mathematics

MDPI

Article

LCA-GAN: Low-Complexity Attention-Generative Adversarial Network for Age Estimation with Mask-Occluded Facial Images

Se Hyun Nam, Yu Hwan Kim, Jiho Choi, Chanhum Park and Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea
* Correspondence: parkgr@dongguk.edu

Abstract: Facial-image-based age estimation is being increasingly used in various fields. Examples include statistical marketing analysis based on age-specific product preferences, medical applications such as beauty products and telemedicine, and age-based suspect tracking in intelligent surveillance camera systems. Masks are increasingly worn for hygiene, personal privacy concerns, and fashion. In particular, the acquisition of mask-occluded facial images has become more frequent due to the COVID-19 pandemic. These images cause a loss of important features and information for age estimation, which reduces the accuracy of age estimation. Existing de-occlusion studies have investigated masquerade masks that do not completely occlude the eyes, nose, and mouth; however, no studies have investigated the de-occlusion of masks that completely occlude the nose and mouth and its use for age estimation, which is the goal of this study. Accordingly, this study proposes a novel low-complexity attention-generative adversarial network (LCA-GAN) for facial age estimation that combines an attention architecture and conditional generative adversarial network (conditional GAN) to de-occlude mask-occluded human facial images. The open databases MORPH and PAL were used to conduct experiments. According to the results, the mean absolution error (MAE) of age estimation with the de-occluded facial images reconstructed using the proposed LCA-GAN is 6.64 and 6.12 years, respectively. Thus, the proposed method yielded higher age estimation accuracy than when using occluded images or images reconstructed using the state-of-the-art method.

Keywords: facial age estimation; conditional GAN; mask-occluded facial images; LCA-GAN; MORPH and PAL

MSC: 68T07; 68U10

Citation: Nam, S.H.; Kim, Y.H.; Choi, J.; Park, C.; Park, K.R. LCA-GAN: Low-Complexity Attention-Generative Adversarial Network for Age Estimation with Mask-Occluded Facial Images. *Mathematics* **2023**, *11*, 1926. https://doi.org/10.3390/math11081926

Academic Editors: Lei Zhang and Wei Wei

Received: 20 February 2023
Revised: 9 April 2023
Accepted: 17 April 2023
Published: 19 April 2023

1. Introduction

In general, age, gender, expression, and race can be derived from facial appearances [1]. Age estimation is being increasingly used in diverse fields, such as statistical marketing analysis based on age-specific product preferences, medical fields such as the beauty industry and telemedicine, and age-based suspect tracking in intelligent surveillance camera systems [2]. Despite continuous efforts in research, including design of age estimation algorithms and models, data collection, system performance tests, and valid evaluation protocols, improving the accuracy of age estimation remains a challenge [3]. Age estimation is challenging because the human face is influenced by internal factors (size, wrinkles, shape, texture, race, etc.) and external factors (health, dietary habits, culture, environment, etc.), and these change over time through complex processes [4]. However, there are general and common features that can explain human facial aging [5]. Age estimation is divided into feature representation, extraction, and age learning stages. Though previous studies used various handcrafted feature-based methods, they require accurate prior knowledge of experts. However, there are no methods to verify the accuracy of the prior knowledge [6]. Unlike conventional methods, a convolutional neural network (CNN) can extract clear and

robust facial features and learn the age on its own [7]. Generally, a CNN for age estimation consists of a convolutional layer and a multi-layer perceptron (MLP). The convolutional layer extracts and represents age information features from the facial images, and the MLP estimates the age with the represented features. Then, the distance between the estimated age and the age label is calculated by the loss function and then backpropagations are performed. The whole process is performed automatically, and unlike conventional methods that rely on prior knowledge, it can utilize information that cannot be extracted with prior knowledge. It provides better age estimation performance than traditional methods, and numerous researchers are actively conducting research on improving the results and accuracy.

Despite the advances in age estimation research, there are several problems when age estimation is applied practically. Facial images obtained in real unrestricted environments frequently have problems that degrade image quality related to resolution, illumination, noise, and occlusion. In particular, mask-occluded facial images have recently become frequent following the COVID-19 pandemic. Masks are worn more frequently because of hygiene, personal privacy concerns, and fashion. Mask-occluded facial images do not contain important features and information for age estimation, which reduces the accuracy of age estimation. Existing de-occlusion research has investigated masquerade masks that do not completely occlude the eyes, nose, and mouth [8–11], but no studies have investigated the de-occlusion of masks that completely occlude the nose and mouth, which is the goal of this study. To this end, this study proposes a novel low-complexity attention-generative adversarial network (LCA-GAN) for facial age estimation that combines an attention architecture and conditional generative adversarial network (conditional GAN) to de-occlude mask-occluded human facial images. Our innovation is for facial image de-occlusion. The present study is new compared to previous studies in four ways:

- This is the first study of its kind on age estimation that considers the de-occlusion of facial images where the nose and mouth are completely occluded by a mask;
- We propose a novel LCA-GAN for mask de-occlusion. LCA-GAN contains low-complexity attention blocks (LCABs) that reduce computation and complexity by combining down and upsampling with the attention module. LCAB comprises low-complexity channel attention (LCCA) and low-complexity spatial attention (LCSA), and it uses attention to assign weights based on the importance of features in channel and spatial dimensions;
- To reconstruct the facial feature information lost by mask occlusion as much as possible in de-occlusion, edge loss and content loss in LCA-GAN were used;
- The trained LCA-GAN and CNN for age estimation and experimental mask generated facial images were published [12], enabling a fair comparison with the performance of other researchers.

The structure of this paper is as follows. Section 2 analyzes existing age estimation studies that have used facial images and de-occlusion methods. Section 3 explains the overall experimental method and LCA-GAN, the de-occlusion network proposed in this paper. Section 4 presents a comparison of the performance and age estimation results using de-occluded images between existing de-occlusion methods and LCA-GAN based on the MORPH and PAL databases. As a final section, Section 5 concludes the paper.

2. Related Works

Facial images contain biological information with diverse attributes, such as race, gender, age, environment, and lifestyle. In [13], the distributions of these attributes and their averages were analyzed, and a method was presented to evaluate the bias of appropriate algorithms and databases. It has influenced various studies using human facial images [14–16]. However, it is difficult to investigate handcrafted feature-based age estimation, which requires consideration of diverse factors. Consequently, most age estimation studies have used CNNs since the emergence of deep learning. As shown in Table 1, age estimation methods are generally classified into five categories [6]. Multi-class classifi-

cation yields high age estimation performance when using limited resources and data with a single model. Hybrid methods, which combine multiple methods, supplement the shortcomings of combined models and yield high age estimation performance for large quantities of data in environments with many available computing resources. To measure the age estimation accuracy, researchers have used mean absolute error (MAE) [17], exact accuracy [18], 1-off [19], and normal score (ϵ-error) [20], among which MAE is the most commonly used.

Table 1. Classification and comparison of existing age estimation methods.

Categories	Method	Database	MAE	Exact	1-Off	ϵ-Error
Classification of multi-class ages	DEX [21]	IMDB-WIKI + LAP2015	3.22	N.A.	N.A.	0.26
	Residual DEX [22]	LAP2015	4.45			N.A.
	Dimensionality reduction + FFNNs [23]	WIKI + AmI-Face + Adience	3.30			
	4C2FC [24]	MORPH	N.A.	46.39		
	RoR [25]	IMDB-WIKI + Adience		67.3	97.51	
	DEX [8]	MORPH	2.68	N.A.	N.A.	
		FG-NET	3.09			
		CACD	6.52			
		IMDB-WIKI + LAP2015	N.A.	64.0		
		Adience		N.A.	96.6	0.26
	4C2FC + dropout [9]	Adience	N.A.	84.8	89.7	N.A.
Regression based on metrics	3NNR [26]	Adience + MORPH + LAP2015	N.A.	N.A.	N.A.	0.37
	OR-CNN [27]	AFAD	3.34			N.A.
		MORPH	3.27			
	VGG + BridgeNet [28]	MORPH	2.38			
		FG-NET	2.56			0.26
		LAP2015	2.98			
Learning by the distribution of deep label	DLDL-v2 [29]	LAP2015	3.14			0.272
		LAP2016	3.45			0.267
		MORPH	1.97			
	Inception v4 [30]	MORPH	1.32			N.A.
		FG-NET	2.19			
Ranking	Ranking-CNN [31]	MORPH	2.96			
	ODFL + OHRank [32]	MORPH	3.12			
		FG-NET	3.89			N.A.
		LAP2016	4.12			0.34
		Adience	N.A.	54.0	88.2	N.A.
	ODL [33]	MORPH	2.92			N.A.
		FG-NET	3.71			0.312
		LAP2016	3.95			
Hybrid methods	Kernel ELM + CNN [34]	LAP2016	N.A.	N.A.	N.A.	0.37
	MRCNN [35]	MORPH	3.48			N.A.
	GA-DFL [36]	MORPH	3.25			N.A.
		FG-NET	3.93			N.A.
		LAP2015	4.21			0.37
	CNN + ELM [37]	MORPH	3.44	N.A.		N.A.
		Adience	N.A.	52.3		
	RAGN [10]	IMDB-WIKI + MORPH	2.61	N.A.		N.A.
		IMDB-WIKI + Adience	N.A.	66.5		0.37
		IMDB-WIKI + LAP2016	N.A.	N.A.		
	AgeNet + divide and rule [11]	FG-NET	4.02	N.A.		N.A.
		MORPH	3.48			
		IMDB-WIKI	3.29			
	MA-ShuffleNet v2 [38]	MORPH	2.68			
		FG-NET	3.81			

As listed in Table 1, age estimation studies have used images obtained in restricted environments, such as the MORPH [39] and FG-NET [40] databases, and those from unre-

stricted environments such as IMDB-WIKI [41], Adience [42], LAP2015 [43], LAP2016 [44], and CACD [45]. MORPH is a database of human facial mugshot images with resolutions ranging from 640 × 480 to 1024 × 768 pixels. It contains various attributes, such as gender, age, and race, and is acquired in a restricted environment such as image resolution, illumination, and pose. FG-NET is a database of human facial images with gender and age information. The database collected facial images that satisfy specific conditions such as image resolution and pose from pictures of people with confirmed ages. In this case, the facial images that satisfy the conditions are similar to the facial images acquired in a restricted environment. On the other hand, the databases in unrestricted environment are collected from the internet, magazines, films, etc. These unrestricted databases contain facial images in natural poses, various image resolutions and illumination changes, and occlusions with various objects. Age estimation research using images from restricted environments has yielded relatively low age estimation accuracy for occluded images. Despite the difficulty of age estimation due to occlusion, previous age estimation studies have not considered the de-occlusion of facial images occluded by a mask that completely covers the nose and mouth. In addition, all datasets in Table 1 have a very small number of masked face images. Therefore, for our experiments, we generated a large number of masked face images from the MORPH and PAL databases. Table 2 presents a comparison of studies that do and do not consider face occlusion with the proposed method.

Table 2. Comparison of the strengths and weaknesses of existing research and proposed methods in age estimation based on consideration of face occlusion.

Categories	Age Learning Technique	Method	Strength	Weakness
Age estimation without considering face occlusion	Handcrafted feature-based	Guo et al. [46]	Age estimation robust to restricted environment	They did not consider face-occluded images for age estimation
		Chen et al. [47]		
	Deep feature-based	Inception v4 [30]		
		MA-ShuffleNet v2 [38]		
Age estimation with considering face occlusion		DEX [8]	Age estimation robust to occluded facial images	They trained simultaneously with occlusion and non-occlusion images, which made network convergence difficult
		AgeNet + divide and rule [11]		
		RAGN [10]		
		4C2FC + dropout [9]		
		Proposed method		Additional procedures are required to train LCA-GAN

A study [30] proposed recurrent age estimation (RAE), which combines inception-v4 [48] and long short-term memory networks (LSTM) [49]. It extracts features from facial images using inception-v4 and learns individual aging patterns with LSTM. To solve the problem of training overfitting, researchers have proposed label distribution learning (LDL), which uses the ambiguity between the label age and predict age. A study [38] proposed mixed attention-ShuffleNet-v2 (MA-SFV2), which combines mixed attention and ShuffleNet-v2 [50]. Classification, regression, and distribution methods were simultaneously applied to learning to transform the output layer of the base model. In [10], CNN2ELM, an ensemble model combining CNN and extreme learning machine (ELM), was proposed for leaning age. It achieved decent results in ChaLearn 2016 [51], a human age estimation competition. In [8], a deep expectation of apparent age (DEX) system was proposed, which uses the softmax expected value refinement of the VGG-16-based network [21]. More specifically, they defined the regression problem of age learning as a classification problem and performed age estimation by multiplying the age label and the

class probability distribution, which is the last softmax output of VGG-16. A study [11] proposed the divide-and-rule architecture and AgeNet, a model based on the method of GoogleNet [52]. AgeNet is a feature extractor and divide-and-rule approach to age learning as an ordinal regression problem. As such, research on age estimation includes studies using restricted environment images [30,39,47,48], studies using unrestricted environment images [9,11], and studies using images of both environments [9,21,38]. This age estimation research includes many studies that excluded occlusion in restricted environments or ignored occlusion in unrestricted environments, so de-occlusion was not considered.

