

Article

LAE-GAN-Based Face Image Restoration for Low-Light Age Estimation

Se Hyun Nam , Yu Hwan Kim , Jiho Choi, Seung Baek Hong, Muhammad Owais  and Kang Ryoung Park  *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro, 1-gil, Jung-gu, Seoul 04620, Korea; nsh6473@dongguk.edu (S.H.N.); taekkuon@dongguk.edu (Y.H.K.); choijh1027@dongguk.edu (J.C.); baek2sm@dongguk.edu (S.B.H.); owais2018@dongguk.edu (M.O.)

* Correspondence: parkgr@dongguk.edu

Abstract: Age estimation is applicable in various fields, and among them, research on age estimation using human facial images, which are the easiest to acquire, is being actively conducted. Since the emergence of deep learning, studies on age estimation using various types of convolutional neural networks (CNN) have been conducted, and they have resulted in good performances, as clear images with high illumination were typically used in these studies. However, human facial images are typically captured in low-light environments. Age information can be lost in facial images captured in low-illumination environments, where noise and blur generated by the camera in the captured image reduce the age estimation performance. No study has yet been conducted on age estimation using facial images captured under low light. In order to overcome this problem, this study proposes a new generative adversarial network for low-light age estimation (LAE-GAN), which compensates for the brightness of human facial images captured in low-light environments, and a CNN-based age estimation method in which compensated images are input. When the experiment was conducted using the MORPH, AFAD, and FG-NET databases—which are open databases—the proposed method exhibited more accurate age estimation performance and brightness compensation in low-light images compared to state-of-the-art methods.



Citation: Nam, S.H.; Kim, Y.H.; Choi, J.; Hong, S.B.; Owais, M.; Park, K.R. LAE-GAN-Based Face Image Restoration for Low-Light Age Estimation. *Mathematics* **2021**, *9*, 2329. <https://doi.org/10.3390/math9182329>

Academic Editors: Daniel Gómez Gonzalez, Javier Montero and Tinguaro Rodriguez

Received: 10 August 2021
Accepted: 16 September 2021
Published: 19 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: age estimation; low-illumination image enhancement; LAE-GAN; CNN

1. Introduction

A human face contains biological information showing various attributes, such as identity, age, gender, emotions, and expressions. Numerous researchers have studied face recognition [1,2], facial expression recognition [3], gender classification [4], facial skin assessment [5], and age estimation [6] by analyzing such information. Specifically, age estimation has a wide range of applications in commercial areas, such as customer prediction and preference surveys according to age, security for controlling access based on age and statistical fields such as age surveys of an audience [6]. However, age estimation using human facial images entails several problems, including the uncontrollable, natural aging process, individual aging patterns, and large inter-class similarity and intra-class variation of subjects' images within age classes [7]. For overcoming these drawbacks, image representation techniques such as the active appearance model (AAM) [8], the active shape model (ASM) [9], the aging pattern subspace model (AGES) [10], feature extraction techniques such as Gabor filters [11], linear discriminant analysis (LDA) [12], and local binary patterns (LBP) [13] have been used in the past. The representative image and extracted features are applied with multi-classification, regression, and hierarchical approaches for age estimation [14]. However, since the emergence of deep learning, where feature extraction and learning are both involved in the process, using a convolutional neural network (CNN) has become popular in age estimation.

Previous studies on age estimation used clear facial images taken during the daytime with high illumination. However, in reality, most of the images are captured in low-light

environments [15,16]. In general, these low-illumination images have a lesser amount of light and a longer exposure time of a camera than images taken during the daytime. Therefore, motion and optical blurs are generated in images, and noise increases in the images due to the characteristics of camera sensors [17,18]. For resolving the problems of low-illumination images captured in low-light or nighttime environments, hardware approaches using a high-performance charge-coupled device (CCD), a complementary metal-oxide-semiconductor (CMOS), low-light compensation circuits, and filters—or using software algorithms to provide flexibility in images and improve the quality of low-light images—can be applied [19]. Hardware approaches, however, increase the cost of the camera and cannot be applied to general cameras; hence, improving software algorithms to enhance image quality is a more practical approach. Existing software algorithms can be classified into: gray transformation methods, histogram equalization, Retinex filtering, frequency-domain methods, image fusion methods, defogging model methods, and machine learning-based methods [19]. Excluding machine learning-based methods, a majority of the methods are conventional image processing-based techniques. In recent years, studies have been actively conducted using machine learning-based techniques in which deep learning based enhancements are quickly gaining attention. Low-light images taken in low-light environments are currently utilized for various purposes, while research on compensating for the aforementioned problems is also actively being conducted. However, there has been no study conducted on compensating low light facial images for age estimation; thus, this study proposes a method for compensating low-light facial images for realistic age estimation.

Instead of researching age estimation with clear images with high illumination taken in a high light or daytime environment, this study performs age estimation using low-light images. A generative adversarial network for low-light age estimation (LAE-GAN) is proposed for removing blur and noise generated due to low illumination and restoring the lost age information based on which low-light images are compensated. Age is then estimated by applying a CNN to the compensated images. Our research is novel in the following four ways compared to previous works:

- It is the first study on age estimation considering low light;
- Without separately applying pre-processing to low-light facial images, images are enhanced using LAE-GAN, which is proposed in this study;
- In LAE-GAN, identity information of input data was preserved by removing an input random noise vector used in a conventional conditional GAN and adding an L2 loss function in the generator. Furthermore, high frequency information of the input image delivered through a skip-connection using a leaky rectified linear unit (ReLU) to the 6th and 7th decoder blocks of the generator was reinforced, and the ReLU was used in the 4th convolution layer of the discriminator;
- Through [20], the trained LAE-GAN and CNN for age estimation are disclosed to be fairly evaluated by other researchers in terms of performance.

This paper is organized as follows. In Section 2, previous studies on low-light image enhancement and facial image age estimation are analyzed and compared with the proposed method. In Section 3, the LAE-GAN proposed for low-illumination facial image enhancement and CNNs for facial image age estimation are explained. In Section 4, the results of the experiment conducted using the method proposed in Section 3 are comparatively analyzed and discussed. Lastly, Section 5 proposes conclusions.

2. Related Works

Age estimation using human facial images is performed by extracting features based on the length, depth, and number of wrinkles, which change over time due to aging and skin condition [21]. Therefore, age estimation involves feature extraction and age learning steps for learning ages based on the extracted represented image. For feature extraction in previous studies, image representation techniques such as AAM [8], ASM [9], and AGES [10] as well as Gabor filters [11], LDA [12], and LBP [13] were applied; multi-class

classification, regression, and hierarchical approaches were taken for age learning. Recently, however, methods using a CNN where feature extraction and age learning proceed end-to-end are more commonly used. Table 1 presents previous studies on age estimation in which deep learning was employed. In previous studies, the mean absolute error (MAE) was used to evaluate the accuracy of age estimation. MAE is the mean absolute error between the estimated and ground-truth ages, and a detailed description of MAE can be found in Equation (14) of Section 4.3.

A study [22] proposed a simple CNN consisting of six layers: three convolutional layers, two pooling layers, and one fully connected layer. The dimension of these extracted features was compressed with principal component analysis (PCA), and age learning was performed using a support vector machine (SVM). A study [23] proposed a CNN consisting of three convolutional layers, two fully connected layers, and one output layer. Most age estimation using a CNN involves a shallow CNN; subsequently, a study [24] improved the performance through fine-tuning deep networks such as a visual geometry group (VGG)-16 [25] and databases such as IMDB-WIKI and ImageNet. In a study [26], a network consisting of the following three steps was proposed for mitigating the learning restricted by a dataset during age estimation: first, data are classified into age groups using an age group classifier; second, age is estimated using the mean value within the age groups; third, errors are revised using the predicted age. In one study [27], a method for predicting age based on rankings between sub-networks was proposed using a network tied with sub-networks predicting a single age label as binary outputs. Likewise, most previous studies conducted age estimation based on various types of databases and networks. However, no study has examined age estimation considering low light, which is more likely to occur.

Table 1. Comparison of previous research on age estimation using deep learning (N.A. means “not available”).

Method	Database	MAE	Accuracy (%)
Wang et al. [22]	MORPH	4.77	N.A.
	FG-NET	4.26	
Levi et al. [23]	Adience	N.A.	84.7
Huerta et al. [28]	MORPH II	4.25	
	FRGC	4.17	
Liu et al. [29]	ICCV2015	3.33	N.A.
Huo et al. [30]	ChaLearn LAP 2016	1.75	
	ICCV2015	N.A.	
Chen et al. [26]	FG-NET	3.49	88.45
	MORPH	N.A.	
Yang et al. [31]	MORPH II	3.23	98.8
Niu et al. [32]	MORPH II	3.27	N.A.
	AFAD	3.34	
Hu et al. [33]	FG-NET	2.8	N.A.
	MORPH	2.78	
Chen et al. [27]	MORPH	2.96	92.9
Li et al. [34]	MORPH II	3.06	N.A.
	WebFace	6.04	
Qawaqneh et al. [35]	Adience	N.A.	62.37
Rodriguez et al. [36]	MORPH II	2.56	61.8 N.A.
	Adience	N.A.	
Duan et al. [37]	MORPH II	3.44	
Wan et al. [38]	CACD	5.22	N.A.
	MORPH II	2.93	
	ChaLearn Lap 2016	3.30	

Table 1. *Cont.*

Method	Database	MAE	Accuracy (%)
Zaghbani et al. [39]	MORPH II	3.34	
	FG-NET	3.75	
Yoo et al. [40]	MORPH II	2.89	
	FG-NET	3.46	
Rattani et al. [41]	Adience	N.A.	80.96
Taheri et al. [42]	MORPH II	3.17	N.A.
	FG-NET	3.29	
Taheri et al. [7]	MORPH II	2.81	
	FG-NET	3.05	

Noise and blur are generated when acquiring low-illumination images in general, which leads to performance degradation in various computer vision fields that rely on human facial images. For enhancing these low-illumination images, the methods employed in previous research can be classified into image processing-based techniques such as histogram equalization methods and Retinex filtering, machine learning-based techniques, and deep learning-based techniques [19]. Well-known cases of image processing-based techniques for improving low illumination in facial images are as follows. In a study [43], a method was proposed for enhancing illumination imbalance using discrete cosine transform (DCT) low frequency coefficients after applying histogram equalization to facial images. In a study [44], adaptive region-based image processing was suggested for compensating low-illumination images that appear differently depending on various lighting conditions. After partitioning an image into various regions according to lighting conditions, contract and edges were used in adaptive region-based histogram equalization. In a study [45], a selective illumination enhancement technique (SIET) was proposed for enhancing low-illumination facial images. SIET was utilized for improving changes in facial images due to the effects of non-uniform illumination; dark regions were isolated and compensated with a correction factor that was determined based on an energy function to enhance illumination. Image processing-based techniques were more commonly used for conventional low-illumination image enhancement than any other techniques. Several studies have been conducted in recent years, as interests in machine learning-based techniques and deep learning-based techniques are on the rise [46–48]. In a study [46], enhancement networks are proposed for preventing performance degradation of facial images being used for the mobile face unlock feature in low-light environments. Networks typically consist of a decomposition part for partitioning input low-illumination facial images into face normals and face albedos, and a reconstruction part for enhancing and reconstructing images using spherical harmonic lighting coefficients. In a study [47], a feature reconstruction network was proposed in which raw face images and illumination-enhanced face images were all used in deep learning-based techniques for face recognition in low illumination. A study [48] proposed REGDet, in which a recurrent exposure generation (REG) module for low-illumination enhancement is combined with a multi exposure detection (MED) module for face detection in low-light environments. These studies on improving low-light conditions are utilized in various fields, not only for facial images. In recent years, a GAN-based method has been actively researched, where data distribution of low-light input images is converted into the data distribution of high-light target images [49–51].

However, no study has examined age estimation considering low-light conditions. This study, therefore, proposes a LAE-GAN-based age estimation method where low-illuminated face images are enhanced, which are then subsequently used as input for a CNN.

Table 2 presents the comparison between the proposed method and previous studies in which low-light facial images were enhanced.

Table 2. Comparisons between the proposed method and previous studies in which low-light facial images were enhanced.

