. Lee, "EDA-78: A novel error detection algorithm for Lempel-Ziv-78 compressed data," *Wireless Personal Communications*, vol. 111, pp. 2177-2189, Apr. 2020.

[84] B. Kwon, H. Song, and S. Lee, "Accurate blind Lempel-Ziv-77 parameter estimation via 1-d to 2-d data conversion over convolutional neural network," *IEEE Access*, vol. 8, pp. 43965-43979, Mar. 2020.

[85] B. Kwon and S. Lee, "Error detection algorithm for Lempel-Ziv-77 compressed data," *Journal of Communications and Networks*, vol. 21, no. 2, pp. 100-112,

Apr. 2019.

[86] B. Kwon and S. Lee, "Cross-antenna interference cancellation and channel estimation for MISO-FBMC/QAM-based eMBMS," *Wireless Networks*, vol. 24, pp. 3281-3293, Nov. 2018.

[87] B. Kwon, M. Gong, J. Huh, and S. Lee, "Identification and restoration of LZ77 compressed data using a machine learning approach," in *Proceedings of the 10th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, HI, Nov. 2018, pp. 1787-1790.

[88] B. Kwon and S. Lee, "Effective interference nulling virtual MIMO broadcasting transceiver for multiple relaying," *IEEE Access*, vol. 5, pp. 20695-20706, Oct. 2017.

[89] B. Kwon, S. Kim, and S. Lee, "Scattered reference symbol-based channel estimation and equalization for FBMC-QAM systems," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3522-3537, Aug. 2017.

[90] B. Kwon, J. Kim, K. Lee, Y. Lee, S. Park, and S. Lee, "Implementation of a virtual training simulator based on 360° multi-view human action recognition," *IEEE Access*, vol. 5, pp. 12496-12511, Jul. 2017.

[91] B. Kwon, M. Gong, and S. Lee, "Machine learning-based compression detection," in *Proceedings of the 32nd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Busan, Korea, Jul. 2017, pp. 164-165.

[92] B. Kwon, M. Gong, and S. Lee, "Novel error detection algorithm for LZSS compressed data," *IEEE Access*, vol. 5, pp. 8940-8947, May 2017.

[93] B. Kwon, S. Kim, D. Jeon, and S. Lee, "Iterative interference cancellation and channel estimation in evolved multimedia broadcast multicast system using filter-bank multicarrier-quadrature amplitude modulation," *IEEE Transactions on Broadcasting*, vol. 62, no. 4, pp. 864-875, Dec. 2016.

[94] B. Kwon, J. Kim, and S. Lee, "An enhanced multi-view human action recognition system for virtual training simulator," in *Proceedings of the 8th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Jeju, Korea, Dec. 2016, pp. 1-4.

[95] B. Kwon, D. Jeon, J. Kim, J. Kim, D. Kim, and H. Song et al., "Framework implementation of image-based indoor localization system using parallel distributed computing," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 41, no. 11, pp. 1490-1501, Nov. 2016.

[96] B. Kwon, J. Park, and S. Lee, "Virtual MIMO broadcasting transceiver design for multi-hop relay networks," *Digital Signal Processing*, vol. 46, pp. 97-107, Nov. 2015.

[97] B. Kwon, S. Kim, H. Lee, and S. Lee, "A downlink power control algorithm for long-term energy efficiency of small cell network," *Wireless Networks*, vol. 21, pp. 2223-2236, Oct. 2015.

[98] B. Kwon, D. Kim, J. Kim, I. Lee, J. Kim, and H. Oh, et al., "Implementation of human action recognition system using multiple Kinect sensors," in *Proceedings of the 16th Pacific-Rim Conference on Multimedia (PCM)*, Gwangju, South Korea, Sep. 2015. pp. 334-343.

[99] B. Kwon, J. Park, and S. Lee, "A target position decision algorithm based on analysis of path departure for an autonomous path keeping system," *Wireless Personal Communications*, vol. 83, pp. 1843-1865, Aug. 2015.

[100] B. Kwon and Y. W. Chung, "An improved energy saving scheme in IEEE 802.16e," *The Journal of Korean Institute of Information Technology*, vol. 10, no. 8, pp. 43-51, Aug. 2012.

AUTHORS

Beom Kwon received the B.S. degree in Electrical and Electronic Engineering from Soongsil University, Seoul, Republic of Korea, in 2012, and the M.S. and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, in 2018. From March 2018 to September 2019, he was a Senior Researcher at the Agency for Defense Development (ADD), Daejeon, Republic of Korea. From October 2019 to August 2021, he was a Staff Engineer at Samsung Electronics Company, Ltd., Suwon City, Gyeonggi Province, Republic of Korea. From September 2021 to August 2023, he was an assistant professor in the Department of Artificial Intelligence at Dongyang Mirae University, Seoul. Since September 2023, he has been an assistant professor in the Division of Interdisciplinary Studies in Cultural Intelligence (Data Science) at Dongduk Women's University, Seoul. His research interests include artificial intelligence and its applications.