However, face occlusion frequently occurs in the real world and is challenging to solve through camera hardware. Most existing de-occlusion methods reconstruct synthesized images because of the lack of databases with non-occluded and occluded image pairs [53–56]. In [53], a two-stage occlusion-aware GAN was proposed, trained with 44 images occluded by sunglasses, hats, scarves, and phones with a random shape, location, and size in a face database, but it did not use facial mask images. This method removes occluded areas with the existing Pix2pix-based [57] GAN architecture and de-occludes the image only using information from non-occluded areas. However, this is an unrealistic occlusion condition, and while de-occlusion is successful for most objects, it fails for glasses and sunglasses. Previous research proposed a method to solve the face recognition problem associated with block occlusion of facial images by two robust feature-based representations, which were designed to fit the errors to a distribution described by a tailored loss function and the reduced rank structure of the errors relative to the image size [58]. The work in [59] proposed MRGAN, a GAN-based two-level network that removed the areas occluded by medical masks and reconstructed the removed areas. Stage one detected masks, and stage two performed de-occlusion. This method is based on a complex network and is computationally intensive. This experiment achieved good de-occlusion results but did not reconstruct color and detail information well in the occluded area. The work in [60] proposed a two-stage GAN-based method for de-occlusion of small objects in facial image such as microphones. Steps one and two were trained similarly to a conventional GAN, but the de-occluded image from step one was used as an input to step two to generate a robust de-occluded image. The de-occlusion results showed that the texture and detail were de-occluded well, but the outline of the occluded object remained.

Moreover, previous study [54] presented a two-stage method; in stage one, the occluded area is detected with an encoder–decoder structure and converted to a mask image as a pre-processing method, and in stage two, the occluded area is transformed through two conditional GAN [61] architectures. For supervised learning, face images were collected from the CelebA database, and the collected images were synthesized with the collected mask images occluding the eyes using Photoshop CC 2080 [54]. This method requires additional binary mask images to be trained, and the de-occlusion results fail on complex and detailed mask images. Another study [55] proposed Swap-R&R that compensates for the lack of paired databases. This method shows very robust de-occlusion performance with facial images occluded by glasses, sunglasses, makeup, and headsets, but it requires an additional 3D face reconstruction network in training and testing, and the network computation is very large [55]. In [56], they reconstructed identity-preserved and de-occluded facial images by CNN, which was supervised with identity labels. By using the additional channel for occlusion detection, a mask for occlusion is computed as a pre-processing method and combined with the reconstructed face. This experiment collected frontal face images between −45 and 45 degrees from the CASIA-WebFace database. The collected face images are used to synthesize multiple objects for supervised learning. There are more than 100 templates for occluding objects such as masks, glasses, and hands [56], and among them, objects such as glasses and masks require accurate position information. Objects that require such location information are synthesized directly using an image program, while other objects are synthesized randomly. However, they only deal with grayscale facial images and produce results including artifacts. Though de-occlusion research has been conducted on masquerade masks that do not completely occlude the eyes, nose, and

mouth [54], no studies have investigated de-occlusion-based age estimation of masks that completely occlude the nose and mouth, which is the goal of this study. To solve the aforementioned problems, this study proposes LCA-GAN-based facial mask de-occlusion and an age estimation method using this technique.

3. Proposed Methods

3.1. Overview of Suggested Method

Figure 1 illustrates the entire procedure of robust age estimation for mask-occluded facial images proposed in this study.

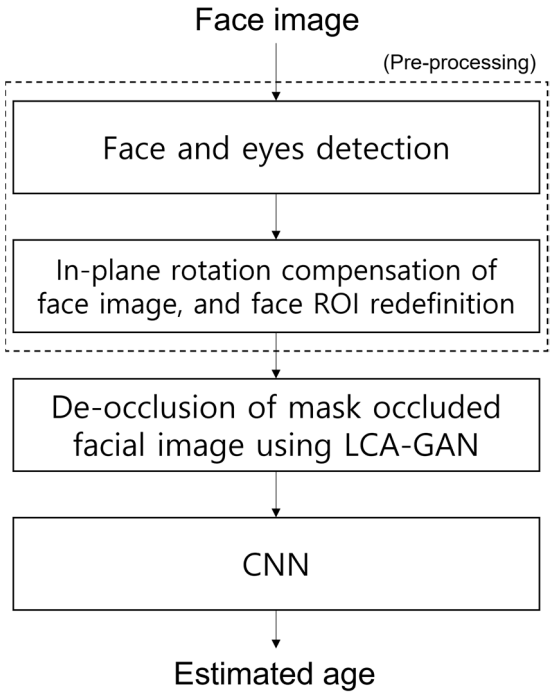


Figure 1. The entire procedure of our LCA-GAN-based de-occlusion and age estimation.

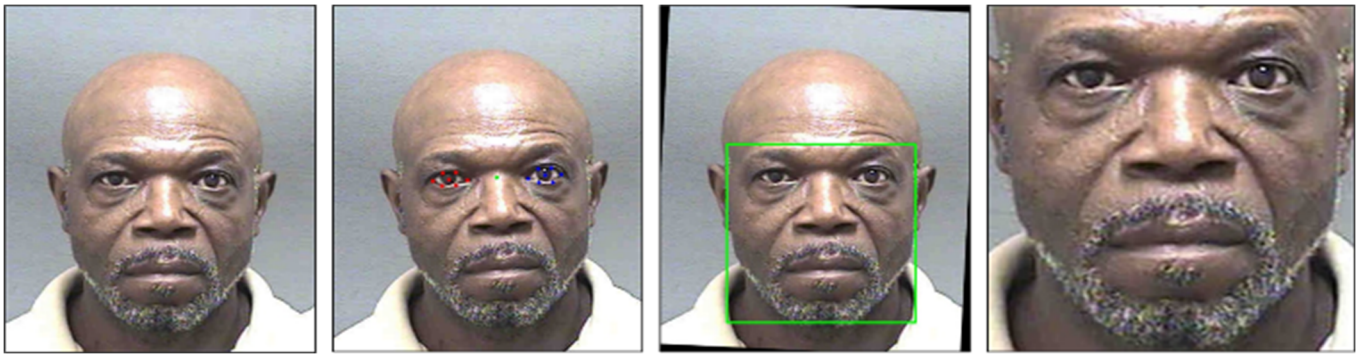
Stages one and two entail pre-processing. First, the positions of the face and eyes are detected, and based on the detected positions, we execute the compensation of in-plane rotation, and the face region of interest (ROI) is re-defined. Following pre-processing, the mask-occluded image is de-occluded using the proposed LCA-GAN. The CNN-based age estimation method is then adopted to predict the age of the person in the de-occluded image.

3.2. Pre-Processing

As shown in Figure 2, the input images are processed. For pre-processing, this study used the dlib facial feature tracker [62], which extracted features with histogram of oriented gradients (HOG) and trained a linear classifier on the extracted information. It does not use any parameters or thresholds except for the upsample_num_time option, which is used for iterating the detection while scaling the input image, and we used the default value of one in our experiment. It was used on original face images of different sizes to detect face landmarks and the face box region. First, the dlib facial feature tracker for facial feature points [62] is used to locate the positions of the eyes, as shown in Figure 2b. Using Equation (1) based on the positions of both detected eyes, in-plane rotation compensation is performed, as shown in Figure 2c. In-plane rotation compensation is performed based on the center position between both eyes (green dot between the eyebrows in Figure 2b).

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

(R_x, R_y) and (L_x, L_y) represent the x- and y-axis positions of the centers of the detected right and left eyes, respectively. In the in-plane rotation compensated image, the face region is found using the dlib facial feature tracker, as shown in Figure 2c, and this face region becomes the face ROI, as shown in Figure 2d. To use the face ROI as an input for LCA-GAN, it is resized to 256 × 256 × 3 by bilinear interpolation.



(**a**) (**b**) (**c**) (**d**)

Figure 2. Exemplary process of face ROI definition. (**a**) Example image of original MORPH database. (**b**) Detected face and eyes region using dlib facial feature tracker. (**c**) In-plane rotated facial image with detected face region. (**d**) Re-defined face ROI image.

3.3. De-Occlusion of Masked Facial Image by LCA-GAN

General image quality problems that distort image information include low resolution, blur, low illumination, noise, etc. However, regarding the occlusion problem, the occluding object reduces image information. Hence, unlike image quality problems where information can be directly extracted from the distorted area, for the occlusion problem, it is more difficult to directly extract information from the occluded area. Consequently, it is difficult to learn the mapping from occluded to de-occluded images using a CNN. To solve this problem, this study proposes LCA-GAN based on adversarial learning. Figure 3 shows the architecture of LCA-GAN. The generator is based on U-net [63]; by using LCAB, which combines up and downsampling with channel and spatial attention, it reduces complexity and computation. For the discriminator, a patch discriminator is used to output with a size of 30 × 30 × 1. It obtains probabilities for each of the 30 × 30 patches and determines whether each of the 900 local patch areas is a real or fake image. Finally, it averages all 900 probabilities to finally determine if the entire global image is real or fake.

3.3.1. Generator

Figure 3a shows the generator for de-occlusion in this study, which uses the U-net [63] architecture comprising an encoder–decoder and skip connection. In the encoder–decoder used for U-net's continuous down and upsampling, the encoder extracts features, and the decoder learns the mapping for image patches corresponding to the extracted features. Additionally, it concatenates high-stage encoder block features with low-stage decoder block features to compensate for lost high-level information and to balance high- and low-level information. However, this architecture is inefficient for the occlusion problem, where it is difficult to directly extract information from the occluded area. To solve this problem, this study proposes LCAB, which combines an attention mechanism [64] with down and upsampling. LCAB is composed of LCCA and LCSA; attention is used to assign weights according to the importance of features in the channel and spatial dimensions. The next subsection describes LCAB architecture in detail. When de-occluding the occluded area, the generator uses edge loss to create a detailed and sharp image as well as content loss to maintain the information of the non-occluded area and the texture in the target image. In this de-occlusion process, $\mathcal{L}_1$ loss function was used for identity loss to maintain the information of the original image. Table 3 presents the overall architecture of the

generator. As shown in Table 3, an image of 256 × 256 × 3 including the mask-occluded area is inputted to the generator of LCA-GAN, and a feature map of 256 × 256 × 64 is obtained via Convolution layer 1 and Spatial Attention. Then, by passing through LCAB 1~5 (including LCCA and LCSA), the feature map of 8 × 8 × 512 is obtained as the final output of encoder. This feature map is again passing through the decoder of LCAB 6~10 (including LCCA, Concatenation, and LCSA except for LCAB 10 including only LCCA and LCSA) and the upsampled feature map of 256 × 256 × 64 is obtained. Then, this feature map is passing through Convolution layer 2 and *Tanh* activation layer, and the final generated (mask-de-occluded) image of 256 × 256 × 3 is obtained as the output of decoder.

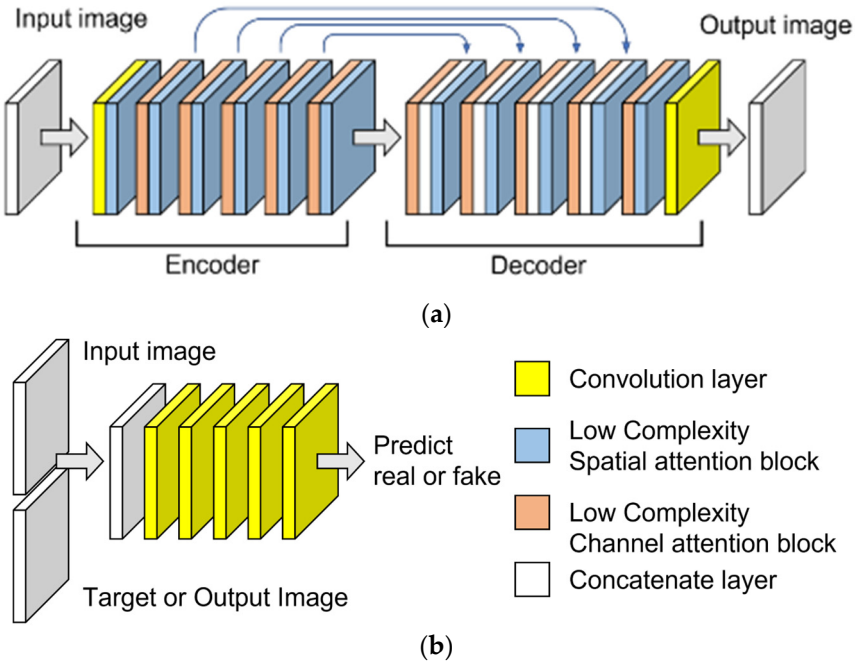


Figure 3. LCA-GAN architecture. (**a**) Generator. (**b**) Discriminator.

3.3.2. The Structure of LCAB

The attention mechanism abstracts the importance between the modalities of the represented features to incorporate high-level information. Representative attention methods used in images include spatial attention and channel attention [64]. Channel and spatial attention determine which features are important in the channel and spatial dimensions, respectively, and assign corresponding weights. In this study, channel and spatial attention were used to detect and de-occlude mask-occluded areas. Moreover, we proposed LCAB, which combines down and upsampling with the attention process to reduce the complexity and computation that increases due to attention and continuous processes after down or upsampling. LCAB comprises LCCA and LCSA, the structures of which are illustrated in Figures 4 and 5.