Categories	Method	Application	Database	Strength	Weakness
Image processing -based techniques	Vishwakarma et al. [43]	Face recognition	Yale Face B	Face recognition is robust to the low-illumination problem	When the environment changes, parameters for enhancement of low-light images need to be manually revised. Does not consider low-illumination images for age estimation.
	Du et al. [44]		Yale Face B, Carnegie Mellon database		
	Vidya et al. [45]		ORL, UMIST, Yale Face B, Extended Yale B, and color FERET		
	Maeng et al. [52]		LDHF database		
	Baradarani et al. [53]		Yale Face B, Extended Yale B, CMU-PIE, FERET, AT&T, and Labeled Face in the Wild (LFW)		
Machine learning -based techniques	Kang et al. [54]	Face detection	LDHF database	Face detection robust to the low illumination problem	Training data for restoration of low-light images, face detection, and recognition need to be trained. Does not consider low-illumination images for age estimation
	Liang et al. [48]		DARK FACE database		
	Shen et al. [55]		Self-constructed database		
Deep learning -based techniques	Cho et al. [56]	Face recognition	Self-constructed database	Face recognition robust to the low illumination problem	Additional procedure for the training of LAE-GAN is necessary
	Le et al. [46]		Self-constructed database		
	Huang et al. [47]		SoF database		
LAE-GAN (proposed method)		Age estimation	MORPH, FG-NET, and AFAD	Age estimation robust to the low illumination problem	

3. Proposed Method

3.1. Overview of the Proposed Method

The age estimation method proposed in this study, which is effective for low-illumination facial images, proceeds according to the four steps shown in Figure 1. The first and second steps are pre-processing for age estimation using facial images effective for low light. In the first step, the face and eye positions are detected in facial images using an adaptive boosting (Adaboost) algorithm [57]. The detected positions become the reference points for aligning facial images in the second step in order to compensate through in-plane rotation and redefine face region of interest (ROI). The pre-processing step is explained in detail in Section 3.2. The pre-processed facial images are input in the third step to LAE-GAN, which has been trained with pairs of low- and high-illumination facial image for low-illumination image enhancement. Finally, the enhanced facial images are used to the trained CNN for age estimation.

3.2. Pre-Processing

In general, the facial region is not aligned in the captured human facial images, which contain parts without age information such as the background. Misalignment in facial images affects the age estimation performance [58]. Therefore, pre-processing, as shown in Figure 2, was performed in this study. First, the Adaboost algorithm [57] is used to detect the face region in the image. Within the detected face region, the exact eye position is detected by designating an exploratory region where eyes may be located. The explored positions of the face and eyes are as shown in Figure 2b, and they are used for the redefinition of ROI and in-plane rotation compensation. Here, Equation (1) is used to proceed with in-plane rotation compensation based on the estimated in-plane rotation angle and bilinear interpolation; then, ROI of the human facial image is redefined with

respect to the center of both eyes for removing the background image. In Equation (1), R_x and R_y are horizontal and vertical positions of the right eye, while L_x and L_y are horizontal and vertical positions of the left eye. The pre-processed image has the size of $256 \times 256 \times 3$ as shown in Figure 2c.

$$\theta = \tan^{-1} \left(\frac{R_y - L_y}{R_x - L_x} \right) \quad (1)$$

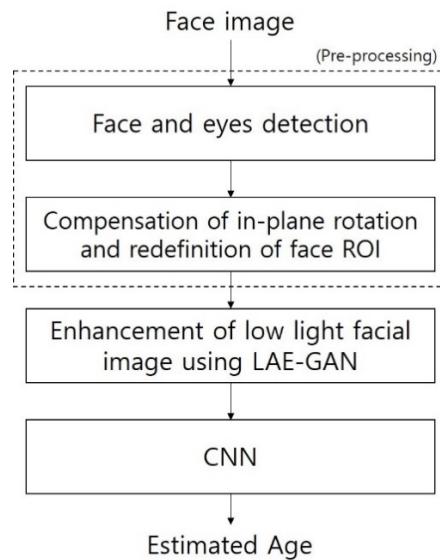


Figure 1. Overall procedure of the proposed method.

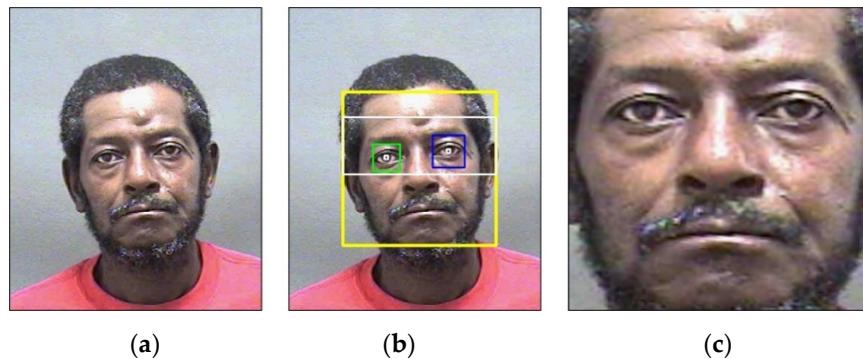


Figure 2. Procedure of pre-processing of the face region. (a) Original PAL database image. (b) Detected face and eye regions using the Adaboost algorithm. (c) In-plane rotation compensation and face ROI redefinition.

3.3. Enhancement of Low-Illuminated Face Image by LAE-GAN

This study proposes a method for compensating a low-illuminated face image using LAE-GAN for age estimation, which is effective for low-light conditions. A conventional conditional GAN [59] performs adversarial learning using paired GAN based on a pair of input and target images. It consists of a generator, which outputs a generated image I^{Out} by receiving the random noise vector z and input image I^{In} , and a discriminator, which distinguishes between real and fake images by receiving I^{In} and I^{Out} or the target image I^{Target} as input. In adversarial learning, the generator tries to deceive the discriminator by generating a realistic image I^{Out} . The discriminator tries to distinguish between the generated image I^{Out} and the target images I^{Target} . The generator has an encoder-decoder structure. The encoder extracts the features of the input image I^{In} , and a decoder maps the patches corresponding to the extracted features. Such learning requires that the data distribution of I^{In} is converted to the distribution of I^{Target} using the loss function shown

in Equation (2) below, where G is the generator, D is the discriminator, \log is the decimal logarithm, and \mathbb{E} is an expected value (mean value).

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{I^{In}, I^{Target}} [\log D(I^{In}, I^{Target})] + \mathbb{E}_{I^{In}, z} [\log(1 - D(I^{In}, G(I^{In}, z)))] \quad (2)$$

This study proposes LAE-GAN for compensating a low-illumination facial image to a corresponding high-illumination facial image. In a study [59], the random noise vector z allows image transformation to be easier and more diverse. The random noise vector z in this study, however, simply acts as noise when compensating from low-illumination facial image I^{In} to high-illumination facial image I^{Out} . Therefore, the loss function after removing random noise vector z is as shown in Equation (3) below.

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{I^{In}, I^{Target}} [\log D(I^{In}, I^{Target})] + \mathbb{E}_{I^{In}} [\log(1 - D(I^{In}, G(I^{In})))] \quad (3)$$

Due to the nature of adversarial learning of the generator and discriminator explained above, the generator aims to deceive the discriminator by generating I^{In} into I^{Out} image having a similar distribution as I^{Target} . This tendency can be trained so as to deceive the discriminator rather than following the data distribution of I^{Target} . Hence, this study adds the new $L2$ loss function, as shown in Equation (4), to the generator for maintaining the identity of the I^{Target} image.

$$\mathcal{L}_{L2}(G) = \mathbb{E}_{I^{In}, I^{Target}} [(I^{Target} - G(I^{In}))^2] \quad (4)$$

Ultimately, the final loss function used in this study is as shown in Equation (5) below. λ is the regularization term. The optimal λ was experimentally determined as 0.9 with training data, which showed the highest accuracy of age estimation with training data. $\arg \min_G \max_D$ represent the arguments of the generator and discriminator, which minimize and maximize the loss functions of the generator and discriminator, respectively.

$$\mathcal{L} = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L2}(G) \quad (5)$$

3.3.1. Generator

The encoder-decoder structure is one of the networks used for generating images [60,61]. U-net [62] is one of the commonly used networks and consists of an encoder for extracting features and a decoder for mapping a patch corresponding to the extracted features of U-net; however, it has a skip connection for preserving the high frequency information of the input image. A skip connection is present between the i^{th} layer and $(n - i)^{th}$ layer of U-net, and concatenates the features extracted in the i^{th} layer to the $(n - i)^{th}$ layer. Therefore, it preserves the high frequency information of the input image as well as the original shape and detail. The U-net generator was used in this study, and its detailed structure is represented in Table 3 below and Figure 3a.

Each encoder consists of blocks comprised of a convolution layer, a batch normalization layer, and a leaky ReLU layer excluding the first encoder since the first encoder does not include a batch normalization layer. Each decoder consists of decoder blocks comprised of a deconvolution layer, a batch normalization layer, and a ReLU layer excluding the sixth, seventh, and last decoders. Concatenation occurs from the skip connection after batch normalization. The sixth and seventh decoder blocks emphasize the features of high frequency information delivered through skip connection using a leaky ReLU layer. The deconvolution layer uses transpose convolution, and the last decoder block consists of $tanh$ function.

Table 3. Generator structure using U-net in LAE-GAN.

Layer Name	Number of Filters	Size of Feature Map (Height × Width × Channel)	Filter Size (Height × Width)	Stride (Height × Width)	Padding (Height × Width)	
Input image		256 × 256 × 3				
1st convolutional layer Leaky ReLU layer	64	128 × 128 × 64	4 × 4 × 3	2 × 2	1 × 1	
2nd convolutional layer Batch normalization Leaky ReLU layer	128	64 × 64 × 128	4 × 4 × 64	2 × 2	1 × 1	
3rd convolutional layer Batch normalization Leaky ReLU layer	256	32 × 32 × 256	4 × 4 × 128	2 × 2	1 × 1	
Encoder	4th convolutional layer Batch normalization Leaky ReLU layer	512	16 × 16 × 512	4 × 4 × 256	2 × 2	1 × 1
	5th convolutional layer Batch normalization Leaky ReLU layer	512	8 × 8 × 512	4 × 4 × 512	2 × 2	1 × 1
	6th convolutional layer Batch normalization Leaky ReLU layer	512	4 × 4 × 512	4 × 4 × 512	2 × 2	1 × 1
	7th convolutional layer Batch normalization Leaky ReLU layer	512	2 × 2 × 512	4 × 4 × 512	2 × 2	1 × 1
	8th convolutional layer Batch normalization Leaky ReLU layer	512	1 × 1 × 512	4 × 4 × 512	2 × 2	1 × 1
	1st deconvolutional layer Batch normalization Concatenation ReLU layer	512	2 × 2 × 512 2 × 2 × 1024	4 × 4 × 512	2 × 2	1 × 1
	2nd deconvolutional layer Batch normalization Concatenation ReLU layer	512	4 × 4 × 512 4 × 4 × 1024	4 × 4 × 1024	2 × 2	1 × 1
	3rd deconvolutional layer Batch normalization Concatenation ReLU layer	512	8 × 8 × 512 8 × 8 × 1024	4 × 4 × 1024	2 × 2	1 × 1
Decoder	4th deconvolutional layer Batch normalization Concatenation ReLU layer	512	16 × 16 × 512 16 × 16 × 1024	4 × 4 × 1024	2 × 2	1 × 1
	5th deconvolutional layer Batch normalization Concatenation ReLU layer	256	32 × 32 × 256 32 × 32 × 512	4 × 4 × 1024	2 × 2	1 × 1
	6th deconvolutional layer Batch normalization Concatenation Leaky ReLU layer	128	64 × 64 × 128 64 × 64 × 256	4 × 4 × 512	2 × 2	1 × 1
	7th deconvolutional layer Batch normalization Concatenation Leaky ReLU layer	64	128 × 128 × 64 128 × 128 × 128	4 × 4 × 256	2 × 2	1 × 1
	8th deconvolutional layer Tanh	3	256 × 256 × 3	4 × 4 × 128	2 × 2	1 × 1
	Generated image		256 × 256 × 3			