Channel attention has two convolutional layers, average pooling, three multi-layered perceptrons (MLPs), a sigmoid layer, multiplication with convolutional layer features, and addition with input features passed through skip-connection. The importance of channel dimension features is arranged using the two-stage convolution layer to re-represent the input-represented features. The features are then compressed by global average pooling in the spatial space, and MLP is used to calculate the importance of modalities between the features in the channel space. Here, the first and third layers of MLP are equal to the channel dimension size of the input features, and the second layer is 1/4 the channel dimension size of the input features. Using sigmoid activation, attention is constructed from the importance of these arranged features. Following the convolutional layer, it is multiplied with the features and then added to the input features with high-level information. This

study proposes LCCA, which combines down and upsampling with channel attention. In the two-stage convolutional layer, the first stage uses a 4×4 size filter to scale the spatial space by 2 in upsampling and by $1/2$ in downsampling. The second stage uses a 3×3 size filter to maintain the size of the spatial space. Moreover, the deformation of high-level information is minimized using bilinear interpolation, and the skip-connection of the input features is passed.

Table 3. Generator architecture in LCA-GAN.

Layer			Size of Feature	Concatenation
Input image			$256 \times 256 \times 3$	-
Encoder	Convolution layer 1		$256 \times 256 \times 64$	-
	Spatial attention		$256 \times 256 \times 64$	-
	LCAB 1	LCCA LCSA	$128 \times 128 \times 64$ $128 \times 128 \times 128$	-
	LCAB 2	LCCA LCSA	$64 \times 64 \times 128$ $64 \times 64 \times 256$	-
	LCAB 3	LCCA LCSA	$32 \times 32 \times 256$ $32 \times 32 \times 512$	-
	LCAB 4	LCCA LCSA	$16 \times 16 \times 512$ $16 \times 16 \times 512$	-
	LCAB 5	LCCA LCSA	$8 \times 8 \times 512$ $8 \times 8 \times 512$	-
Decoder	LCAB 6	LCCA Concatenation LCSA	$16 \times 16 \times 512$ $16 \times 16 \times 1024$ $16 \times 16 \times 512$	LCAB4
	LCAB 7	LCCA Concatenation LCSA	$32 \times 32 \times 512$ $32 \times 32 \times 1024$ $32 \times 32 \times 512$	LCAB3
	LCAB 8	LCCA Concatenation LCSA	$64 \times 64 \times 512$ $64 \times 64 \times 768$ $64 \times 64 \times 256$	LCAB2
	LCAB 9	LCCA Concatenation LCSA	$128 \times 128 \times 256$ $128 \times 128 \times 384$ $128 \times 128 \times 128$	LCAB1
	LCAB 10	LCCA LCSA	$256 \times 256 \times 128$ $256 \times 256 \times 64$	-
	Convolution layer 2 *Tanh* activation layer		$256 \times 256 \times 3$	-
Generated image			$256 \times 256 \times 3$	

As shown in Figure 5, spatial attention is composed of two convolutional layers, global average pooling, a convolutional layer, a sigmoid layer, multiplication with convolutional layer features, and addition with input features passed through skip-connection. The importance of spatial dimension features is arranged using the two-stage convolution layer to re-represent the input-represented features. The features are then compressed by global average pooling in the channel space, and the importance of modalities between the features in the spatial space is calculated through the convolutional layer. Using sigmoid activation, attention is created from the importance of these arranged features. Following the convolutional layer, it is multiplied with the features and then added to the input features with high-level information. This study proposes LCSA, which combines down and upsampling with spatial attention. The two-stage convolutional layer maintains the

spatial space using a 3 × 3 size filter but adjusts the channel dimension to the desired size. After global average pooling, the convolutional layers use a 7 × 7 size filter to calculate the importance of modalities between features in the spatial space and then represent them as probability values using the sigmoid activation layer. Additionally, the size of the channel space of the input features is adjusted using bilinear interpolation, the deformation of high-level information is minimized, and the information is passed. Through LCCA and LCAS, channel and spatial attention preserve the original goal of representing the importance of the represented features from the perspectives of "what" and "where", combining the down and upsampling processes. Moreover, they reduce computation and complexity caused by the use of continuous down and upsampling and attention mechanisms, and they are effective for the de-occlusion process.

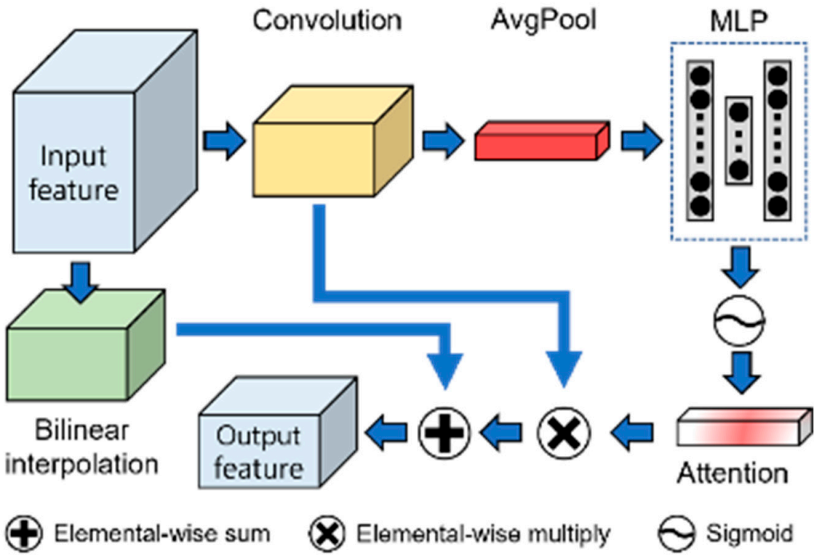


Figure 4. Structure of LCCA (AvgPool and MLP mean average pooling and multi-layered perceptron, respectively).

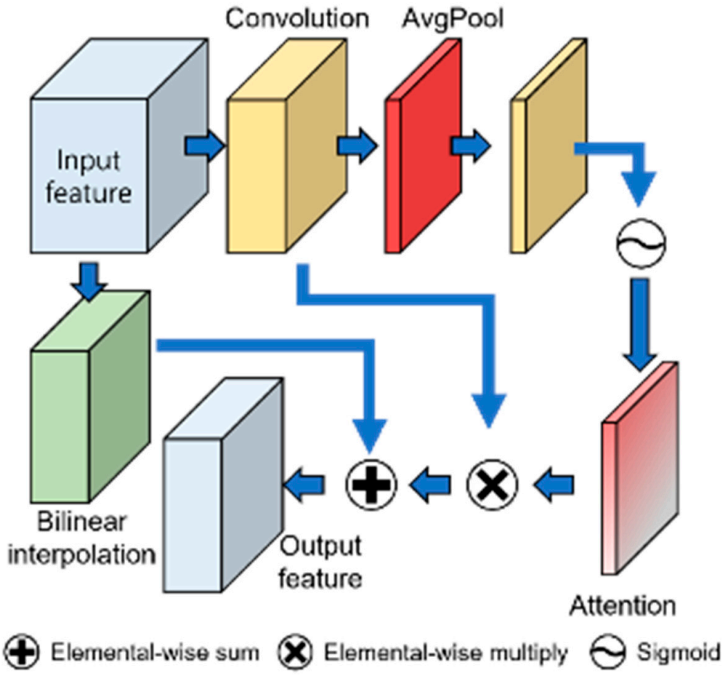


Figure 5. Structure of LCSA.

3.3.3. Discriminator

LCA-GAN uses a patch discriminator. It receives a fixed input image and a random target image and output image, and then it concatenates them. Convolution is then performed, and each grid of extracted features has a receptive field according to the computational structure. In this study, individual grids of $1 \times 1 \times 1$ units of the final output, which has a size of $30 \times 30 \times 1$, have a receptive field of 70×70, use individual grids to determine the local area, and then distinguish the global area with the average of all grids. In this process, identity loss and content loss are used to maintain the continuity of information in the input image. Table 4 lists the detailed structure of the discriminator.

Table 4. Discriminator architecture in LCA-GAN. CL and BN mean convolution layer and batch normalization, respectively.

Layer		Size of Feature
Input image		$256 \times 256 \times 3$
Target or de-occluded image		$256 \times 256 \times 3$
Concatenate		$256 \times 256 \times 6$
CL 1	Convolution BN ReLU	$128 \times 128 \times 64$
CL 2	Convolution BN ReLU	$64 \times 64 \times 128$
CL 3	Convolution BN ReLU	$32 \times 32 \times 256$
CL 4	Zero padding Convolution BN Leaky ReLU	$34 \times 34 \times 256$ $31 \times 31 \times 512$
CL 5	Zero padding Convolution Sigmoid Average pooling	$33 \times 33 \times 512$ $30 \times 30 \times 1$ $1 \times 1 \times 1$
Output		Real or fake

LCA-GAN proposed in this study performs learning using pairs of images comprising the mask-occluded image (input image) and original un-occluded image (target image). Conditional GAN [57] learns the mapping using the loss function in Equation (2), which receives an input image I^{In} and generates an output image I^{Out} that is similar to the target image I^{Target}.

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

This adversarial learning method generates smooth pixel-wise images [65]. In this study, the generator used content loss to maintain the texture of the original face when de-occluding the mask-occluded area. For this purpose, the $\mathcal{L}_2$ loss function of Equation (3) is applied to the VGG-16 (pre-trained with ImageNet) to measure the dissimilarity of I^{Out} and I^{Target}.

$$\mathcal{L}_{cont} = \mathbb{E}_{I^{Out},I^{Target}}\left[\left(I^{Target} - I^{Out}\right)^2\right] \quad (3)$$

The proposed LCA-GAN preserves the area other than the mask-occluded area in the mask-occluded image and performs de-occlusion. The discriminator concatenates

pairs of I^{In} and I^{Out} or I^{In} and I^{Target}, which enables learning to reinforce high-frequency information in the mask-occluded area of I^{Target}. This is accomplished by applying the edge loss function in Equation (4) to generate a detailed de-occluded image. Δ denotes the Laplasian operation, and ε is 10^{-3}.

$$\mathcal{L}_{Edge} = \mathbb{E}_{I^{Out},I^{Target}}\left[\sqrt{\left(\Delta I^{Target} - \Delta I^{Out}\right)^2 + \varepsilon^2}\right] \tag{4}$$

Rather than learning the data distribution of I^{Target}, adversarial learning sometimes strongly tends toward receiving the discriminator's determination of the real image. To prevent this and preserve the identity of the image, we added identity loss, which uses the $\mathcal{L}_1$ loss function, as shown in Equation (5).

$$\mathcal{L}_{Iden} = \mathbb{E}_{I^{Out},I^{Target}}\left[\left|\left|I^{Target} - I^{Out}\right|\right|\right] \tag{5}$$

Finally, our final loss function is in Equation (6). Using the training data, 1.5, 2, and 2 were determined as the optimal values of λ_1, λ_2, and λ_3, respectively, to obtain the best age estimation accuracy.

$$\mathcal{L}_{LCA} = \arg\min_{G}\max_{D}\mathcal{L}_{GAN}(G,D) + \lambda_1\mathcal{L}_{cont} + \lambda_2\mathcal{L}_{Edge} + \lambda_3\mathcal{L}_{Iden} \tag{6}$$

3.4. Age Estimator

The DEX model [8] was used to predict the age of de-occluded facial images with LCA-GAN, which exhibited good results in the Looking at People (LAP) 2015 [66] competition and previous research results on age estimation accuracy [67]. DEX is an age estimation model based on VGG-16 [68], an existing classification network. For age estimation, VGG16 pre-trained on ImageNet was additionally pre-trained using the IMDB and WIKI databases. Since human aging typically involves sequential changes over time, the similarity between adjacent classes in DEX is high, so the probability of the trained model is considered to have a normal distribution. Therefore, rather than estimating the class label showing the highest probability score as age, the age was predicted as the product of the class label and probability value, as shown in Equation (7).

$$Estimated\ age(I) = \sum_1^n l_i p_i \tag{7}$$

where I denotes the input facial image, n indicates the number of classes, l_i denotes the ith class label, and p_i corresponds to the ith output probability value. The detailed age estimation methods of DEX are illustrated in Figure 6.

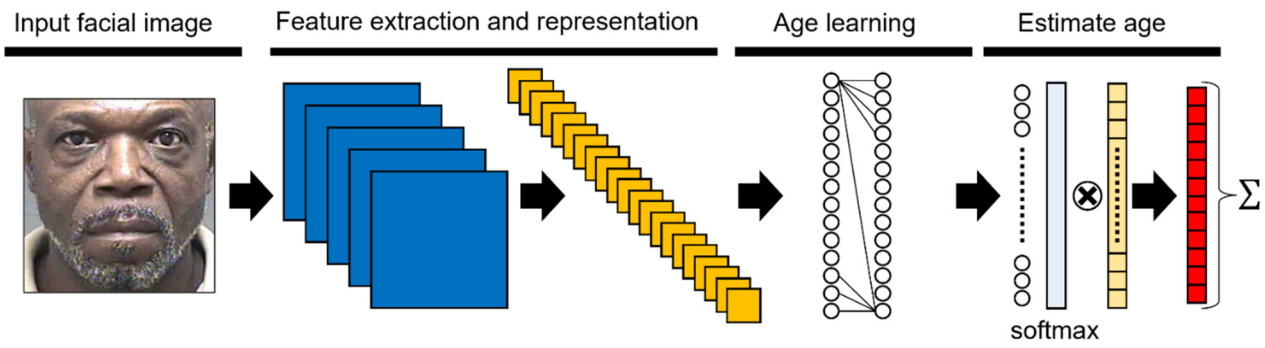


Figure 6. The overall procedure of age estimation with DEX.

DEX [8] is a VGG-16-based network with a convolution layer, which consists of a convolution filter, batch normalization, an ReLU activation function, two MLP layers with 4096 nodes, and an output MLP layer equal to the size of the age class labels. The

convolutional layers extract features with age information and represent the features for age learning. MLP layers learn age from the represented features. The last MLP layer outputs a probability value through a softmax activation function. Finally, to improve the age estimation performance, DEX applies the age estimation method described in Equation (7). In this experiment, we used the $\mathcal{L}_2$ loss function in DEX to emphasize the relation of close age classes, thus making the age estimation method using Equation (7) more robust.

4. Experimental Results

4.1. Data and Environment for Experiments

As shown in Figure 7, this study used MORPH [39] and PAL [69] as the databases to de-occlude mask-occluded facial images. Given the lack of open databases of mask-occluded facial images obtained in real environments that include existing age information, as shown in Figure 8, we generated mask-occluded facial images using mask images without a background image directly acquired from MORPH and PAL, which are existing human facial databases. Figure 8a shows the original facial image, and Figure 8b shows the mask image with no background. Subsequently, in the facial image, the dlib facial feature tracker [62] explained in Section 3.2 detects the eye area, as shown in Figure 8c. Using the center position of the eyes, the method described in Section 3.2 is used to perform in-plane rotation, as shown in Figure 8d; in this image, the dlib tracker for facial feature points finds the position of the face landmark corresponding to the annotated point of the mask image. Based on this position information, the annotated mask image is geometrically transformed and warped, as shown in Figure 8e. As shown in Figure 8f, the transformed mask image is then occluded on the aligned facial image, and the face ROI is detected again using the dlib tracker. The ROI is re-defined with the detected face ROI, and the final face ROI image of the mask-occluded image is created, as shown in Figure 8g.

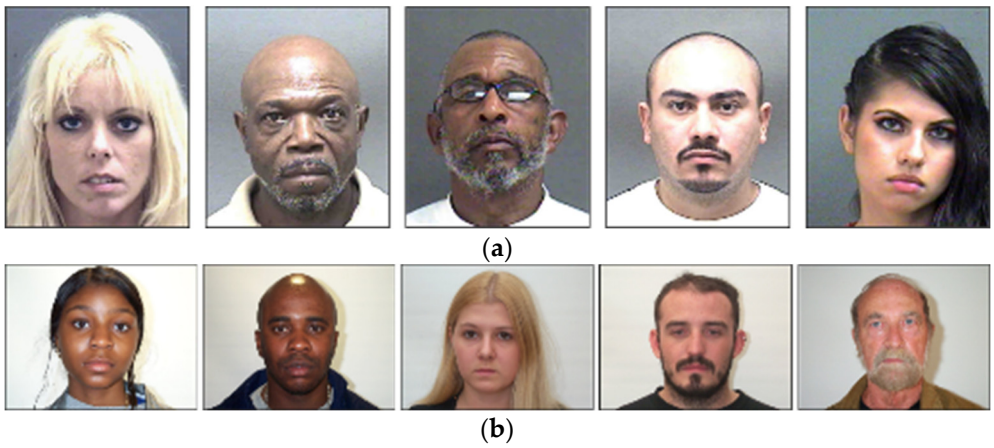


Figure 7. Examples of (**a**) MORPH and (**b**) PAL database.

Figure 8. Procedure of mask inclusion of facial images. (**a**) Original facial image, (**b**) original mask image, (**c**) eyes detected using dlib tracker, (**d**) in-plane rotation compensated image with detected jawlines, (**e**) warped mask image using geometric transformation, (**f**) mask occlusion image and detected face ROI, and (**g**) re-defined face ROI image of mask occlusion image.

The experiments were performed with two-fold cross validation, and in each fold, we used 5% of the training images as a validation set. To locate the face ROI, we used the python library (version 3.5.2) [70] and OpenCV (version 4.2.0) [71]. The specification of the desktop computer for our experiments is as follows: 3.5 GHz CPU (Intel® Core™ i7-3770K) and 24 GB RAM, Windows Tensorflow (version 2.2.0) [72], and Nvidia graphics processing unit (GPU) card (Nvidia GeForce GTX 1070 [73]).

4.2. Training of LCA-GAN for Masked Image De-Occlusion and CNN for Age Estimation

LCA-GAN proposed in this study performs learning using the mask-occluded facial image as the input image and the original facial image without a mask as the target image. During training, through online augmentation, the input images were resized to 286 × 286 × 3 and then randomly cropped to 256 × 256 × 3. The adaptive moment estimation (Adam) optimizer [74] was used during training, with a learning rate of 0.0002, beta_1 of 0.5, and beta_2 of 0.999. Training was conducted for 100 epochs; Figure 9 shows the training and validation loss graphs of the generator and discriminator of LCA-GAN. It is evident that the generator and discriminator converged, indicating that the training data were sufficiently learned. For validation loss, the results of the generator and discriminator converged, indicating that LCA-GAN was not overfitted to the training data. In the case of GAN, mode collapse usually occurs when a generator tries to map an input (training data set) to the same output (generator function). The discriminator and generator should be learning together and interacting with each other, but one becomes too well trained (learning imbalance), and mode collapse occurs [75]. As shown in Figure 9, the discriminator becomes too well trained compared to the generator in our experiment, which is usually the case for conventional GAN [75–77]. Therefore, although the mouth areas are different in the input images, those in the generated output images are somewhat similar, which represents a small level of mode collapse. However, other areas of face in the generated output images were different according to our LCA-GAN, which confirms that overfitting and mode collapse were not severe in our LCA-GAN.

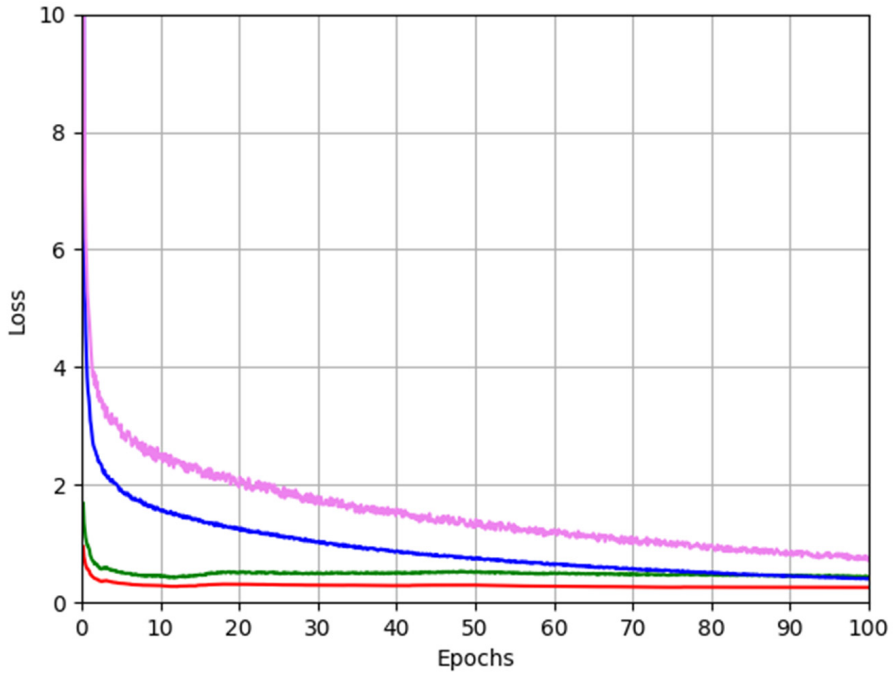


Figure 9. Graphs of LCA-GAN losses. Blue and violet lines represent the generator training and validation loss graphs, respectively. Red and green lines represent the discriminator training and validation loss graphs, respectively.

The slow convergence of the validation loss of the generator in Figure 9 was due to the insufficient number of MORPH databases used. In addition, considering the age labels ranging from 16 to 77 years old, various ethnicities, and unbalanced gender ratio in the MORPH database, the distribution of the data is very complex and unbalanced. These factors create a condition where the loss of validation data using only 5% of the training data can lead to late convergence. In this case, the difficulty of learning generally increases, and overfitting, underfitting, and divergence of losses are likely to occur. We described our training and validation sets according to the distribution of our experimental database in Table 5 of Section 4.2. As shown in Table 5, there exists a class imbalance in some age ranges, races, and gender. To solve this problem, we increased the training and validation data by data augmentation that included translation, cropping, and horizontal flipping for classes whose numbers of images were much smaller than those of other classes (e.g., the images of ages from 56~65 and 66~77). From that, we could make all the images of each class distributed equally by gender, age, and race for the training and validation. In our experiment, the convergence of the validation loss was slightly delayed, as shown in Figure 9, but the result was well learned without overfitting the training data.

The reason why the validation loss of the discriminator increased after 10 epochs is as follows. The Adam optimizer used in this experiment has various advantages, but it has the disadvantages of poor conditioning problems and slow initial learning speed that depends on the size of the database, the value of the hyperparameter, and the loss function used [74]. In Figure 9, the slightly slower convergence speed of the generator loss is likely due to the initially slow learning speed of the Adam optimizer, while the relatively fast convergence of the discriminator loss is due to the fact that it is relatively easier to learn than the generator [76]. However, the brief increase in discriminator loss is a result of the poor conditioning problem mentioned above. The first and second moments used by the Adam optimizer are the mean of the sample mean and sample square of the input data, respectively. In this experiment, mean squared error (MSE) is used as the content loss function, which causes a poor conditioning problem in the second moment during back-propagation. Several studies have proposed methods to solve this problem, and in our paper, we applied L_2 regularization [78], used the largest possible batch size in the experimental environment [79], used a learning rate of 0.0002, which is smaller than the 0.001 usually used for the Adam optimizer [80], and applied a weight decay every 10 epochs [78]. Therefore, the discriminator loss in Figure 9 increases slightly after 10 epochs, decreases again after 50 epochs, and gradually converges, which can be seen as a good response to the problem of the Adam optimizer in our experiment.

Subsequently, the images de-occluded using LCA-GAN were learned by DEX [8], an age estimation CNN model. The same random cropping outlined above was applied through online augmentation, and learning was conducted for 200 epochs using the Adam optimizer, with a learning rate of 0.0002, beta_1 of 0.5, and beta_2 of 0.999 [74]. Figure 10 illustrates the training loss and accuracy graphs of DEX as well as the validation loss and accuracy graphs of DEX. The convergence of the training loss and accuracy graphs demonstrates that the DEX age estimator was sufficiently trained on the de-occluded training data generated by LCA-GAN. Moreover, the convergence of the validation loss and accuracy graphs demonstrates that the DEX age estimator was not overfitted to the de-occluded training data generated by LCA-GAN.

Table 5. Distribution of our experimental database for two-fold cross validation.

	Gender	Age	White	Black	Hispanic	Asian	Other	Total
Training	Male	16~25	952	5834	401	44	4	7235
		26~35	864	4184	231	13	5	5296
		36~45	1090	4300	92	3	4	5490
		46~55	579	1973	24	4	7	2588
		56~65	90	268	2	0	0	360
		66~77	7	16	0	0	0	23
		Total	3582	16,574	750	63	20	20,990
	Female	16~25	283	710	20	5	0	1017
		26~35	367	761	19	0	2	1149
		36~45	395	818	6	0	5	1224
		46~55	106	275	0	0	1	383
		56~65	18	25	0	0	1	45
		66~77	1	1	0	0	0	2
		Total	1169	2591	46	6	9	3820
Validation	Male	16~25	212	1296	89	10	1	1608
		26~35	192	930	51	3	1	1177
		36~45	242	956	20	1	1	1220
		46~55	129	439	5	1	2	575
		56~65	20	60	0	0	0	80
		66~77	2	4	0	0	0	5
		Total	796	3683	167	14	4	4665
	Female	16~25	63	158	4	1	0	226
		26~35	82	169	4	0	1	255
		36~45	88	182	1	0	1	272
		46~55	24	61	0	0	0	85
		56~65	4	6	0	0	0	10
		66~77	0	0	0	0	0	1
		Total	260	576	10	1	2	849
Total	Male		7961	36,832	1667	141	44	46,645
	Female		2598	5757	102	13	19	8489

4.3. Testing with MORPH Database

4.3.1. Comparisons of the Quality of Images Generated by Proposed Method and State-of-the-Art Methods

To compare the performance of the de-occlusion model for mask-occluded facial images in this experiment with other models, the structural similarity index measure (SSIM) [81] and peak signal-to-noise ratio (PSNR) [82] were used to measure the similarity between the original image and the generated de-occluded image. SSIM is expressed in Equation (9), and PSNR is expressed in Equation (10). Larger values of both SSIM and PSNR indicate better performance of the de-occlusion model.