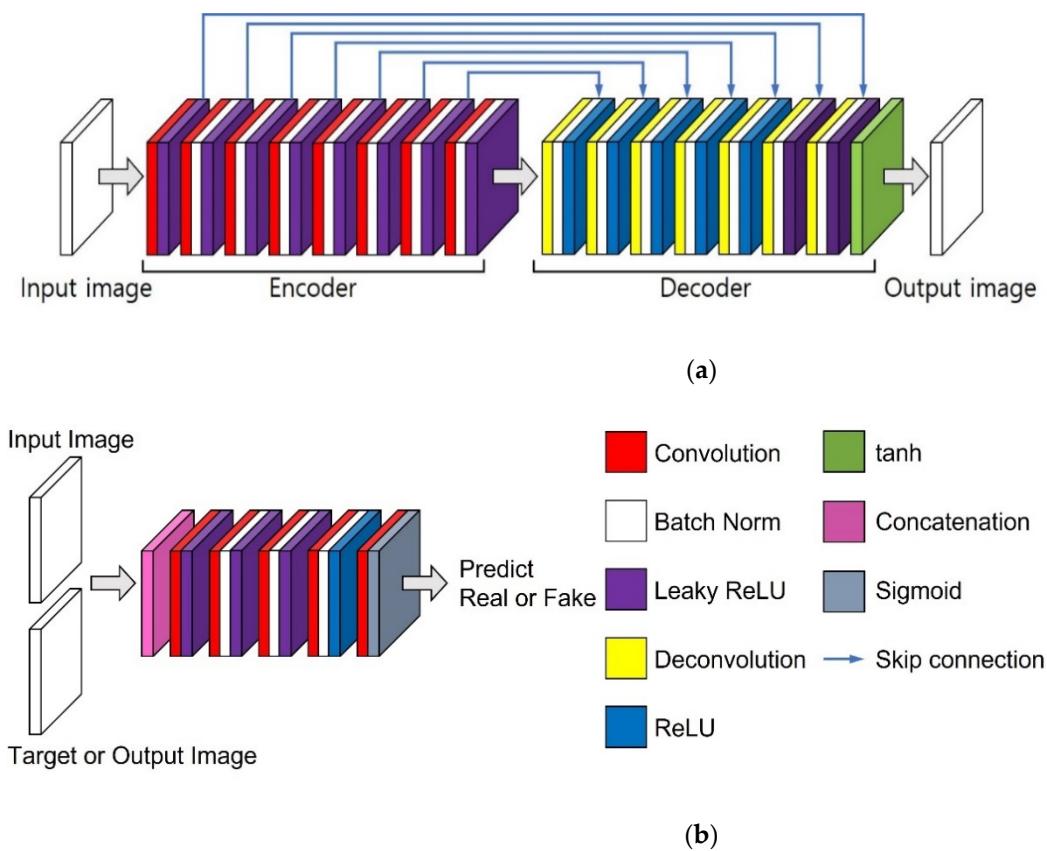


Figure 3. The structure of LAE-GAN. (a) Generator. (b) Discriminator. The generator uses a U-net architecture that combines an encoder-decoder structure with a skip-connection. The discriminator consists of convolutional layers that receive a concatenated image. Detailed explanations are given in Sections 3.3.1 and 3.3.2.

3.3.2. Discriminator

The discriminator in this study concatenates I^{Target} and I^{Out} that are randomly input with I^{In} through convolution layers and proceeds with feature extraction to generate a feature map of $30 \times 30 \times 1$ in the last layer. The generated feature map can be considered as a set of $1 \times 1 \times 1$ grids. The grids are used to analyze local information of a 70×70 receptive field instead of global information in which the local information that may be lost in the global information is utilized to adequately express detail and shape of the image. Therefore, such learning can reduce blurry results rather than applying $L1$ loss or $L2$ loss to the entire features; further, the information of the original image can be preserved as much as possible. For maintaining the disposition of the original image and discerning the authenticity of the input image, the discriminator consistently receives I^{In} as input. The features extracted from I^{In} will express the information that the image must consistently maintain and thus prevent improper learning of the generator between adversarial learning. The detailed structure of the discriminator is presented in Table 4 and Figure 3b.

3.4. Difference of Conditional GAN

The LAE-GAN proposed in this study has the following differences from the conventional conditional GAN [59]:

- A random noise vector was used in the conventional conditional GAN for inducing image transformation, but it has been removed in this study as it has a stronger negative effect than noise in a 1:1 mapping structure between input data and target data for low-illumination image compensation;
- L2 loss function was used in the generator to preserve the identifiable information of the input data;

- Leaky ReLU was used in the 6th and 7th decoder blocks of the generator to strengthen the high frequency information of the input image delivered through skip connections;
- ReLU was used in the 4th convolution layer of the discriminator.

Table 4. Discriminator structure in LAE-GAN.

Layer Name	Number of Filters	Size of Feature Map (Height × Width × Channel)	Filter Size (Height × Width)	Stride (Height × Width)	Padding (Height × Width)
Input image		256 × 256 × 3			
Generated or target image		256 × 256 × 3			
Concatenation		256 × 256 × 6			
1st convolutional layer Leaky ReLU layers	64	128 × 128 × 64	4 × 4 × 6	2 × 2	1 × 1
2nd convolutional layer Batch normalization Leaky ReLU layers	128	64 × 64 × 128	4 × 4 × 64	2 × 2	1 × 1
3rd convolutional layer Batch normalization Leaky ReLU layers	256	32 × 32 × 256	4 × 4 × 128	2 × 2	1 × 1
4th convolutional layer Batch normalization ReLU layers	512	31 × 31 × 512	4 × 4 × 256	1 × 1	1 × 1
5th convolutional layer	1	30 × 30 × 1	4 × 4 × 512	1 × 1	1 × 1
Sigmoid layer		30 × 30 × 1			

3.5. Age Estimation

In this study, age estimation was performed by training various CNNs using facial images enhanced by LAE-GAN. Training was performed using VGG [25], which achieved high accuracy in conventional image classification. The residual network (ResNet) [63], various networks that produced good accuracy in age estimation [25,29,63,64], and age estimation performance were compared according to the compensation of low-illumination facial images.

3.5.1. VGG

VGG [25] is a well-known classification network that has achieved high performance in ImageNet and is used or applied in various age estimation studies [29,64]. In general, classification performance tends to improve in deep learning networks as the depth increases. The performance of VGG was compared by implementing CNNs of different depths. Filters of 5×5 size and 7×7 size can be replaced with continuous filters of 3×3 size while reducing computational complexity; non-linearity of a network was secured by using a 1×1 convolution. In this study, age estimation performance was evaluated using VGG-16, which is fairly well-known among various VGG networks.

3.5.2. DEX

In a study [64], a VGG-16-based network was used to produce good performance in the age estimation field in the ChaLearn competition. DEX is an ImageNet database for which VGG-16 was pre-trained using an extensive number of databases, including IMDB and Wiki. Moreover, instead of estimating age based on the probability value of a class,

age was estimated as the sum of the product of a class label and the probability of the respective label, as shown in Equation (6):

$$Age(X) = \sum_1^n c_i p_i \quad (6)$$

where X is the input image, while n is the entire class (age range). Accordingly, c_i and p_i are the label and probability of the i^{th} class, respectively. As described above, DEX [64] is a VGG-based network which has 13 convolution layers and 3 fully connected layers. Like DEX, we used categorical cross entropy loss [65], as shown in Equations (7) and (8).

$$f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^C e^{s_j}} \quad (7)$$

$$\mathcal{L}_{CE} = - \sum_i^C t_i \log(f(s)_i) \quad (8)$$

In Equations (7) and (8), $f(\cdot)$ is a softmax activation function, e represents an exponential function, t is a ground-truth age, and s is an estimated age. In addition, C is the number of classes, i is the i^{th} class, and \log is the decimal logarithm. An adaptive moment estimation (Adam) optimizer [66] was used in our experiments, whereas DEX adopts a stochastic gradient descent (SGD) optimizer.

3.5.3. ResNet

ResNet [63] is a prototypical classification network that has achieved high performance in ImageNet. Furthermore, it has been widely used in various studies that researched age estimation—particularly in studies that use unique residual blocks and skip connections. It consists of continuous filters having 1×1 , 3×3 , and 1×1 sizes, and has a bottleneck structure for giving reduction and expansion effects on the dimension of a feature map. A weights sum is applied to the feature maps before and after the residual block to resolve the vanishing gradient problem. A skip connection is also present for maintaining the identity of the input image. ResNet is a network which has various depths depending on the number of residual blocks; in this study, ResNet-50 and ResNet-152 pre-trained with the ImageNet database are used in the experiment.

3.5.4. Age-Net

In a study [29], VGG and Age-Net were used for age estimation, which resulted in excellent age estimation performance in the ChaLearn competition. Training included the first step involving VGG and the second step involving Age-Net in which VGG—pre-trained with ImageNet—is fine-tuned using the MORPH database [67]. Then, various open databases are mixed and classified into two types to be trained using the KL divergence loss and softmax loss function. This process creates four fine-tuning models where a concatenated feature map is generated in the last layer of each model using a distance-based voting ensemble method. Secondly, Age-Net is trained with various open databases for which Kullback–Leibler (KL) divergence loss function is used. VGG and Age-Net have the same output dimension where the average of the two networks was estimated as the predicted age if the difference between the two networks was 11 or below; or, if the difference was greater than 11, the result of the first network (VGG) was then estimated as the predicted age.

3.5.5. Inception with Random Forest

In a study [68], the Inception v2 network [69] was applied with the random forest (RF) for age estimation. Inception v2 is a network that extracts features using convolution filters of various sizes and concatenates the extracted features to ensure the balance between a sparse nature and a dense nature of network training. Features are extracted using

Inception v2 pre-trained with various databases as a feature extractor, and RF is used to perform age learning.

4. Experimental Results

4.1. Experimental Data and Environment

In this study, the experiment was conducted using the MORPH [67], FG-NET [70], and AFAD [71] databases, which are open databases, as shown in Figure 4. The MORPH database has 55,134 facial images of 13,617 individuals aged between 16 and 77. In addition, the FG-NET database contains 1002 images of 82 individuals aged between 0 and 69. The AFAD database contains 164,432 facial images of individuals aged between 15 and 40.

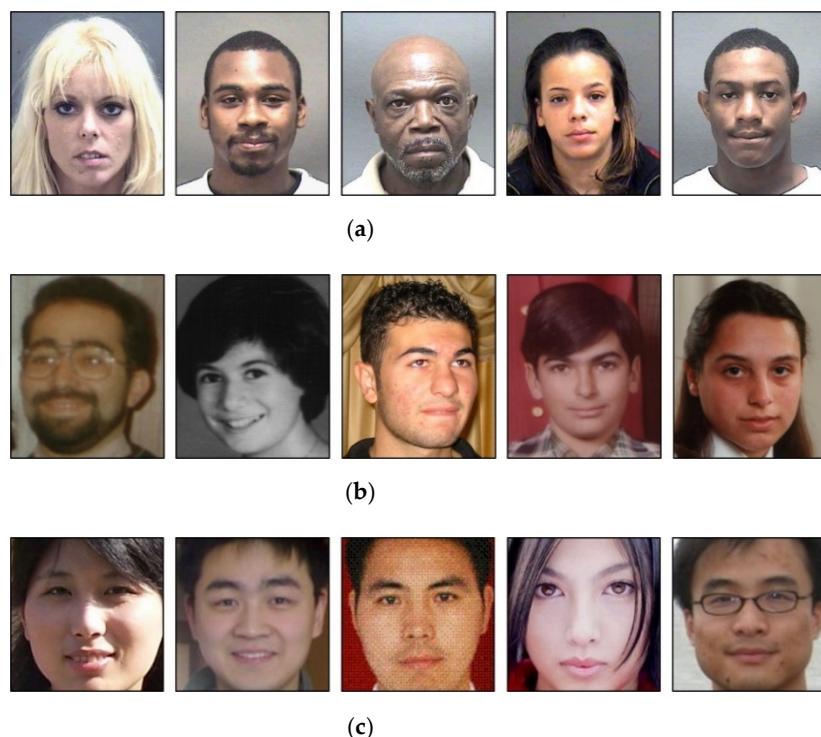


Figure 4. Examples of face image databases. (a–c) are examples of the MORPH, FG-NET, and AFAD databases, respectively.

Since open facial databases acquired in low-light environments and containing age information do not exist, the aforementioned open databases were transformed to low-illumination images to proceed with training and testing in this study. The same pre-processing explained in Section 3.2 was applied to the training images to redefine the ROI of facial images. The pre-processed low-light image and the original image are used as input images and target images for training. When illumination decreases in the actual environment, pixels with a large brightness value experience significant changes, while pixels with a small brightness value experience relatively smaller changes. For representing such a non-linear nature, a gamma correction [72] technique was applied in this study to generate low-illumination facial images. Original RGB images were converted to HSV images, which consist of hue, saturation, and value channels, expressed as H, S, and V channels, respectively. Gamma correction was applied to the V channel to decrease the non-linear brightness value. Blurry images are generated due to the exposure time of a camera in low-light environments in which noise due to a camera sensor is also generated. For applying these elements, a Gaussian blur was applied to generate a blurry image,

while Gaussian and Poisson noises were applied to generate noise in this experiment. Equation (9) below shows the effects used for generating low-illumination facial images.

$$I_o = B_G(S \cdot (I_v)^\gamma) + N_G + N_P \quad (9)$$

In Equation (9), I_v is the V channel value of the HSV image, while I_o is the V channel value of the low-illumination image generated as above. S and γ are gamma correction parameters for which S is 0.06 and γ is 2.5. B_G is the Gaussian blur kernel, for which the standard deviation σ was randomly applied between 1.5 and 2. We selected these values based on previous studies [73,74]. Lastly, N_G and N_P are Gaussian and Poisson noise, respectively. Figure 5 shows the examples of the original facial images and low-illumination facial images generated for the experiment. Figure 5c shows the corresponding histogram-equalized images of low-illumination facial images of Figure 5b. Although the low-illumination images of Figure 5b are difficult to discriminate via the human eye, we can confirm that they have rough information of face images as shown in Figure 5c. Therefore, the algorithm does not estimate age from non-visible/non-usable images.

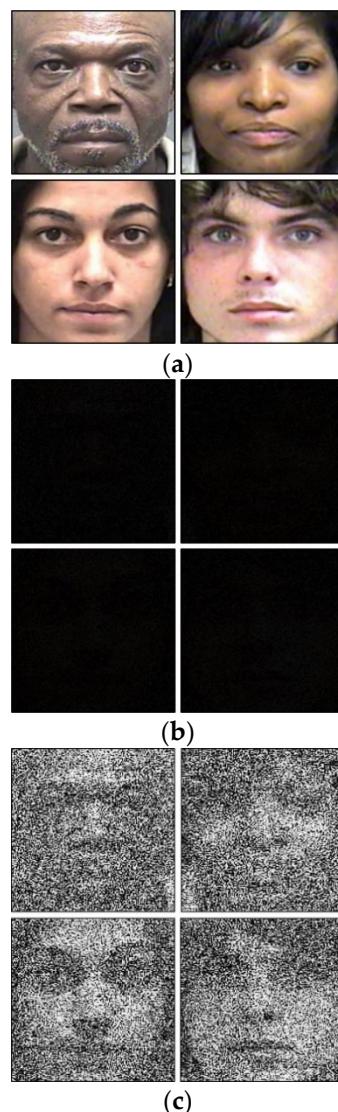


Figure 5. Examples of original and generated facial images. (a) Original images, (b) corresponding generated facial images in low illumination, and (c) corresponding histogram-equalized facial images in low illumination.

For the experiment, we used a desktop computer, which was equipped with a 3.5 GHz CPU (Intel Core™ i7-3770K) and 24 GB RAM. Windows TensorFlow (version 2.2.0) [75] was utilized for the training and testing procedure. We used an NVIDIA graphics processing unit (GPU) card including 1920 compute unified device architecture (CUDA) cores and 8 GB memory (Nvidia GeForce GTX 1070 [76]). To extract the face ROI, we used the Python program (version 3.5.2) [77] and the OpenCV (version 4.2.0) library [78].

4.2. Training of LAE-GAN for Image Enhancement of Low Illumination and CNN for Age Estimation

LAE-GAN, explained in Section 3.3, was used to enhance low-illumination images into high-illumination images, and various age estimation networks explained in Section 3.4 were used to estimate ages. LAE-GAN was trained with low-illumination images as input images and high-illumination images as target images. As explained in Section 4.1, pre-processed training data were resized into $286 \times 286 \times 3$ and then randomly cropped to $256 \times 256 \times 3$ through online augmentation for training. An Adam optimizer [66] was used during training. Learning rate, beta_1, and beta_2 were set to 0.0002, 0.5, and 0.999, respectively, for training, which was conducted over 100 epochs. The optimal parameters of learning rate, beta_1, beta_2, and the number of epoch were experimentally determined with training data, which showed the highest accuracy of age estimation with the training data.

Figure 6 shows the training loss graphs of the generator and discriminator when LAE-GAN was trained using the MORPH database. Figure 6a shows the loss graph of the generator, and Figure 6b shows the loss graph of the discriminator. In general, when the loss function converges to 0, the training can be regarded as progressing well. The discriminator has a binary classification problem that discriminates real and fake images, and the network is simple. On the other hand, the generator that enhances the image has a deep network. Therefore, the discriminator has a lower learning complexity than the generator. Consequently, the discriminator loss converges relatively quickly compared to the generator loss, and the converged loss value of discriminator is usually lower than that of generator. In this study, by adding the L2 loss, the loss of the discriminator temporarily increases. However, the discriminator loss converges at a similar time to the generator loss. As shown in Figure 6a,b, both generator and discriminator loss converged, which indicates that LAE-GAN was properly trained. Subsequently, the CNN was trained for age estimation using the facial images enhanced with trained LAE-GAN. Various age estimation networks explained in Section 3.4 were used for training. Previously trained networks were fine-tuned, in which the training was conducted for 200 epochs. Figure 7 shows the training loss and accuracy graphs of DEX [64], which exhibited the highest age estimation performance. The convergence of the loss function means that the error is reduced, so the accuracy should be improved. In Figure 7, as training loss stably converged and accuracy stably increased, the network could be considered adequately trained.

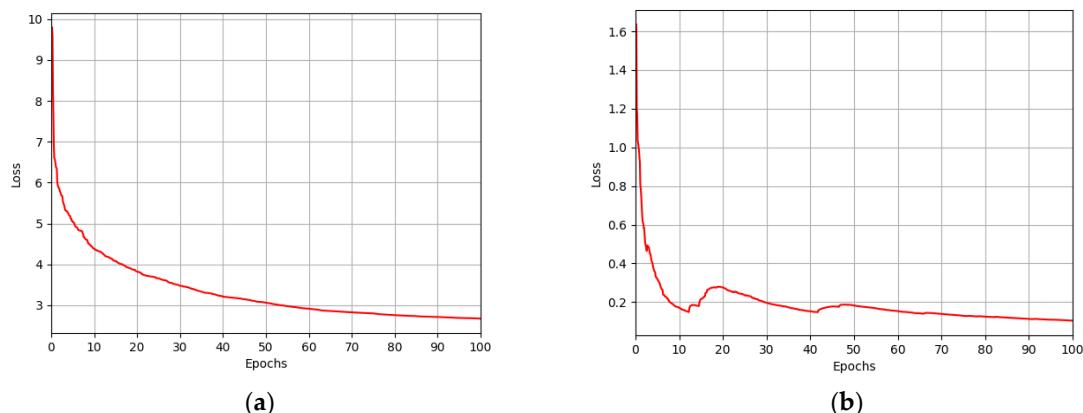


Figure 6. Graphs of LAE-GAN losses. (a) Generator loss and (b) discriminator loss.

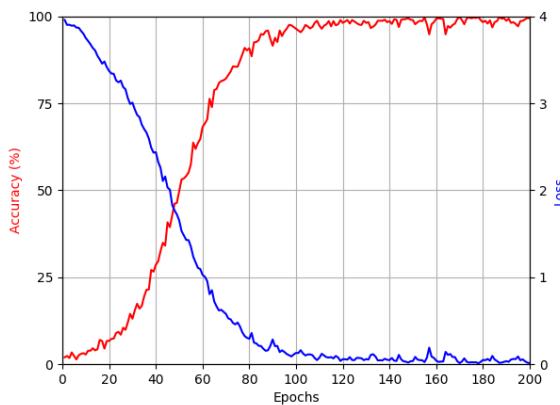


Figure 7. DEX training loss and accuracy graphs with enhanced images.

4.3. Testing with the MORPH Database

In the first experiment, the image enhancement performances of the LAE-GAN proposed in this study and other state-of-the-art networks were compared. CycleGAN [79], Attention GAN [80], Attention cGAN [81], and conditional GAN [59] were used to compare the illumination enhancement performance with LAE-GAN; the signal-to-noise ratio (SNR) [82], peak signal-to-noise ratio (PSNR) [83], and structural similarity (SSIM) [84] were used for comparing the similarity between the original image and the generated enhanced image. Equations (10)–(13) represent the equations for MSE, SNR, PSNR, and SSIM, respectively. SNR, PSNR, and SSIM values tend to be higher if the similarity between two images is higher.

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_o(i, j) - I_e(i, j)]^2 \quad (10)$$

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_o(i, j)]^2}{\text{MSE}} \right) \quad (11)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \quad (12)$$

I_o is an original image of high illumination and I_e is the generated image. m and n show the width and height of the image, respectively.

$$\text{SSIM} = \frac{(2\mu_e\mu_o + C1)(2\sigma_{eo} + C2)}{(\mu_e^2 + \mu_o^2 + C1)(\sigma_e^2 + \sigma_o^2 + C2)} \quad (13)$$

μ_o and σ_o show the mean and standard deviation of the pixel values of an original image of high illumination, respectively. μ_e and σ_e show the mean and standard deviation of the pixel values of a generated image, respectively; σ_{eo} is the covariance of the two images. $C1$ and $C2$ are the positive constant values, which make the denominator non-zero.

As shown in Table 5, there exist other methods that exhibited better performance than LAE-GAN in SNR and PSNR, whereas LAE-GAN resulted in the best performance in SSIM. However, PSNR and SNR cannot accurately evaluate the similarity and difference in the visual definitions of humans [85,86]. SSIM, on the other hand, is more suitable for evaluating similarities in definitions since it is a measurement designed for improving PSNR and SNR [84]. Accordingly, it can be confirmed that the proposed method resulted in the highest accuracy.

Table 5. Comparative accuracies of enhancement by our network and the state-of-the-art methods.

Methods	SNR	PSNR	SSIM
CycleGAN [79]	1.2971	19.0120	0.5024
Attention GAN [80]	1.1808	16.3112	0.5011
Attention cGAN [81]	1.2734	18.5221	0.5631
Conditional GAN [59]	1.4802	19.8352	0.6207
LAE-GAN	1.3924	18.9404	0.6223

Figure 8 illustrates the images enhanced by various networks presented in Table 5. Figure 8c shows the corresponding histogram-equalized images of the low-illumination facial images of Figure 8b. Although the low-illumination images of Figure 8b are difficult to discriminate by the human eye, we can confirm that they have rough information of face images as shown in Figure 8c. Therefore, the algorithms are not getting better images from completely random/black images. In addition, as shown in Figure 8h, the proposed LAE-GAN successfully transforms the low-illumination facial images for Figure 8b. The LAE-GAN proposed in this study has more outstanding image enhancement effects compared to other networks, as shown in Figure 8.

For the next experiment, age estimation accuracy was compared using various networks explained in Section 3.4 for the images enhanced by LAE-GAN, as shown in Table 6. For evaluating the age estimation accuracy, MAE, which is the most often-used measure, is used as shown in Equation (14). A lower MAE value indicates higher age estimation accuracy.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |p_i - y_i| \quad (14)$$

Table 6. Comparisons of age estimation accuracies by various methods on the MORPH database (unit: years).

Method	MAE
VGG-16 [25]	13.99
ResNet-50 [63]	12.83
ResNet-152 [63]	12.76
DEX [64]	12.46
AgeNet [29]	15.33
Inception with RF [68]	15.01
Original	5.8
Low illumination (without LAE-GAN)	19.02
Enhanced by LAE-GAN (proposed)	12.46
CycleGAN [79]	16.97
Attention GAN [80]	19.00
Attention cGAN [81]	18.60
Conditional GAN [59]	13.01
LAE-GAN	12.46

In the equation, n is the number of images, p_i is the estimated age, and y_i is the ground-truth age.