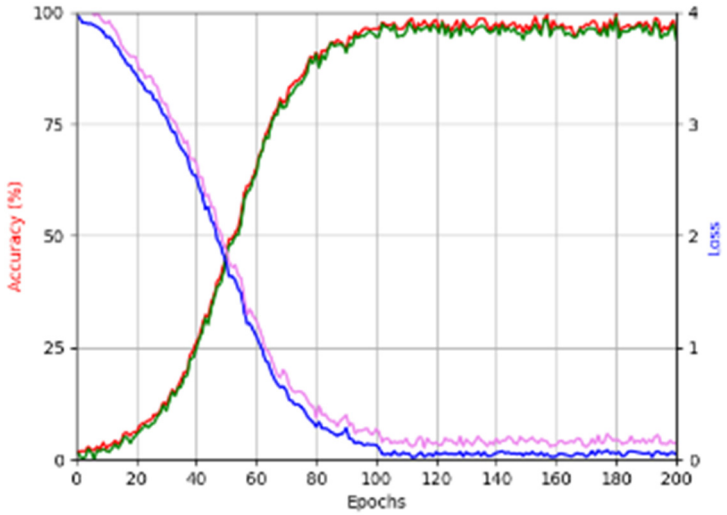


Figure 10. Graphs of DEX loss and accuracy. Blue and red lines represent the training loss and accuracy graphs, respectively. Violet and green lines represent the validation loss and accuracy graphs, respectively.

$$\text{MSE} = \frac{1}{WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}[I_o(i,\ j) - I_d(i,\ j)]^2 \tag{8}$$

$$\text{SSIM} = \frac{(2\mu_d\mu_o + C1)(2\sigma_{do} + C2)}{(\mu_d{}^2 + \mu_o{}^2 + C1)(\sigma_d{}^2 + \sigma_o{}^2 + C2)} \tag{9}$$

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

I_o represents the original image, and I_d represents the mask de-occluded image. Moreover, W and H denote the width and height of the image, respectively. μ_o and σ_o indicate the mean and standard deviation of the pixel values of the original image, respectively. μ_d and σ_d indicate the mean and standard deviation of the pixel values of a mask de-occluded image, respectively, and μ_{do} denotes the two images' covariance. $C1$ and $C2$ correspond to the positive constant offsets.

As presented in Table 6, Pix2pix [57] and MPRNet [83] yielded the best performance for SSIM and PSNR according to the de-occlusion results, while the proposed LCA-GAN exhibit the fourth and third highest performance for SSIM and PSNR, respectively. Nevertheless, SSIM and PSNR are values that represent the image quality according to de-occlusion; the primary goal of this study is to improve age estimation accuracy (not image quality) through de-occlusion. As presented in Sections 4.3.2 and 4.4.2, the proposed LCA-GAN yielded the highest age estimation accuracy.

Table 6. Comparative SSIM and PSNR of original image and de-occluded images.

	LCA-GAN	**AFD-StackGAN [54]**	**CFR-GAN [55]**	**MPRNet [83]**	**CycleGAN [84]**	**Pix2pix [57]**
SSIM	0.6962	0.6769	0.7107	0.7031	0.5630	0.7225
PSNR (unit: dB)	19.0302	16.3121	18.3067	19.6427	19.3321	18.8731

4.3.2. Comparisons of Age Estimation Accuracy

Ablation Studies

As shown in Equation (11), we used the mean absolute error (MAE), the most frequently applied metric [85,86], to evaluate age estimation accuracy. A lower MAE value shows better performance of age estimation.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|p_i - y_i| \tag{11}$$

In the above equation, n denotes the number of images, p_i denotes the predicted age, and y_i denotes the ground-truth age. For the first ablation study, we compared the age estimation accuracy of the de-occluded images depending on the application of LCCA and LCSA, which form LCAB, and edge and content losses in the proposed LCA-GAN. According to the results in Table 7, the proposed LCA-GAN yields the highest age estimation performance in the generated de-occluded images when using LCCA, LCSA, and edge and content losses, respectively.

Table 7. Ablation study with or without LCSA, LCCA, and Edge and Content losses in our LCA-GAN (unit: years).

LCSA	LCCA	Edge Loss + Content Loss	MAE
×	×	×	7.72
○	×	×	7.11
×	○	×	7.09
×	×	○	7.83
○	○	×	6.82
○	○	○	6.64

We performed an experiment to conduct an additional ablation study, as shown in Table 8. We compare the age estimation accuracy when applying various backbone models as the LCA-GAN generator. The last row in Table 8 is the method that subtracts the original image and masked image, concatenates the input image with the image where only the occluded area remains, and uses this input image in Pix2pix learning. Evidently, the best age estimation performance was achieved when using the Pix2pix backbone generator in LCA-GAN. In addition, we performed additional comparisons of the accuracy of the state-of-the-art age estimation method in Table 8. As shown in Table 8, the MAEs with original non-occluded and mask-occluded face images by the state-of-the-art age estimation method were 5.80 years (baseline 1) and 10.45 years (baseline 2) years, respectively. Although the MAE by our LCA-GAN is 6.64 years, it is much lower than that with mask-occluded face images without our LCA-GAN (baseline 2), which confirms the effectiveness of our proposed LCA-GAN.

Figure 11 shows examples of the generated images according to the ablation study in Table 8. In Table 8, the age estimation performance of baseline 1 using non-occluded facial images and baseline 2 using mask-occluded facial images (occluding nose and mouth) were 5.80 and 10.45, respectively, which is a difference of 4.65 years, indicating that the nose and mouth have important information for age estimation. In addition, the areas in the facial image that have important information for age estimation are shown, with significant activation in the nose and mouth. Figure 11a,b show the masked images and original images, and Figure 11c–f display the de-occluded images processed in the order in Table 9. According to the results in Figure 11, the proposed LCA-GAN using only Pix2pix yielded the best de-occlusion performance. The de-occlusion networks including Pix2pix and CycleGAN used in Table 8 and Figure 11, except for U-net, are adversarial networks using U-net as a generator. These adversarial learning-based de-occlusion methods generate

robust and realistic de-occluded facial images by capturing complex patterns, but age information is somewhat lost due to unnecessary deformation in non-occluded areas of the eyes. On the other hand, CNN-based U-net is trained with a pixel-wise loss function and has less deformation in the non-occluded facial area of the eyes, so it preserves age information well. However, learning the mapping from input image to target image in the mask-occluded area is difficult and generates blurred images. Consequently, adversarial learning-based methods are weak at preserving information in non-occluded areas of the eyes and are strong at de-occlusion, while U-net is strong at preserving information in non-occluded areas of the eyes and is weak at de-occlusion. The areas with significant age information are shown in the human face image. The areas with high activation are the eyes, nose, and mouth, where the eyes are a non-occluded area and the nose and mouth are occluded areas. As a result, U-net preserved the age information in the non-occluded area of the eyes well, but the non-occluded areas were smaller than the whole face area, and the consequent de-occlusion performance was lower than the other methods because it cannot restore the age information.

Table 8. Ablation study according to various backbone generators in LCA-GAN. Baselines 1 and 2 show the MAEs with original non-occluded and mask-occluded face images by DEX, respectively. Baseline 2 shows the MAE with original non-occluded face images by DEX. (Pix2pix* indicates the method that subtracts the original image and masked image, concatenates the input image with the image where only the occluded area remains, and uses this input image in Pix2pix learning) (unit: years).

Method	MAE
Baseline 1	5.80
Baseline 2	10.45
U-net	7.70
Pix2pix (LCA-GAN)	6.64
CycleGAN	7.15
Pix2pix*	6.91

Table 9. Comparative accuracies of age estimation by LCA-GAN and various de-occlusion methods (MPRNet* is a two-stage model that reduces one stage using the input image with the smallest size in MPRNet) (unit: years).

	LCA-GAN	AFD-StackGAN [54]	CFR-GAN [55]	MPRNet [83]	MPRNet* [83]	Pix2pix [57]	CycleGAN [84]
MAE	6.64	6.92	7.13	6.95	7.83	7.72	8.18

(**a**)

Figure 11. *Cont.*

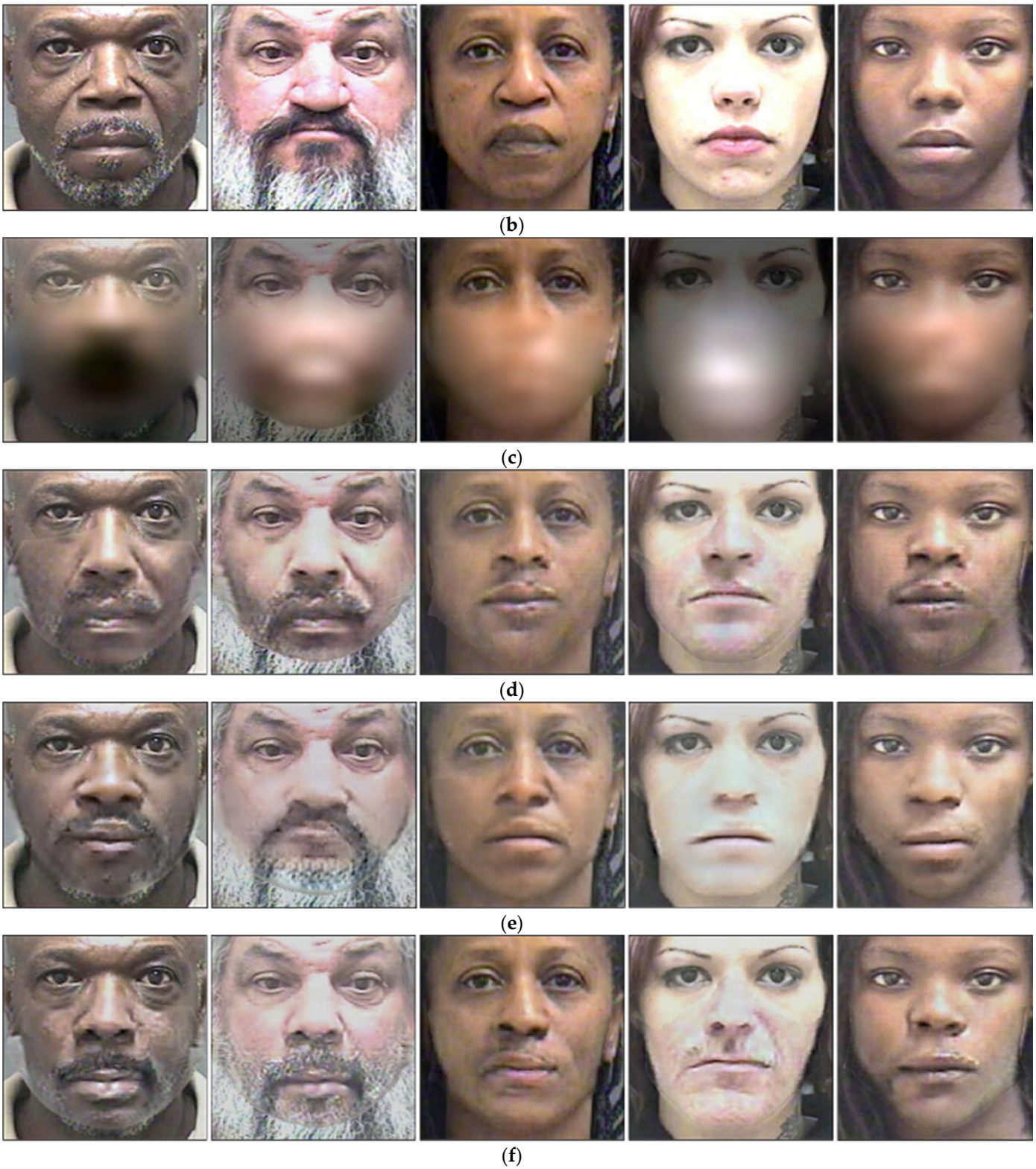


Figure 11. Examples of de-occluded mask images. (**a**) Masked images and (**b**) original images. De-occluded images are shown using (**c**) U-net structure, (**d**) CycleGAN structure, (**e**) Pix2pix structure (LCA-GAN), and (**f**) Pix2pix* (a method that subtracts the original image and masked image, concatenates the input image with the image where only the occluded area remains, and uses this input image in Pix2pix learning).

Comparisons of Our LCA-GAN with Existing Methods

This subsection compares the proposed method with state-of-the-art methods. MPRNet is a three-stage model with an iterative structure that receives input images through the multi-scale approach [83]. MPRNet* in Table 9 is a two-stage model that reduces one stage using the input image with the smallest size in MPRNet. According to the experimental results listed in Table 9, LCA-GAN, the de-occlusion network proposed in this study, yielded the best age estimation performance.