The experiment results showed that DEX had the best performance in age estimation. The age estimation performance of other networks were better than the age estimation performance based on low-illumination facial images, as shown in Table 6. Therefore, it can be concluded that the LAE-GAN used in this study performed better in enhancing low-illumination facial images for age estimation.

In Table 6, age estimation performance, or baseline performance, was measured in original images of high-illumination and low-illuminated images with or without LAE-GAN using DEX—which had the best performance in Table 6. In each case, DEX was fine-tuned using the training data, and accuracy was evaluated using the testing data.

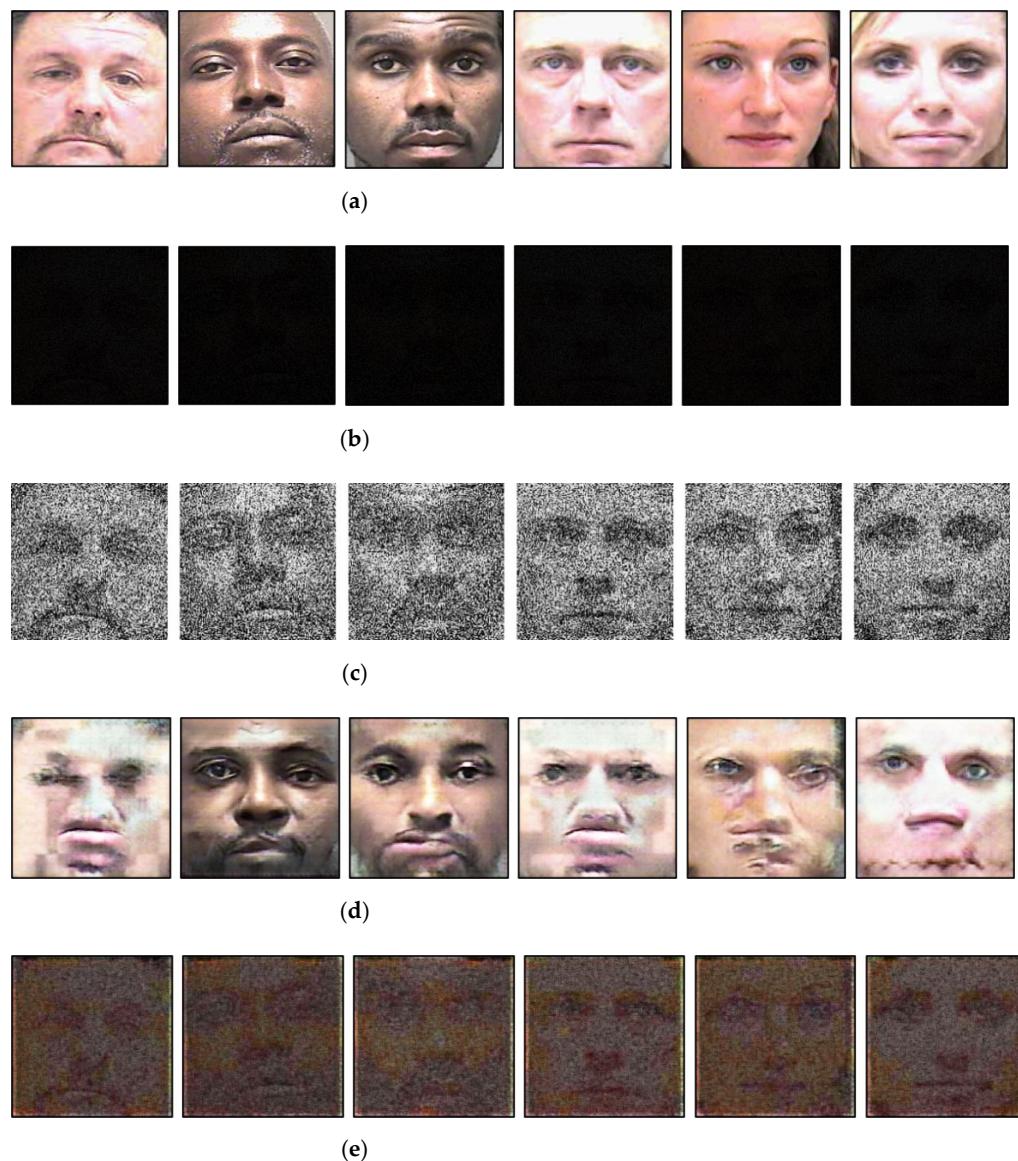


Figure 8. Cont.

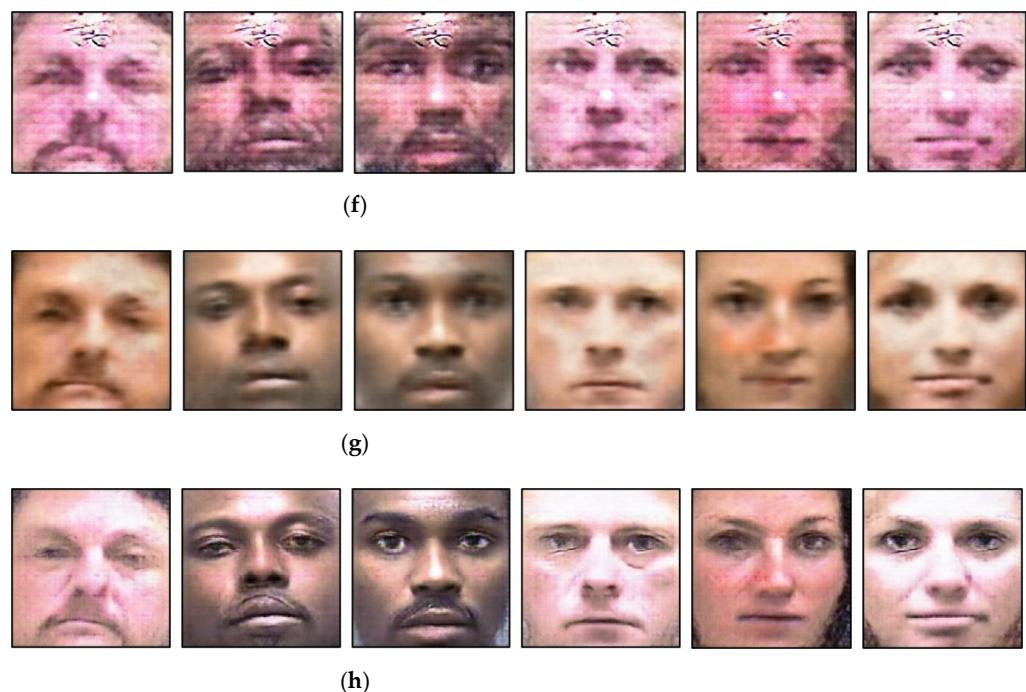


Figure 8. Examples of (a) original images, (b) low-illuminated images, (c) histogram equalized images, and enhanced images by (d) CycleGAN, (e) Attention GAN, (f) Attention cGAN, (g) conditional GAN, and (h) LAE-GAN.

As shown in Table 6, a MAE of 5.8 years was found in the original images of high illumination, whereas a MAE of 19.02 years was found in the low-illuminated images without LAE-GAN. However, the MAE was significantly reduced to 12.46 years when LAE-GAN was used. In Table 6, in the case of “Original” images, we trained and tested with the original dataset. In the case of “Low illumination (without LAE-GAN)”, we trained and tested with the low-illumination dataset. In case of “Enhanced by LAE-GAN (proposed)”, we trained and tested with the image dataset enhanced by LAE-GAN. Therefore, they were fair comparisons, since the model was trained on one set of images and its performance was also evaluated on the same set.

For the next experiment, the age estimation performance of the LAE-GAN and other state-of-the-art networks were compared. For a fair evaluation, DEX was used as an age estimator for all cases. As shown in Table 6, LAE-GAN had the greatest effect on low-illumination facial image enhancement and age estimation performance improvement.

Figures 9 and 10 show good cases and bad cases, respectively, of age estimation performance when age is estimated using DEX and LAE-GAN. The first and second rows of Figures 9 and 10 are original images and low-illumination images, respectively. The third rows of Figures 9 and 10 are facial images enhanced using LAE-GAN. In Figure 9, the images were enhanced to be very similar to the original images, unlike Figure 10, where high frequency information such as wrinkles or detailed information such as the skin texture of the original images were not adequately restored in the enhancement images. Consequently, a higher portion of bad cases was found when low-illumination images of older individuals were enhanced to appear as images of younger individuals—which resulted in less accurate age estimation.

4.4. Testing with the AFAD Database

For verifying the generality of the proposed method, an experiment was conducted using a different open database—the AFAD database. For the first experiment, age estimation accuracy was compared using the various networks explained in Section 3.4 for the

images enhanced by LAE-GAN, as shown in Table 7. The experiment results showed that the best performance was exhibited by Inception with RF, unlike the MORPH database.

Table 7. Comparisons of age estimation accuracies by various methods on the AFAD database (unit: years).

Method	MAE
VGG-16 [25]	14.10
ResNet-50 [63]	16.31
ResNet-152 [63]	14.35
DEX [64]	14.12
AgeNet [29]	15.17
Inception with RF [68]	13.81
Original	7.08
Age estimation using various age estimators with LAE-GAN	
Low illumination (without LAE-GAN)	16.10
Enhanced by LAE-GAN (proposed)	13.81

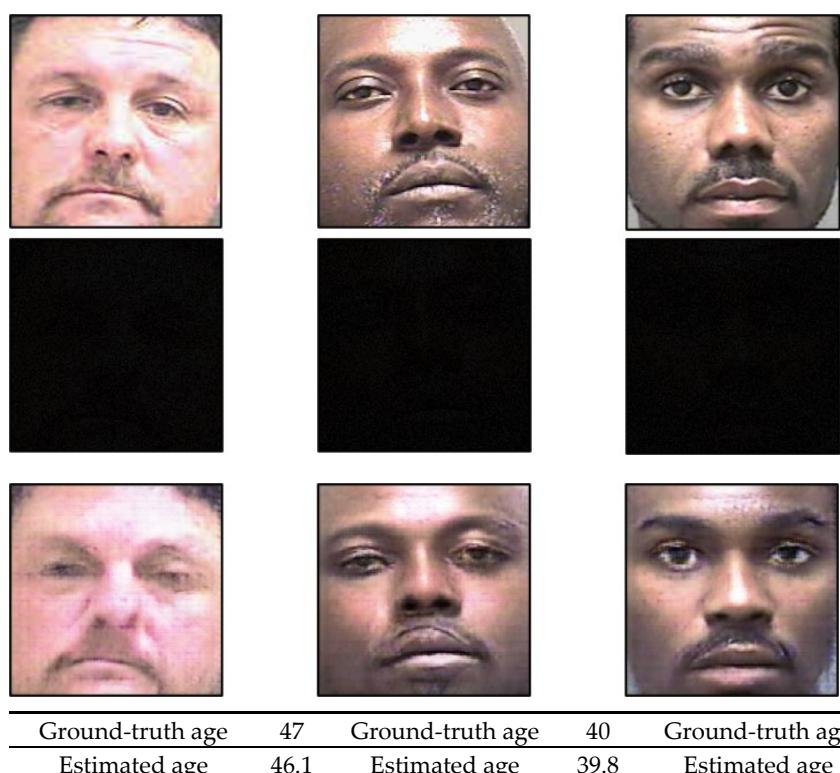


Figure 9. Good cases of age estimation by proposed method. The 1st, 2nd and 3rd rows show the original, low-illuminated images, and the enhanced one by LAE-GAN, respectively.

In Table 7, age estimation performance, or baseline performance, was measured in original images of high illumination and low-illuminated images with or without LAE-GAN using Inception with RF, which had the best performance in Table 7. In each case, Inception with RF was fine-tuned using the training data, and accuracy was evaluated using the testing data.

As shown in Table 7, a MAE of 7.08 years was found in the original images of high illumination, where a MAE of 16.10 years was found in the low illuminated images without LAE-GAN. However, MAE was reduced to 13.81 when LAE-GAN was used.