Figure 12 shows examples of mask de-occluded images obtained using the proposed LCA-GAN and state-of-the-art methods. Figure 12a displays masked facial images created by the method described in Section 4.1, and Figure 12b shows the original facial images. De-occluded images are shown by (c) the proposed LCA-GAN, (d) AFD-Stack GAN, (e) CFR-GAN, (f) MRPNet, (g) CycleGAN, and (h) Pix2pix. As shown in Figure 12, the mask de-occluded image generated by LCA-GAN is the nearest to the original image.

(**a**)

(**b**)

(**c**)

Figure 12. *Cont.*

(**d**)

(**e**)

(**f**)

(**g**)

(**h**)

Figure 12. Examples of mask de-occluded images. (**a**) Masked images and (**b**) original images without a mask. De-occluded images are shown by using (**c**) the proposed LCA-GAN, (**d**) AFD-Stack GAN, (**e**) CFR-GAN, (**f**) MRPNet, (**g**) CycleGAN, and (**h**) Pix2pix.

4.4. Testing with PAL Database

4.4.1. Comparisons of the Quality of Images Generated by Proposed Method and the State-of-the-Art Methods

We performed additional experiments using the open database PAL to confirm the generality of the proposed LCA-GAN performance. As presented in Table 10, CFR-GAN [55] and AFD-StackGAN [54] exhibited the best performance for SSIM and PSNR, respectively, while the proposed LCA-GAN yielded the third and fourth highest performance for SSIM and PSNR, respectively. However, SSIM and PSNR are values that represent image quality according to de-occlusion, but the primary goal of this study is to improve age estimation accuracy (not image quality) through de-occlusion. According to a comparison of age estimation accuracy in Table 11, the proposed LCA-GAN yielded the highest accuracy.

Table 10. Comparative SSIM and PSNR of original image and de-occluded images.

	LCA-GAN	**AFD-StackGAN [54]**	**CFR-GAN [55]**	**MPRNet [83]**	**Pix2pix [57]**	**CycleGAN [84]**
SSIM	0.7042	0.6983	0.7207	0.7002	0.7134	0.6892
PSNR	18.3302	19.4423	18.3043	18.9742	19.4211	17.9443

Table 11. Comparative accuracies of age estimation by LCA-GAN and various de-occlusion methods (MPRNet* is a two-stage model that reduces one stage using the input image with the smallest size in MPRNet) (unit: years).

	LCA-GAN	**AFD-StackGAN [54]**	**CFR-GAN [55]**	**MPRNet [83]**	**MPRNet* [83]**	**Pix2pix [57]**	**CycleGAN [84]**
MAE	6.12	6.94	6.52	8.21	8.70	7.12	9.02

4.4.2. Comparisons of Age Estimation Accuracy by Our LCA-GAN and the Existing Methods

For the next experiment, we de-occluded images using LCA-GAN and compared the age estimation accuracy using DEX. MPRNet* in Table 9 is a two-stage model that reduces one stage using the input image with the smallest size in MPRNet. As shown in Table 11, LCA-GAN, the de-occlusion network proposed in this study, yielded the best age estimation performance. In addition, we performed additional experiments using different age estimation methods after the same use of LCA-GAN. As shown in Table 12, DEX showed the best accuracy among all the different age estimation methods.

Table 12. Comparative accuracies of different age estimation methods after the same use of LCA-GAN (unit: years).

	VGG-16 [68]	**ResNet-50 [87]**	**ResNet-152 [87]**	**DEX [8]**	**AgeNet [11,88]**	**Inception with Random Forest [20]**
MAE	6.20	7.22	6.32	6.12	6.19	6.42

Figure 13 illustrates examples of mask-de-occluded images obtained by the proposed LCA-GAN and the state-of-the-art methods. As evidenced in Figure 13, the mask de-occluded image generated by LCA-GAN is the closest to the original image.

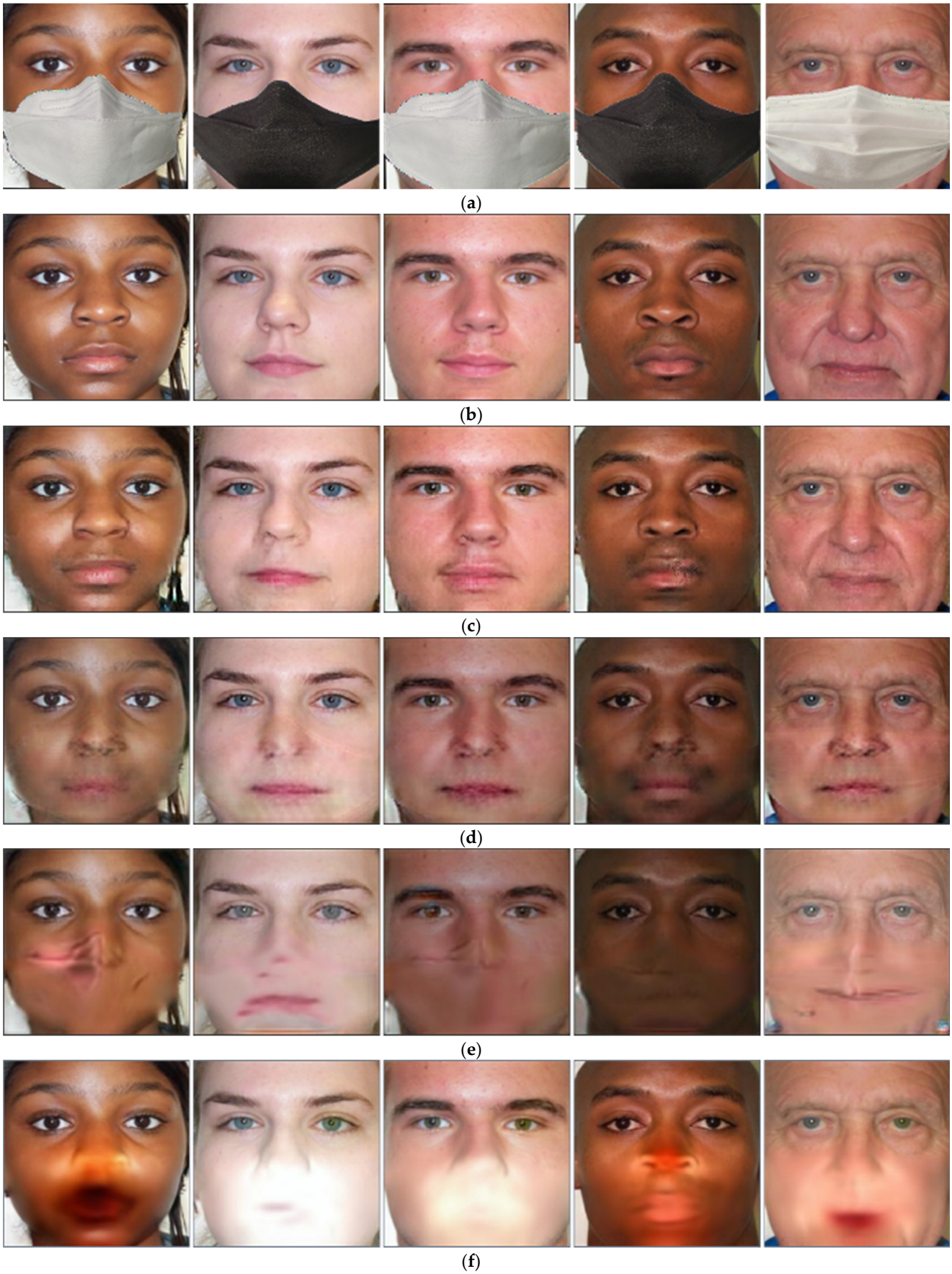


Figure 13. *Cont.*

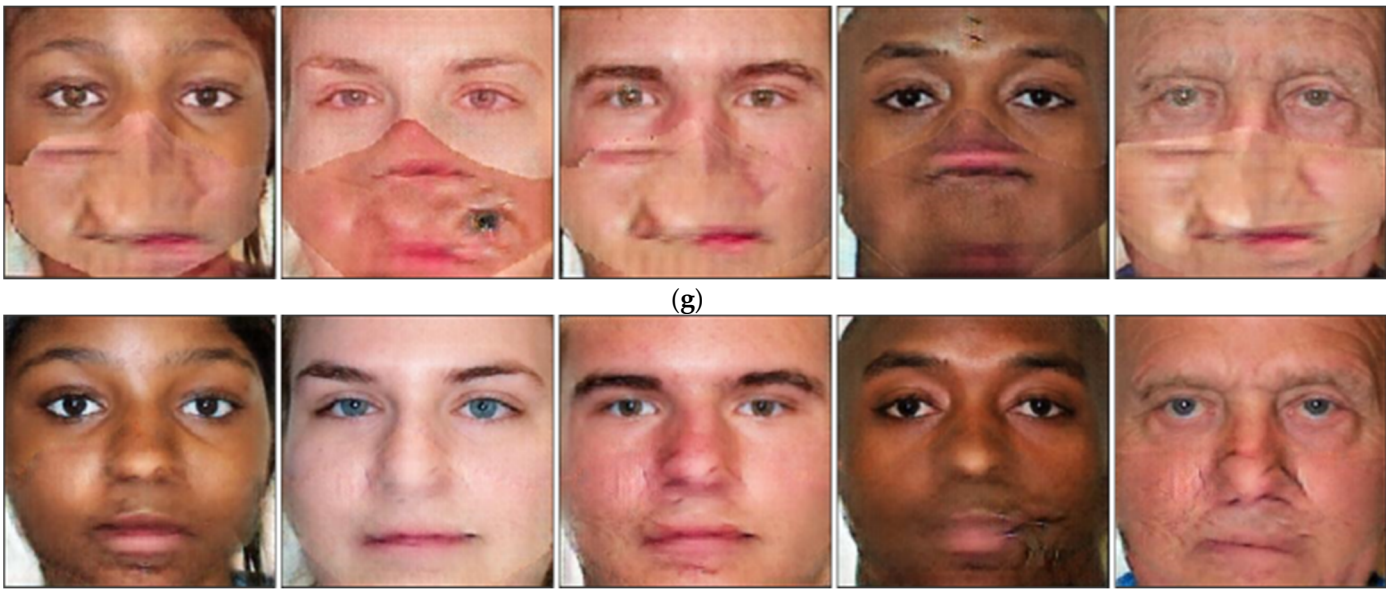


Figure 13. Examples of mask-de-occluded images. (**a**) Masked images and (**b**) original images without a mask. De-occluded images are shown by (**c**) the proposed LCA-GAN, (**d**) AFD-Stack GAN, (**e**) CFR-GAN, (**f**) MRPNet, (**g**) CycleGAN, and (**h**) Pix2pix.

4.5. Processing Speed

In this subsection, we measured and compared the processing times of the proposed LCA-GAN and state-of-the-art methods in the desktop environment described in Section 4.1 and a Jetson TX2 board [89], as shown in Figure 14. Table 13 lists the measured processing times. LCA-GAN yielded a faster processing speed in the desktop environment and embedded environment than all state-of-the-art methods, except Pix2pix [57]. Furthermore, the processing speed did not greatly differ from Pix2pix [57]. This indicates that the proposed LCA-GAN can be operated even in an embedded system environment with limited computing resources. Table 14 compares the number of parameters, giga floating point operations per second (GFLOPs), and memory usage between the proposed LCA-GAN and state-of-the-art methods. LCA-GAN exhibited the smallest number of parameters, second lowest GFLOPs, and third lowest memory usage compared to the state-of-the-art methods. However, as indicated in Tables 9 and 11, the proposed LCA-GAN yielded the best age estimation performance compared to the previous methods.

Table 13. Comparative average processing time of one image by LCA-GAN and state-of-the-art methods (unit: ms).

	Desktop Computer	**Jetson TX2 Board**
LCA-GAN	11.94	177.52
AFD-Stack GAN [54]	23.03	342.8
CFR-GAN [55]	35.6	541.5
MPRNet [83]	21.12	318.02
MPRNet* [83]	14.04	212.24
Pix2pix [57]	11.2	171.5
Cycle GAN [84]	23.2	353.1

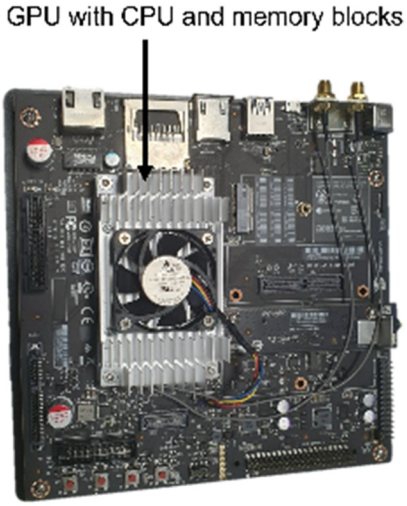


Figure 14. Jetson TX2 board.

Table 14. Comparative model complexities of LCA-GAN and state-of-the-art methods.