Figures 11 and 12 show good cases and bad cases, respectively, of age estimation performance when age is estimated using Inception with RF with LAE-GAN. The first and second rows of Figures 11 and 12 are original images and low-illumination images, respectively. The third rows of Figures 11 and 12 are facial images enhanced using LAE-GAN.

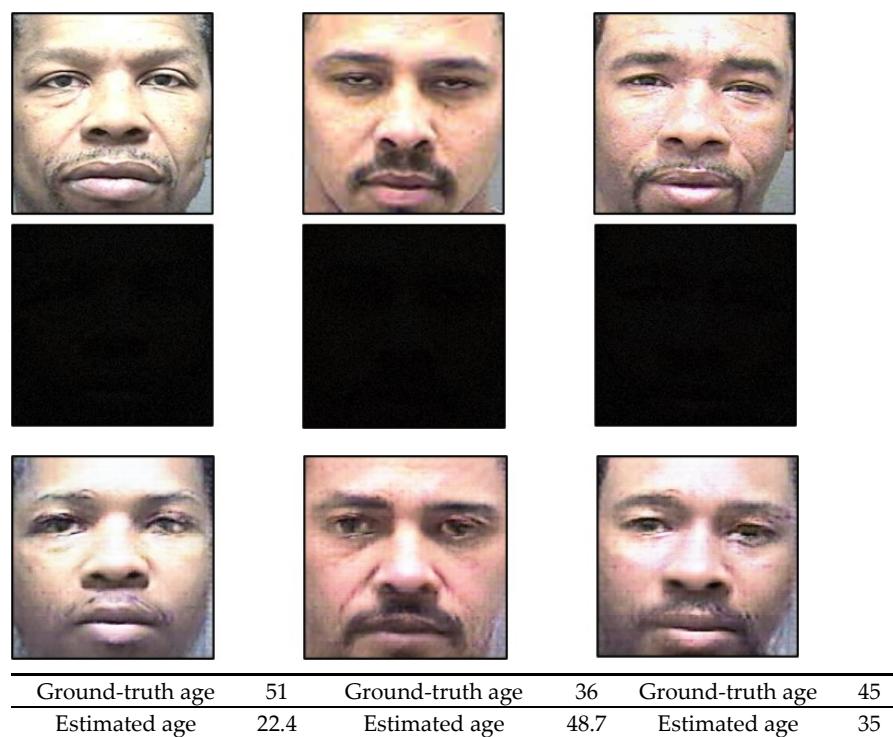


Figure 10. Bad cases of age estimation by the proposed method. The 1st, 2nd and 3rd rows show the original, low-illuminated images, and the images enhanced by LAE-GAN, respectively.

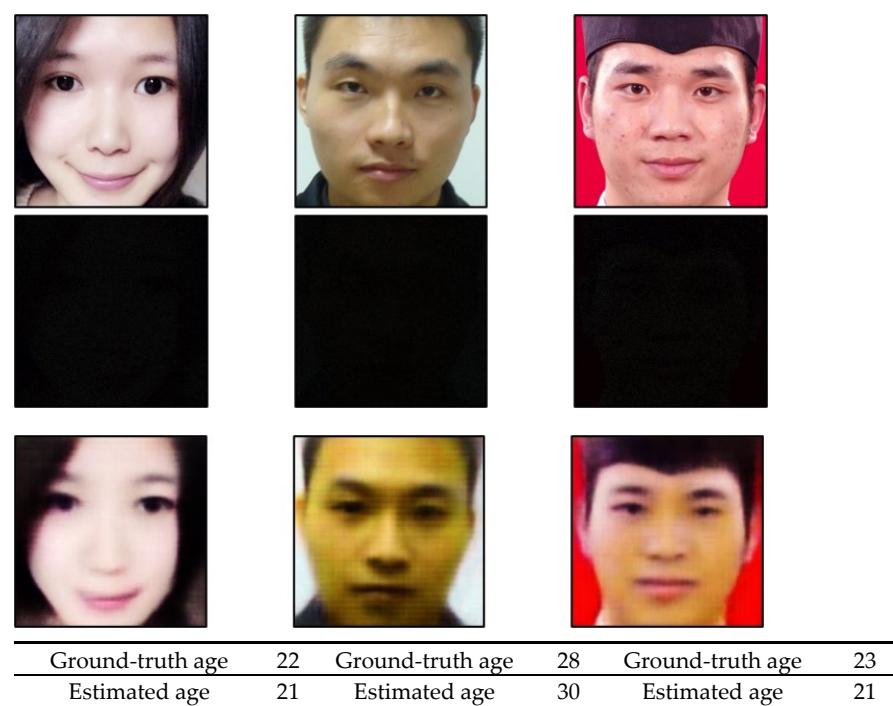


Figure 11. Good cases of age estimation by the proposed method. The 1st, 2nd and 3rd rows show the original images, the low-illuminated images, and the images enhanced by LAE-GAN, respectively.

Ground-truth age 26	Ground-truth age 21	Ground-truth age 19
Estimated age 19	Estimated age 31	Estimated age 30

Figure 12. Bad cases of age estimation by the proposed method. The 1st, 2nd and 3rd rows show the original images, low-illuminated images, and images enhanced by LAE-GAN, respectively.

Figures 11 and 12 show blurs in the enhanced facial images. However, blur is more severe in bad cases compared to good cases in Figure 11, and many enhanced images with severe noise were observed, which ultimately led to degradation in age estimation performance.

4.5. Testing with the FG-NET Database

For verifying the generality of the proposed method, an experiment was conducted using another open database—the FG-NET database. For the first experiment, age estimation accuracy was compared using various networks explained in Section 3.4 for the images enhanced by LAE-GAN, as shown in Table 8. The experiment results showed that the best performance was exhibited by DEX, similar to the MORPH database.

Table 8. Comparisons of age estimation accuracies by various methods on the FG-NET database (unit: years).

Method	MAE
VGG-16 [25]	10.22
ResNet-50 [63]	11.00
ResNet-152 [63]	9.74
DEX [64]	9.55
AgeNet [29]	10.40
Inception with RF [68]	10.14
Original	6.42
Age estimation using various age estimators with LAE-GAN	
Low illumination (without LAE-GAN)	11.31
Enhanced by LAE-GAN (proposed)	9.55
Age estimation using original facial images or low-illuminated facial images without or with LAE-GAN	

In Table 8, age estimation performance, or baseline performance, was measured in original images of high illumination and low-illuminated images with or without LAE-GAN using DEX, which had the best performance in Table 8. In each case, DEX was fine-tuned using the training data, and accuracy was evaluated using the testing data.

As shown in Table 8, a MAE of 6.42 years was found in the original images of high illumination, whereas a MAE of 11.31 years was found in the low-illuminated images without LAE-GAN. However, MAE was reduced to 9.55 when LAE-GAN was used.

Figures 13 and 14 show good cases and bad cases, respectively, of age estimation performance when age is estimated using DEX and LAE-GAN. The first and second rows of Figures 13 and 14 are original images and low-illumination images, respectively. The third rows of Figures 13 and 14 are facial images enhanced using LAE-GAN.

As shown in Figures 13 and 14, when LAE-GAN was trained using the FG-NET database, the overall color of the images changed, but detailed information and overall shape were expressed adequately in good cases compared to the bad cases. An enhanced image different from the original image was generated in some bad cases, which increased errors in age estimation.

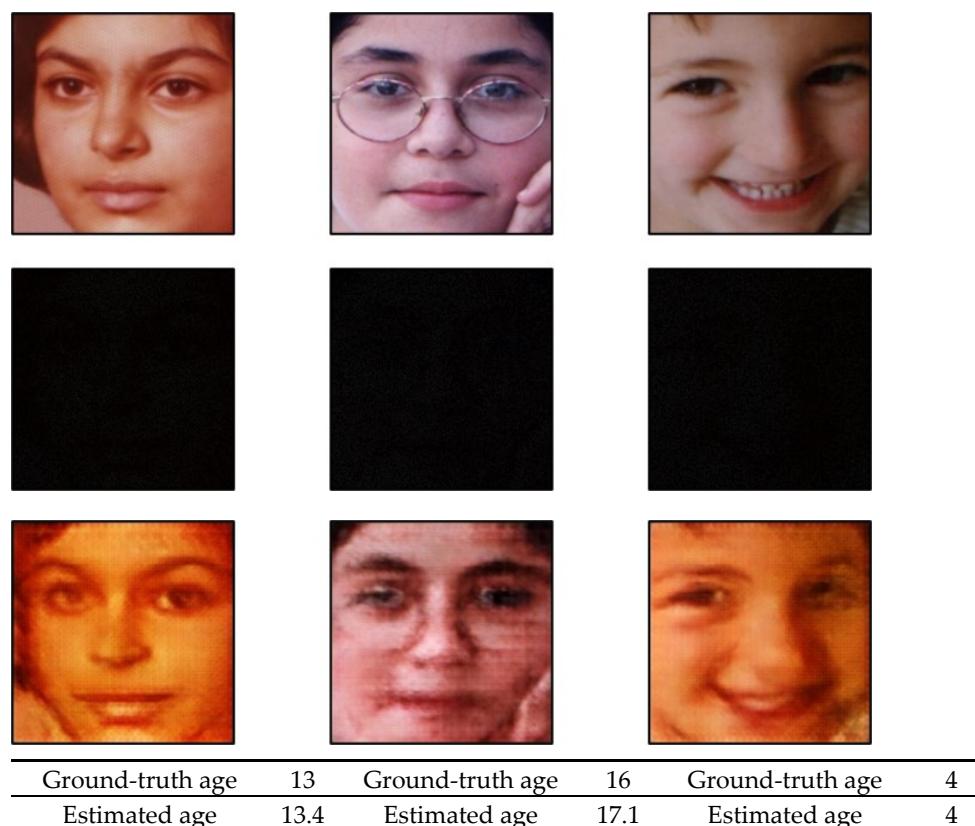


Figure 13. Good cases of age estimation by the proposed method. The 1st, 2nd and 3rd rows show the original images, low-illumination images, and the images enhanced by LAE-GAN, respectively.

4.6. Discussion and Analysis of Grad CAM

In our experiments, we used the AFAD database, which already includes images with severe slant angles (in-plane and out-plane rotations) and illumination variations as shown in Figure 15a. The number of images of these severe slant angles and illumination variations are almost 20% of the total number of images of the AFAD database. However, our LAE-GAN successfully transformed the low-illumination images (Figure 15b) of these severe slant angles and illumination variations into enhanced ones as shown in Figure 15c, and our method shows a higher accuracy of age estimation than the state-of-the-art methods, as shown in Table 7.

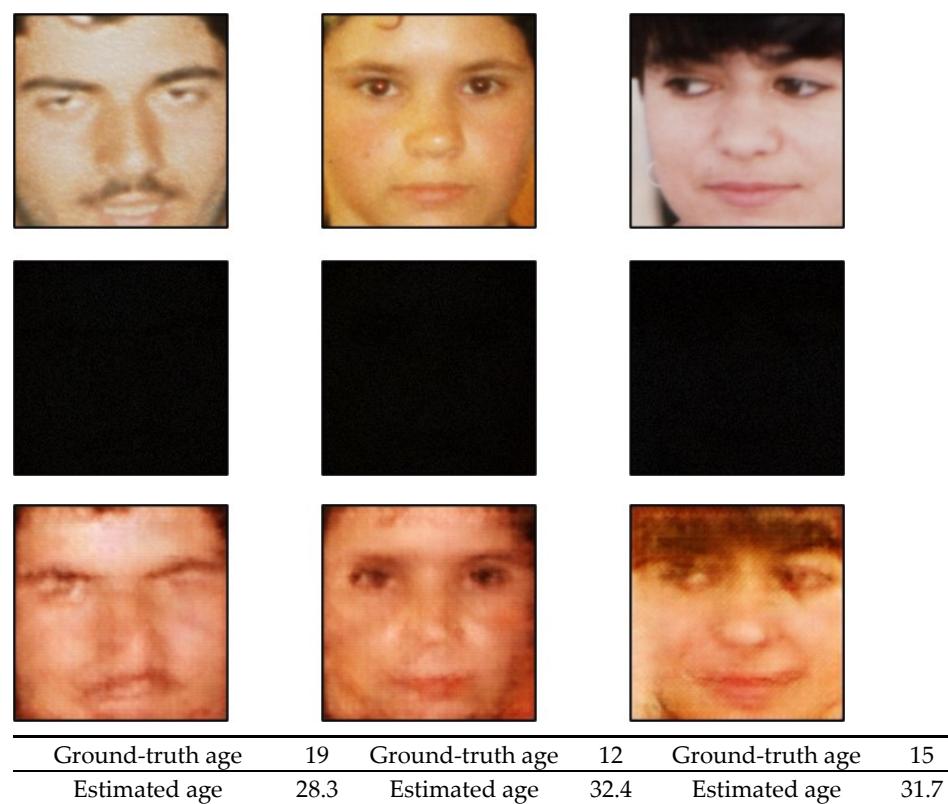


Figure 14. Bad cases of age estimation by the proposed method. The 1st, 2nd and 3rd rows show the original images, low-illuminated images, and the images enhanced by LAE-GAN, respectively.