	Number of Parameters	GFLOPs	Memory Usage (GB)
LCA-GAN	57,118,684	1.4668	0.5913
AFD-Stack GAN [54]	102,325,149	2.6764	0.5961
CFR-GAN [55]	171,588,876	2.8730	1.0925
MPRNet [83]	102,725,856	9.1092	4.5598
MPRNet* [83]	68,114,344	6.0728	3.0399
Pix2pix [57]	57,196,292	0.972	0.2062
Cycle GAN [84]	114,392,584	1.944	0.4124

4.6. Discussion

In this subsection, we present the extraction and analysis of the attention map of the attention module used in the LCA-GAN de-occlusion process (Figure 15) and the gradient class activation map (Grad-CAM) [90] of DEX used for age estimation (Figure 16). Figure 15 displays the (a) original facial images and (b) the mask-occluded facial images. Figure 15c–e show the attention maps of LCAB 1, LCAB 3, and LCAB 5 of the Table 3 encoder, respectively, and Figure 15f–h show the attention maps of LCAB 6, LCAB 8, and LCAB 10 of the Table 3 decoder, respectively. The attention maps show that in the proposed LCA-GAN, as the encoder de-occlusion progresses, attention is activated from the entire face area to the mask area. Subsequently, as the decoder de-occlusion progresses, high activation is shown in the detailed areas of major facial elements, such as the eyes, nose, mouth, and chin.

Subsequently, we examined the Grad-CAM images of DEX, the age estimation network used in this experiment. Figure 16 shows (a) the original images, (b) the mask-occluded images, and (c) the de-occluded images. In Figure 15d–g, the Grad-CAM images of DEX's 4th, 8th, and 11th convolution layers and the last max pooling layers are overlapped with the mask de-occluded images. As illustrated in Figure 16, as DEX learns to estimate the age from mask de-occluded facial images, in Grad-CAM, which showed high activation for high-frequency information in large areas of the image, elements such as the eyes, nose, mouth, and the surrounding textures are activated with the deepening of layers.

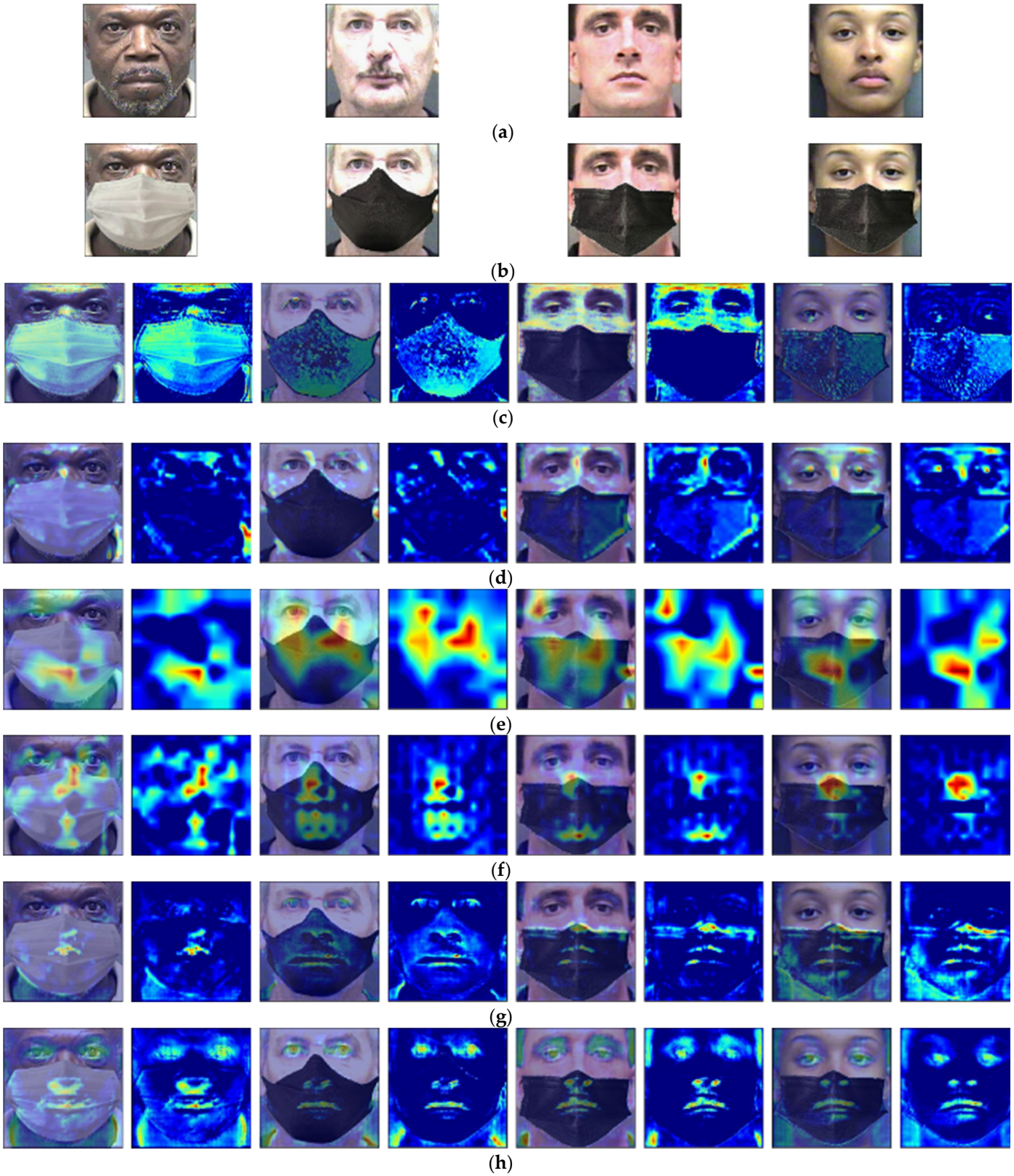


Figure 15. LCA-GAN attention map. (**a**) Original image, (**b**) mask-occluded image, (**c**–**e**) attention maps from Table 3 encoder LCAB 1, LCAB 3, and LCAB 5, and (**f**–**h**) attention maps from Table 3 decoder LCAB 6, LCAB 8, and LCAB 10.

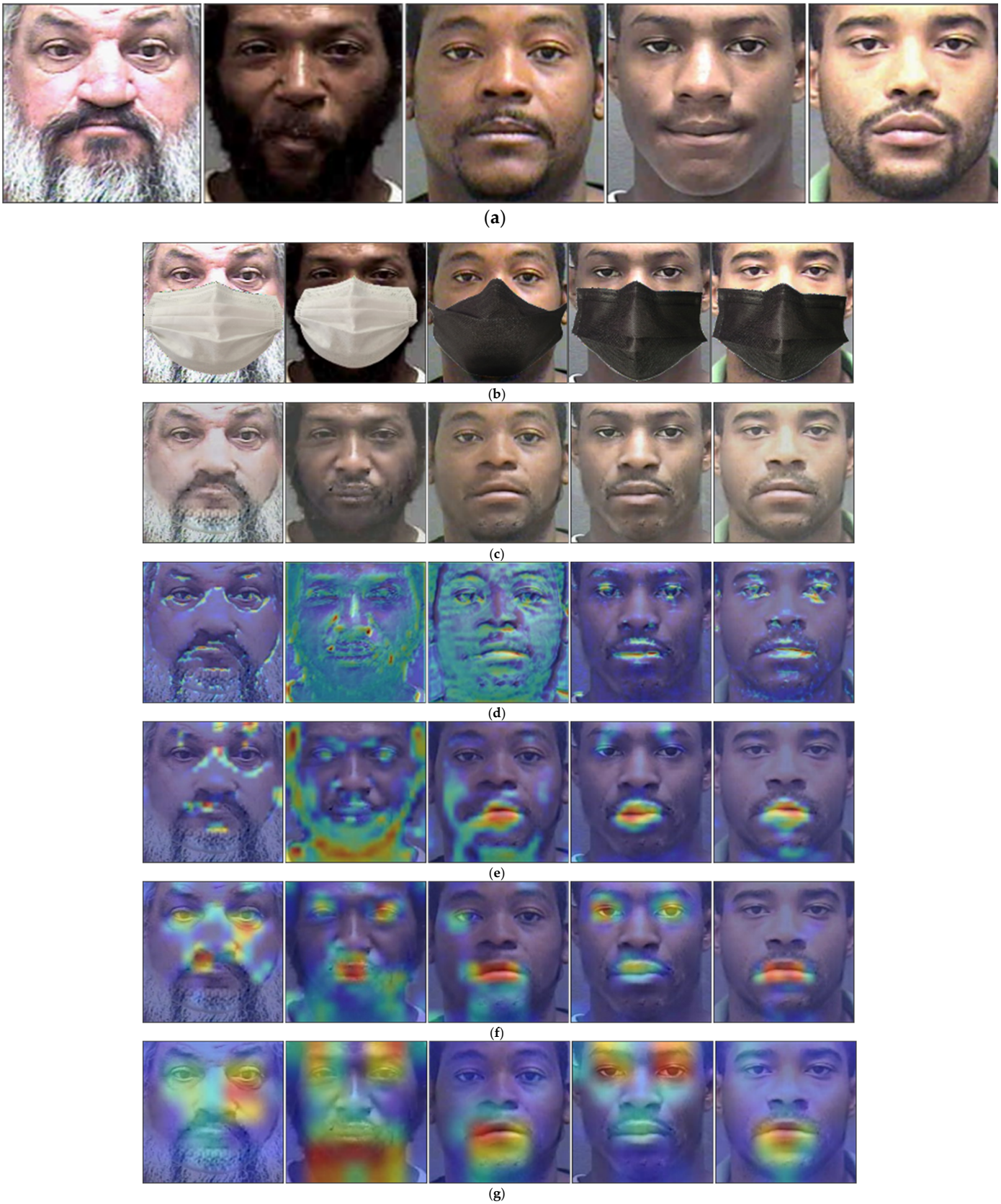
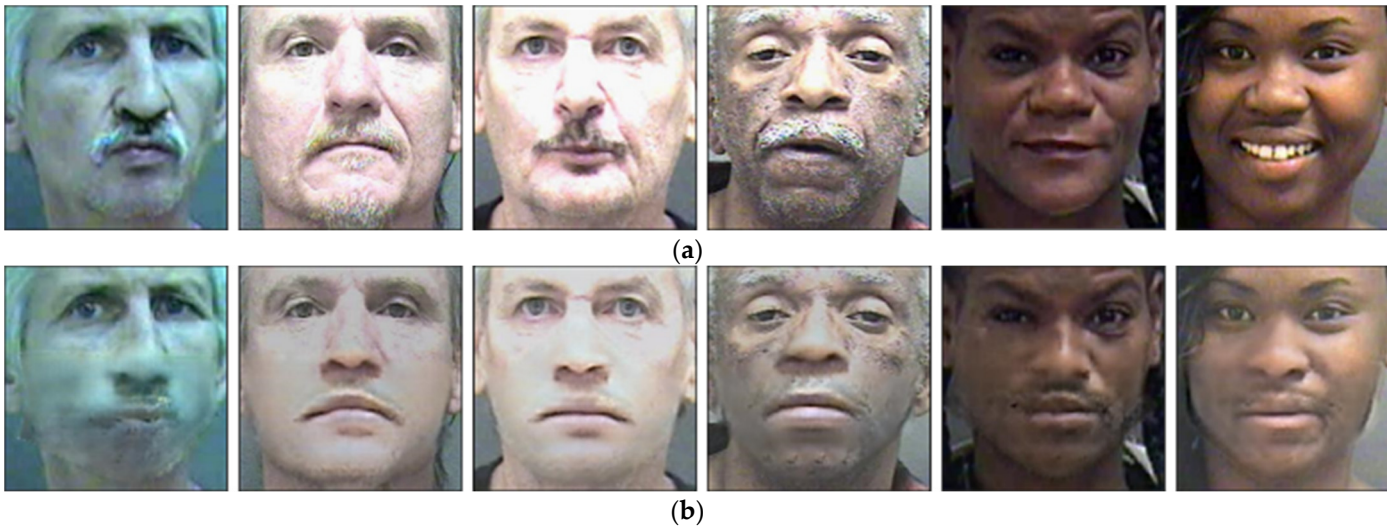


Figure 16. Grad-CAM images from DEX with LCA-GAN. (**a**) Original image, (**b**) mask-occluded images, (**c**) de-occluded image, (**d**–**g**) overlapped Grad-CAM images extracted from the 4th, 8th, 11th convolution layers as well as the last pooling layers of DEX, respectively.

Figure 17 shows examples of de-occluded images that were incorrectly generated by the proposed LCA-GAN. These incorrectly generated de-occluded images can be attributed to several problems: first, the difference in convergence speed between the generator and discriminator during adversarial learning; second, the use of a single generator; and finally, the class imbalance according to gender and race in the learning images as well as non-uniform lighting in the test images.



(**a**)

(**b**)

Figure 17. Bad case images generated by LCA-GAN. (**a**) Original images and (**b**) de-occluded images by LCA-GAN.

5. Conclusions

Mask-occluded images that occur in real environments cause a loss of information required for age estimation, thereby degrading age estimation performance. This study proposed a novel de-occlusion network LCA-GAN. Through experiments using MORPH and PAL, open databases of human facial images, the proposed network achieved higher age estimation performance than existing state-of-the-art de-occlusion networks. Furthermore, the proposed LCA-GAN contains 57,118,684 parameters, which is fewer than existing methods. This indicates that it can be operated even in an embedded system with limited computing resources. Moreover, from the attention maps in LCA-GAN and Grad-CAM images of DEX for images de-occluded with LCA-GAN, LCA-GAN and DEX effectively extracted features for de-occlusion and age estimation, respectively. However, as shown in Figure 17, LCA-GAN occasionally incorrectly generated de-occluded images.