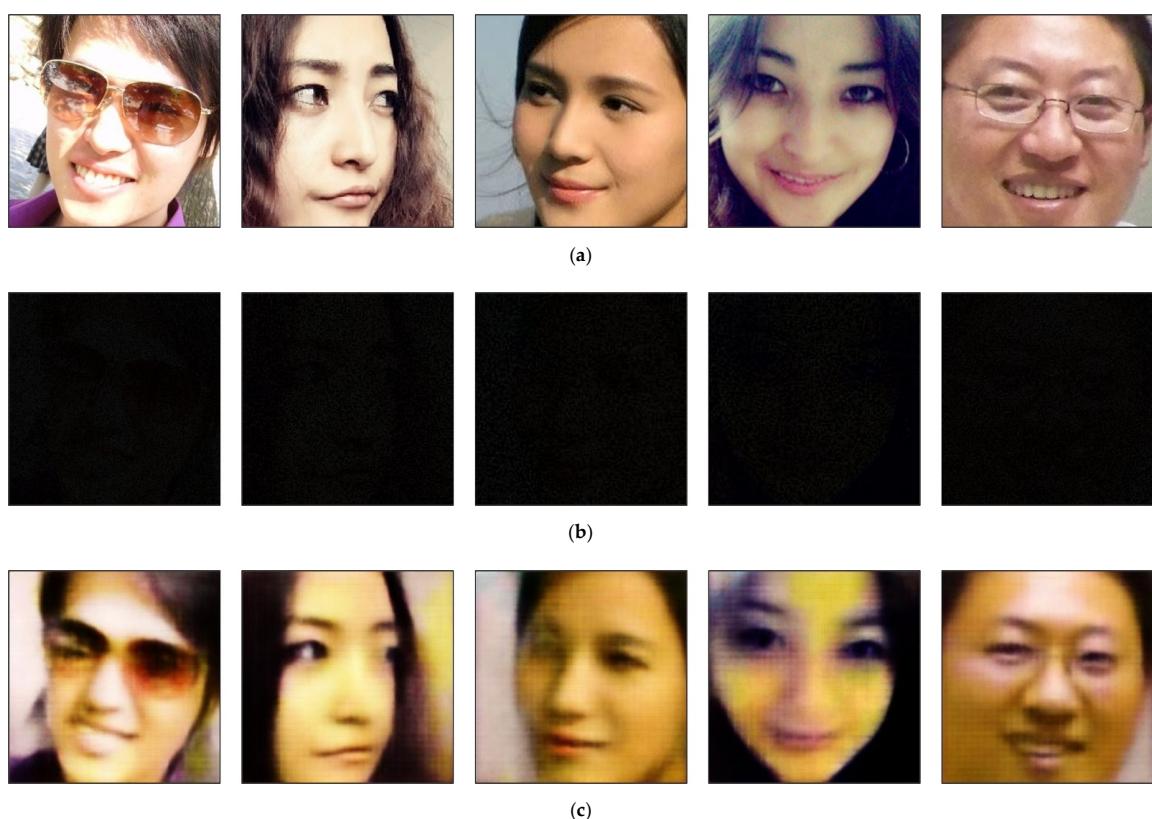


Figure 15. Various angles and illumination images in the AFAD database. (a) is the original database, (b) is the low-illumination images, and (c) is the images enhanced with LAE-GAN.

In addition, gradient-weighted class activation mapping (Grad-CAM) [87] images extracted from each layer of DEX, with the images enhanced using LAE-GAN as input, were analyzed. Figure 16a is the original facial image, while the pictures on the left and right sides in Figure 16b are low-illumination images and the images enhanced by LAE-GAN, respectively. Figure 16c through Figure 16g are Grad-CAM images extracted from the first, fourth, eighth, and eleventh convolutional layers and the last max pooling layers. The pictures on the left in Figure 16c–g are Grad-CAM images, while the pictures on the right are the LAE-GAN-enhanced images overlapped with the Grad-CAM images.

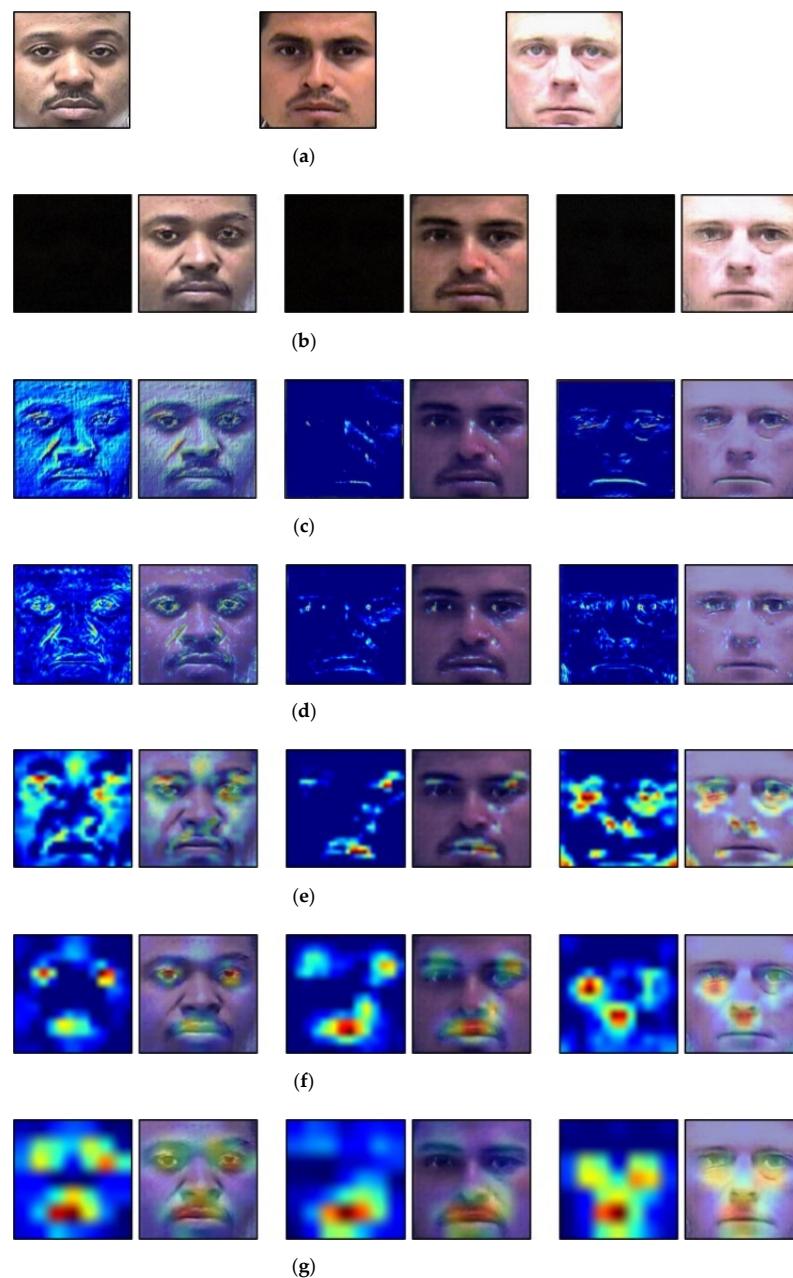


Figure 16. Examples of Grad-CAM images extracted from DEX where LAE-GAN enhanced images are used as input. (a) Original images, (b) low-illumination images (left) and LAE-GAN enhanced images (right). (c–g) are Grad-CAM images extracted from the first, fourth, eighth, and eleventh convolutional layers and the last max pooling layers of DEX. The left pictures in Figure 16c–g are Grad-CAM images, while the right pictures are the LAE-GAN-enhanced images overlaid with the Grad-CAM images.

As shown in Figure 16c,d, high activation areas, mostly in the high frequency areas such as the eyes, nose, mouth, and lines, in the Grad-CAM images are extracted from the front convolutional layers of DEX. As convolution proceeds, it can be observed in Figure 16e–g that activation areas are found in more global areas of the face, including the eyes, nose, and mouth. As shown in Figure 16g, the features effective for age estimation are adequately extracted through the eye, nose, and mouth areas of the face using the proposed method.

5. Conclusions

Human facial images acquired in low-illumination environments lose the information required for age estimation because various kinds of noise and blur are generated. Therefore, to overcome the problem of degradation in age estimation performance of human facial images captured in low-light environments, this study proposed a new LAE-GAN for enhancing low-illumination images and performed a CNN-based age estimation on the enhanced images. The results of the experiments conducted using open databases—including the MORPH database, FG-NET database, and AFAD database—showed that low-illumination images enhanced with the LAE-GAN proposed in this study produced better age estimation performance compared to state-of-the-art enhancement networks. However, in the case of enhancement by LAE-GAN, high-frequency information such as wrinkles and detailed information such as the skin texture of the original image were not fully restored in the enhanced image, or different images from the original image were generated. In addition, the restored images are a little blurred and include additional noise through the transformation by LAE-GAN.

For solving these issues in the future, adding a loss function to fully restore skin texture or strengthening an identity loss to prevent different enhanced images from being generated will be investigated further. Moreover, more research will be conducted on age estimation and image enhancement using images of various illumination and angles in addition to facial image compensation that is more effective against various environments that are found in the real world. Although the proposed method shows high performance, the processing time is increased by operating two models of LAE-GAN and an age estimator. In future work, we intend to investigate a method to combine these two models into one, which can enhance the processing speed without reducing the accuracy of age estimation.

Author Contributions: Methodology, S.H.N.; conceptualization, Y.H.K.; validations, J.C., S.B.H., M.O.; supervision, K.R.P.; writing—original draft, S.H.N.; writing—review and editing, K.R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program (NRF-2021R1F1A1045587), in part by the NRF funded by the MSIT through the Basic Science Research Program (NRF-2020R1A2C1006179), and in part by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01789) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, X.; Yap, M.H.; Palmer, I. Face recognition in the presence of expressions. *J. Softw. Eng. Appl.* **2012**, *5*, 321–329. [[CrossRef](#)]
2. Meyers, E.; Wolf, L. Using biologically inspired features for face processing. *Int. J. Comput. Vis.* **2008**, *76*, 93–104. [[CrossRef](#)]
3. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comp.* **2009**, *27*, 803–816. [[CrossRef](#)]

4. Wang, Y.; Ricanek, K.; Chen, C.; Chang, Y. Gender classification from infants to seniors. In Proceedings of the 4th IEEE International Conference on Biometrics: Theory, Applications, Systems, Washington, DC, USA, 27–29 September 2010; pp. 1–6.
5. Alarifi, J.S.; Goyal, M.; Davison, A.K.; Dancey, D.; Khan, R.; Yap, M.H. Facial skin classification using convolutional neural networks. In Proceedings of the International Conference Image Analysis and Recognition, Montreal, QC, Canada, 5–7 July 2017; pp. 479–485.
6. Punyani, P.; Gupta, R.; Kumar, A. Neural networks for facial age estimation: A survey on recent advances. *Artif. Intell. Rev.* **2020**, *53*, 3299–3347. [[CrossRef](#)]
7. Taheri, S.; Toygar, Ö. On the use of DAG-CNN architecture for age estimation with multi-stage features fusion. *Neurocomputing* **2019**, *329*, 300–310. [[CrossRef](#)]
8. Cootes, T.F.; Edwards, G.J.; Taylor, C.J. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 681–685. [[CrossRef](#)]
9. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active shape models—their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59. [[CrossRef](#)]
10. Geng, X.; Zhou, Z.-H.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240. [[CrossRef](#)] [[PubMed](#)]
11. Gabor, D. Theory of communication. Part 1: The analysis of information. *J. Inst. Electr. Eng. Part. III Radio. Commun. Eng.* **1946**, *93*, 429–441. [[CrossRef](#)]
12. Fisher, R.A. The statistical utilization of multiple measurements. *Ann. Eugen.* **1938**, *8*, 376–386. [[CrossRef](#)]
13. Gunay, A.; Nabihev, V.V. Automatic age classification with LBP. In Proceedings of the 23th International Symposium on Computer and Information Sciences, Istanbul, Turkey, 27–29 October 2008; pp. 1–4.
14. Choi, S.E.; Lee, Y.J.; Lee, S.J.; Park, K.R.; Kim, J. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognit.* **2011**, *44*, 1262–1281. [[CrossRef](#)]
15. Conde, M.H.; Zhang, B.; Kagawa, K.; Loffeld, O. Low-light image enhancement for multiaperture and multitap systems. *IEEE Photonics J.* **2016**, *8*, 1–25. [[CrossRef](#)]
16. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [[CrossRef](#)] [[PubMed](#)]
17. Aditya, K.P.; Reddy, V.K.; Ramasangam, H. Enhancement technique for improving the reliability of disparity map under low light condition. *Procedia Technol.* **2014**, *14*, 236–243. [[CrossRef](#)]
18. Qi, M.; Yanyan, W.; Jiao, L.; Hongan, L.; Zhanli, L. Research on the improved Retinex algorithm for low illumination image enhancement. *J. Harbin Eng. Univ.* **2018**, *39*, 2001–2010.
19. Wang, W.; Wu, X.; Yuan, X.; Gao, Z. An experiment-based review of low-light image enhancement methods. *IEEE Access* **2020**, *8*, 87884–87917. [[CrossRef](#)]
20. LAE-GAN with Algorithm. Available online: <https://github.com/nsh6473/LAE-GAN> (accessed on 26 May 2021).
21. Chao, W.-L.; Liu, J.-Z.; Ding, J.-J. Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognit.* **2013**, *46*, 628–641. [[CrossRef](#)]
22. Wang, X.; Guo, R.; Kambhamettu, C. Deeply-learned feature for age estimation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 534–541.
23. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 11–12 June 2015; pp. 34–42.
24. RMalli, C.; Aygun, M.; Ekenel, H.K. Apparent age estimation using ensemble of deep learning models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 9–16.
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
26. Chen, J.-C.; Kumar, A.; Ranjan, R.; Patel, V.M.; Alavi, A.; Chellappa, R. A cascaded convolutional neural network for age estimation of unconstrained faces. In Proceedings of the 8th IEEE International Conference on Biometrics Theory, Applications and Systems, Niagara Falls, NY, USA, 6–9 September 2016; pp. 1–8.
27. Chen, S.; Zhang, C.; Dong, M.; Le, J.; Rao, M. Using ranking-CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5183–5192.
28. Huerta, I.; Fernández, C.; Segura, C.; Hernando, J.; Prati, A. A deep analysis on age estimation. *Pattern Recognit. Lett.* **2015**, *68*, 239–249. [[CrossRef](#)]
29. Liu, X.; Li, S.; Kan, M.; Zhang, J.; Wu, S.; Liu, W.; Han, H.; Shan, S.; Chen, X. Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 16–24.
30. Huo, Z.; Yang, X.; Xing, C.; Zhou, Y.; Hou, P.; Lv, J.; Geng, X. Deep age distribution learning for apparent age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 17–24.
31. Yang, Y.; Chen, F.; Chen, X.; Dai, Y.; Chen, Z.; Ji, J.; Zhao, T. Video system for human attribute analysis using compact convolutional neural network. In Proceedings of IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016; pp. 584–588.

32. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928.
33. Hu, Z.; Wen, Y.; Wang, J.; Wang, M.; Hong, R.; Yan, S. Facial age estimation with age difference. *IEEE Trans. Image Process.* **2016**, *26*, 3087–3097. [[CrossRef](#)]
34. Li, K.; Xing, J.; Hu, W.; Maybank, S.J. D2C: Deep cumulatively and comparatively learning for human age estimation. *Pattern Recognit.* **2017**, *66*, 95–105. [[CrossRef](#)]
35. Qawaqneh, Z.; Mallouh, A.A.; Barkana, B.D. Age and gender classification from speech and face images by jointly fine-tuned deep neural networks. *Expert Syst. Appl.* **2017**, *85*, 76–86. [[CrossRef](#)]
36. Rodríguez, P.; Cucurull, G.; Gonfaus, J.M.; Roca, F.X.; González, J. Age and gender recognition in the wild with deep attention. *Pattern Recognit.* **2017**, *72*, 563–571. [[CrossRef](#)]
37. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNN–ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [[CrossRef](#)]
38. Wan, J.; Tan, Z.; Lei, Z.; Guo, G.; Li, S.Z. Auxiliary demographic information assisted age estimation with cascaded structure. *IEEE Trans. Cybern.* **2018**, *48*, 2531–2541. [[CrossRef](#)] [[PubMed](#)]
39. Zaghbani, S.; Boujneh, N.; Bouhlel, M.S. Age estimation using deep learning. *Comput. Electr. Eng.* **2018**, *68*, 337–347. [[CrossRef](#)]
40. Yoo, B.; Kwak, Y.; Kim, Y.; Choi, C.; Kim, J. Deep facial age estimation using conditional multitask learning with weak label expansion. *IEEE Signal. Process. Lett.* **2018**, *25*, 808–812. [[CrossRef](#)]
41. Rattani, A.; Reddy, N.; Derakhshani, R. Convolutional neural network for age classification from smart-phone based ocular images. In Proceedings of the IEEE International Joint Conference on Biometrics, Denver, CO, USA, 1–4 October 2017; pp. 756–761.
42. Taheri, S.; Toygar, Ö. Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images. *IET Biom.* **2018**, *8*, 124–133. [[CrossRef](#)]
43. Vishwakarma, V.P.; Pandey, S.; Gupta, M.N. A novel approach for face recognition using DCT coefficients re-scaling for illumination normalization. In Proceedings of the 15th International Conference on Advanced Computing and Communications, Guwahati, India, 18–21 December 2007; pp. 535–539.
44. Du, S.; Ward, R.K. Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions. *IEEE Trans. Circuits Syst. Video Technol.* **2010**, *20*, 1165–1175. [[CrossRef](#)]
45. Vidya, V.; Farheen, N.; Manikantan, K.; Ramachandran, S. Face recognition using threshold based DWT feature extraction and selective illumination enhancement technique. *Procedia Technol.* **2012**, *6*, 334–343. [[CrossRef](#)]
46. Le, H.A.; Kakadiaris, I.A. SeLENNet: A semi-supervised low light face enhancement method for mobile face unlock. In Proceedings of the International Conference on Biometrics, Crete, Greece, 4–7 June 2019; pp. 1–8.
47. Huang, Y.-H.; Chen, H.H. Face recognition under low illumination via deep feature reconstruction network. In Proceedings of the IEEE International Conference on Image Processing, Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2161–2165.
48. Liang, J.; Wang, J.; Quan, Y.; Chen, T.; Liu, J.; Ling, H.; Xu, Y. Recurrent exposure generation for low-light face detection. *arXiv* **2020**, arXiv:2007.10963v1.
49. Kim, G.; Kwon, D.; Kwon, J. Low-lightgan: Low-light enhancement via advanced generative adversarial network with task-driven training. In Proceedings of the IEEE International Conference on Image Processing, Taipei, Taiwan, 22–25 September 2019; pp. 2811–2815.
50. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Gool, L.V. Wespe: Weakly supervised photo enhancer for digital cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 691–700.
51. Meng, Y.; Kong, D.; Zhu, Z.; Zhao, Y. From night to day: GANs based low quality image enhancement. *Neural Process. Lett.* **2019**, *50*, 799–814. [[CrossRef](#)]
52. Maeng, H.; Liao, S.; Kang, D.; Lee, S.W.; Jain, A.K. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In Proceedings of the Asian Conference on Computer Vision, Daejeon, Korea, 5–9 November 2012; pp. 708–721.
53. Baradarani, A.; Wu, Q.M.J.; Ahmadi, M. An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering. *Pattern Recognit.* **2013**, *46*, 57–72. [[CrossRef](#)]
54. Kang, D.; Han, H.; Jain, A.K.; Lee, S.W. Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching. *Pattern Recognit.* **2014**, *47*, 3750–3766. [[CrossRef](#)]
55. Shen, J.; Li, G.; Yan, W.; Tao, W.; Xu, G.; Dia, D.; Green, P. Nighttime driving safety improvement via image enhancement for driver face detection. *IEEE Access* **2018**, *6*, 45625–45634. [[CrossRef](#)]
56. Cho, S.W.; Baek, N.R.; Kim, M.C.; Koo, J.H.; Kim, J.H.; Park, K.R. Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network. *Sensors* **2018**, *18*, 2995. [[CrossRef](#)] [[PubMed](#)]
57. Viola, P.; Jones, M.J. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [[CrossRef](#)]
58. Wang, H.L.; Wang, J.-G.; Yau, W.-Y.; Chua, X.L.; Tan, Y.P. Effects of facial alignment for age estimation. In Proceedings of the 11th International Conference on Control Automation Robotics & Vision, Singapore, 7–10 December 2010; pp. 644–647.
59. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976.
60. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.

61. Zhou, Y.; Berg, T.L. Learning temporal transformations from time-lapse videos. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 262–277.
62. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
63. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
64. Rothe, R.; Timofte, R.; Gool, L.V. Dex: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 11–12 December 2015; pp. 252–257.
65. Cox, D.R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1958**, *20*, 215–242. [CrossRef]
66. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
67. MORPH Database. Available online: https://ebill.uncw.edu/C20231_ustores/web/store_main.jsp?STOREID=4 (accessed on 17 May 2021).
68. Zhu, Y.; Li, Y.; Mu, G.; Guo, G. A study on apparent age estimation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 11–12 December 2015; pp. 267–273.
69. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015; pp. 1–9.
70. FGNET Database. Available online: https://yanweifu.github.io/FG_NET_data/index.html (accessed on 17 May 2021).
71. AFAD Database. Available online: <https://afad-dataset.github.io> (accessed on 17 May 2021).
72. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2010.
73. Cho, S.W.; Baek, N.R.; Koo, J.H.; Park, K.R. Modified perceptual cycle generative adversarial network-based image enhancement for improving accuracy of low light image segmentation. *IEEE Access* **2021**, *9*, 6296–6324. [CrossRef]
74. Koo, J.H.; Cho, S.W.; Baek, N.R.; Park, K.R. Multimodal human recognition in significantly low illumination environment using modified EnlightenGAN. *Mathematics* **2021**, *9*, 1934. [CrossRef]
75. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467v2.
76. NVIDIA GeForce GTX 1070. Available online: <https://www.nvidia.com/en-in/geforce/products/10series/geforce-gtx-1070/> (accessed on 17 May 2021).
77. Python. Available online: <https://www.python.org/> (accessed on 17 May 2021).
78. OpenCV. Available online: <http://opencv.org> (accessed on 17 May 2021).
79. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 24–27 October 2017; pp. 2242–2251.
80. Mejjati, Y.A.; Richardt, C.; Tompkin, J.; Cosker, D.; Kim, K.I. Unsupervised attention-guided image-to-image translation. In Proceedings of the 32nd Conference on Neural Information Processing Systems, Montreal, Canada, 4–6 December 2018; pp. 1–11.
81. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. *arXiv* **2018**, arXiv:1805.08318v2.
82. Stathaki, T. *Image Fusion: Algorithms and Applications*; Academic: Cambridge, MA, USA, 2008.
83. Salomon, D. *Data Compression: The Complete Reference*, 4th ed.; Springer: New York, NY, USA, 2006.
84. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
85. Huynh-Thu, Q.; Ghanbari, M. The accuracy of PSNR in predicting video quality for different video scenes and frame rates. *Telecommun. Syst.* **2012**, *49*, 35–48. [CrossRef]
86. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]
87. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 24–27 October 2017; pp. 618–626.