To solve this, it is necessary to research solutions for several problems: the difference in convergence speed between the generator and discriminator during adversarial learning, the use of a single generator, the class imbalance according to gender and race in the learning images, and non-uniform lighting in the test images. Moreover, we will investigate solutions for cases where mask occlusion simultaneously occurs with other factors, such as low light and image blurring. Furthermore, we will research a shallower model to achieve faster processing speeds in an embedded platform.

Author Contributions: Methodology, S.H.N.; supervision, K.R.P.; validation, Y.H.K., J.C. and C.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-2022R1F1A1064291), in part by the MSIT, Korea, under the Information Technology Research Center (ITRC) support program (IITP-2023-2020-0-01789) supervised by the IITP (Insti-

tute for Information and Communications Technology Planning and Evaluation), and in part by the National Supercomputing Center with supercomputing resources including technical support (TS-2023-RE-0025).

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. Gallagher, A.C.; Chen, T. Estimating age, gender, and identity using first name priors. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
2. Angulu, R.; Tapamo, J.R.; Adewumi, A.O. Age estimation via face images: A survey. *EURASIP J. Image Video Process.* **2018**, *2018*, 42. [CrossRef]
3. 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.
4. Farkas, J.P.; Pessa, J.E.; Hubbard, B.; Rohrich, R.J. The Science and Theory behind Facial Aging. *Plast. Reconstr. Surg. Glob. Open* **2013**, *1*, e8–e15. [CrossRef]
5. Albert, A.M.; Ricanek, K., Jr.; Patterson, E. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Sci. Int.* **2007**, *172*, 1–9. [CrossRef]
6. Olatunbosun, A.-A.; Serestina, V. Deep learning approach for facial age classification: A survey of the state-of-the-art. *Artif. Intell. Rev.* **2020**, *54*, 179–213.
7. Antipov, G.; Baccouche, M.; Berrani, S.-A.; Dugelay, J.-L. Apparent age estimation from face images combining general and children-specialized deep learning models. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 801–809.
8. Rothe, R.; Timofte, R.; Van Gool, L. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [CrossRef]
9. Agbo-Ajala, O.; Viriri, S. Face-based age and gender classification using deep learning model. In Proceedings of the 2019 International Workshops, Sydney, NSW, Australia, 18–22 November 2019; pp. 125–137.
10. Duan, M.; Li, K.; Li, K. An ensemble CNN2ELM for age estimation. *IEEE Trans. Inf. Forensic Secur.* **2018**, *13*, 758–772. [CrossRef]
11. Liao, H.; Yan, Y.; Dai, W.; Fan, P. Age Estimation of Face Images Based on CNN and Divide-and-Rule Strategy. *Math. Probl. Eng.* **2018**, *2018*, 1712686. [CrossRef]
12. LCA-GAN with Algorithm. (Model and Algorithm to Be Uploaded on Github). Available online: https://github.com/nsh6473/LCA-GAN/ (accessed on 1 February 2023).
13. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77–91.
14. Hiba, S.; Keller, Y. Hierarchical attention-based age estimation and Bias estimation. *arXiv* **2021**, arXiv:2103.09882.
15. Nimhed, C. Estimation of Height, Weight, Sex and Age from Magnetic Resonance Images Using 3D Convolutional Neural Networks. Master's Thesis, Linköping University, Linköping, Sweden, 2022; pp. 1–60.
16. Yaman, D.; Eyiokur, F.I.; Ekenel, H.K. Multimodal soft biometrics: Combining ear and face biometrics for age and gender classification. *Multimedia Tools Appl.* **2021**, *81*, 22695–22713. [CrossRef]
17. Onifade, O.F.W.; Akinyemi, J.D. A GW ranking approach for facial age estimation. *Egypt. Comput. Sci. J.* **2014**, *38*, 63–74.
18. 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.
19. 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 IEEE 8th International Conference on Biometrics Theory, Applications and Systems, Niagara Falls, NY, USA, 6–9 September 2016; pp. 1–8.
20. 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.
21. 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.
22. Agustsson, E.; Timofte, R.; Escalera, S.; Baro, X.; Guyon, I.; Rothe, R. Apparent and real age estimation in still images with deep residual regressors on appa-real database. In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 87–94.
23. Anand, A.; Labati, R.D.; Genovese, A.; Munoz, E.; Piuri, V.; Scotti, F. Age estimation based on face images and pre-trained convolutional neural networks. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence, Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–7.

24. Aydogdu, M.F.; Demirci, M.F. Age classification using an optimized CNN architecture. In Proceedings of the International Conference on Compute and Data Analysis, Lakeland, FL, USA, 19–23 May 2017; pp. 233–239.
25. Zhang, K.; Gao, C.; Guo, L.; Sun, M.; Yuan, X.; Han, T.X.; Zhao, Z.; Li, B. Age Group and Gender Estimation in the Wild With Deep RoR Architecture. *IEEE Access* **2017**, *5*, 22492–22503. [CrossRef]
26. Ranjan, R.; Zhou, S.; Chen, J.C.; Kumar, A.; Alavi, A.; Patel, V.M.; Chellappa, R. Unconstrained age estimation with deep convolutional neural networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop, Santiago, Chile, 7–13 December 2015; pp. 351–359.
27. 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.
28. Li, W.; Lu, J.; Feng, J.; Xu, C.; Zhou, J.; Tian, Q. Bridgenet: A continuity-aware probabilistic network for age estimation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1145–1154.
29. Gao, B.B.; Zhou, H.Y.; Wu, J.; Geng, X. Age estimation using expectation of label distribution learning. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 712–718.
30. Zhang, K.; Liu, N.; Yuan, X.; Guo, X.; Gao, C.; Zhao, Z.; Ma, Z. Fine-Grained Age Estimation in the Wild With Attention LSTM Networks. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3140–3152. [CrossRef]
31. Chen, S.; Zhang, C.; Dong, M.; Le, J.; Rao, M. Using ranking-CNN for age estimation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 742–751.
32. Liu, W.; Chen, L.; Chen, Y. Age Classification Using Convolutional Neural Networks with the Multi-class Focal Loss. *IOP Conf. Series: Mater. Sci. Eng.* **2018**, *428*, 012043. [CrossRef]
33. Liu, H.; Lu, J.; Feng, J.; Zhou, J. Ordinal Deep Learning for Facial Age Estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *29*, 486–501. [CrossRef]
34. Gurpinar, F.; Kaya, H.; Dibeklioglu, H.; Salah, A.A. Kernel ELM and CNN based facial age estimation. In Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 785–791.
35. Liu, K.-H.; Yan, S.; Kuo, C.-C.J. Age Estimation via Grouping and Decision Fusion. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2408–2423. [CrossRef]
36. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11–26. [CrossRef]
37. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNNELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [CrossRef]
38. Liu, X.; Zou, Y.; Kuang, H.; Ma, X. Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network. *Symmetry* **2020**, *12*, 146. [CrossRef]
39. MORPH Database. Available online: https://ebill.uncw.edu/C20231_ustores/web/store_main.jsp?STOREID=4 (accessed on 17 May 2022).
40. FGNET Database. Available online: https://yanweifu.github.io/FG_NET_data/index.html (accessed on 17 May 2022).
41. IMDB Database. Available online: https://www.imdb.com/interfaces/ (accessed on 17 May 2022).
42. Adience Database. Available online: https://talhassner.github.io/home/projects/Adience/Adience-data.html/ (accessed on 17 May 2022).
43. LAP 2015 Database. Available online: https://chalearnlap.cvc.uab.cat/dataset/18/description/ (accessed on 17 May 2022).
44. LAP 2016 Database. Available online: https://chalearnlap.cvc.uab.cat/dataset/19/description/ (accessed on 17 May 2022).
45. CACD Database. Available online: https://bcsiriuschen.github.io/CARC/ (accessed on 17 May 2022).
46. Guo, G.; Mu, G. Human age estimation: What is the influence across race and gender. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 71–78.
47. Chen, K.; Gong, S.; Xiang, T.; Change Loy, C. Cumulative attribute space for age and crowd density estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2467–2474.
48. Szegedy, C.; Loffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv* **2016**, arXiv:1602.07261v2. [CrossRef]
49. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
50. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131. [CrossRef]
51. Looking at People CVPR Challenge—Track1: Age Estimation. Available online: https://chalearnlap.cvc.uab.cat/challenge/13/description/ (accessed on 17 May 2022).
52. 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, 7–12 June 2015; pp. 1–9.
53. Dong, J.; Zhang, L.; Zhang, H.; Liu, W. Occlusion-aware GAN for face de-occlusion in the wild. In Proceedings of the IEEE International Conference on Multimedia and Expo, London, UK, 6–10 July 2020; pp. 1–6.

54. Jabbar, A.; Li, X.; Assam, M.; Khan, J.A.; Obayya, M.; Alkhonaini, M.A.; Al-Wesabi, F.N.; Assad, M. AFD-StackGAN: Automatic nask generation network for face de-occlusion using StackGAN. *Sensors* **2022**, *22*, 1747. [CrossRef]
55. Ju, Y.-J.; Lee, G.-H.; Hong, J.-H.; Lee, S.-W. Complete Face Recovery GAN: Unsupervised joint face rotation and de-occlusion from a single-view image. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 3711–3721.
56. Zhao, F.; Feng, J.; Zhao, J.; Yang, W.; Yan, S. Robust LSTM-Autoencoders for Face De-Occlusion in the Wild. *IEEE Trans. Image Process.* **2017**, *27*, 778–790. [CrossRef] [PubMed]
57. 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.
58. Iliadis, M.; Wang, H.; Molina, R.; Katsaggelos, A.K. Robust and Low-Rank Representation for Fast Face Identification With Occlusions. *IEEE Trans. Image Process.* **2017**, *26*, 2203–2218. [CrossRef] [PubMed]
59. Din, N.U.; Javed, K.; Bae, S.; Yi, J. A Novel GAN-Based Network for Unmasking of Masked Face. *IEEE Access* **2020**, *8*, 44276–44287. [CrossRef]
60. Khan, M.K.J.; Din, N.U.; Bae, S.; Yi, J. Interactive Removal of Microphone Object in Facial Images. *Electronics* **2019**, *8*, 1115. [CrossRef]
61. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
62. Dlib C++ Library. Available online: http://dlib.net/ (accessed on 17 May 2022).
63. 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.
64. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
65. Jiang, J.; Wang, C.; Liu, X.; Ma, J. Deep Learning-based Face Super-resolution: A Survey. *ACM Comput. Surv.* **2021**, *55*, 1–36. [CrossRef]
66. Looking at People ICCV Challenge—Track1: Age Estimation. Available online: https://chalearnlap.cvc.uab.cat/challenge/12/description/ (accessed on 17 May 2022).
67. 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. [CrossRef]
68. 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.
69. PAL database. Available online: http://agingmind.utdallas.edu/download-stimuli/face-database/ (accessed on 17 May 2022).
70. Python. Available online: https://www.python.org/ (accessed on 1 October 2019).
71. OpenCV. Available online: http://opencv.org (accessed on 1 October 2022).
72. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467v2.
73. NVIDIA GeForce GTX 1070. Available online: https://www.nvidia.com/en-in/geforce/products/10series/geforce-gtx-1070/ (accessed on 21 April 2022).
74. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, Sandiego, CA, USA, 7–9 May 2015; pp. 1–15.
75. Goodfellow, I. NIPS 2016 Tutorial: Generative adversarial networks. *arXiv* **2017**, arXiv:1701.00160v4.
76. Salimans., T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X.; Chen, X. Improved techniques for training GANs. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2234–2242.
77. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017; pp. 1–17.
78. Loshchilov, I.; Hutter, F. Fixing weight decay regularization in ADAM. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–4.
79. Smith, S.L.; Kindermans, P.-J.; Ying, C.; Le, Q.V. Don't decay the learning rate increase the batch size. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–11.
80. Reddi, S.J.; Kale, S.; Kumar, S. On the convergence of Adam and beyond. *arXiv* **2019**, arXiv:1904.09237v1.
81. 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]
82. Antkowiak, J.; Baina, T.J. Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. ITU-T Standards Contribution COM, 2000. Available online: https://www.vqeg.org/publications-and-software/ (accessed on 17 May 2022).
83. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.
84. 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, 22–29 October 2017; pp. 2223–2232.

85. Sharma, N.; Sharma, R.; Jindal, N. Face-Based Age and Gender Estimation Using Improved Convolutional Neural Network Approach. *Wirel. Pers. Commun.* **2022**, *124*, 3035–3054. [CrossRef]
86. Zhang, B.; Bao, Y. Age Estimation of Faces in Videos Using Head Pose Estimation and Convolutional Neural Networks. *Sensors* **2022**, *22*, 4171. [CrossRef] [PubMed]
87. 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.
88. 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.
89. Jetson TX2 Module. Available online: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems-dev-kits-modules/ (accessed on 15 September 2022).
90. 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 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626. [CrossRef]