The Journal of Supercomputing (2023) 79:7132–7159
https://doi.org/10.1007/s11227-022-04954-x

L-GAN: landmark-based generative adversarial network for efficient face de-identification

Sung-su Jang[1] · Cheol-jin Kim[1] · Seong-yeon Hwang[1] · Myung-jae Lee[1] · Young-guk Ha[1]

Accepted: 12 November 2022 / Published online: 22 November 2022

Abstract

A large amount of high-quality data are collected through autonomous vehicles, CCTVs, guidance service robots, and web map services (Google Street View). However, the data collected through them include personal information such as peoples' faces and vehicle license plates. Currently, personal information contained in data is de-identified using methods such as face blur, pixelation, and masking. Consequently, it loses value as data for artificial intelligence (AI) learning. Therefore, in this study, we propose a model to generate a fake face that maintains the basic structure of the human face. There are several methods for generating faces. One is to generate them using a generative adversarial network (GAN) model. The GAN is an AI algorithm used for unsupervised learning and is implemented by a system in which two neural networks compete. However, because GAN operates as an input of a random noise vector, it is difficult to obtain results for the desired face angle and shape. Therefore, pre- and post-processing is required to generate a fake face that maintains the basic structure and angles; however, the calculation is complicated, and it is difficult to generate a natural image. To solve this problem, we propose a method for generating a fake face that maintains the basic structure and angle of the real face by applying a facial landmark. Using the proposed method, it was possible to generate a fake face with a different impression while maintaining the basic structure and angle of the face.

Keywords Facial landmark · GAN · Face de-identification · Face generation

✉ Young-guk Ha
ygha@konkuk.ac.kr

Extended author information available on the last page of the article

1 Introduction

Recently, innovative technologies such as autonomous vehicles and chatbots have attracted attention in the field of artificial intelligence (AI). The development of such AI offers a method for solving existing problems. Although with the machine learning method, good recognition rate has been achieved in specific situations, this approach is limited by its inability to solve diverse and complex problems. However, with the advancements in hardware, fast calculations are now possible, and deep learning technology is being rapidly developed to solve such problems. Computer vision processes images and videos [1, 2]. It began in the 1960s with the study of AI and robotics. During the research that Hubel and Wiesel conducted in the field of electrophysiology, the question they posed was 'What is the visual processing mechanism in primates and mammals?' that is, how do primates and mammals visually process images? To this end, they studied the brains of cats [3]. Electrodes were inserted into the visual cortex of a cat's brain, and the study focused on the simplest cell to determine what stimulated the neurons in the cat's brain. The results showed that even if the shape of the neurons responsible for the cat's vision was the same, the response of the cat's brain changed based on the direction of the stimulus. Neurons in an animal's visual cortex respond to outline representations in specific areas within the visual field of view. Thus, they conclude that the human brain appears to process visual information hierarchically. In 1982, Fukushima [4] created a neural network that behaved like a brain, based similarly on the theory above. This implies that research on computer vision using artificial neural networks has been conducted for decades. However, the problem of gradient loss [5] associated with information loss in the neural learning process has not yet been solved; hence, research has not yet proceeded in this area.

As mentioned above, with an increase in the depth of networks, learning is not performed because of the extinction gradient problem, which is a prominent issue. This problem can be solved in deep networks through various methods such as the advent of backpropagation learning methods and changing the activation function. Based on these ideas, Yann-LeCun [6] published LeNet-5 in 1990, which trains convolutional neural networks (CNNs) to recognize a database of handwritten digits (Modified National Institute of Standards and Technology (MNIST)) using backpropagation, which is the starting point of the CNN network architecture. These techniques have been applied in various fields such as for object detection and semantic segmentation. Subsequently, owing to the development of new hardware, GPUs, and numerous deep learning studies, Krizhevsky developed AlexNet [7] in 2012, which is a model that is significantly larger than the existing CNN models. It was trained on the ImageNet dataset with a high-performance GPU, and significantly outperformed existing best-performing models. Since then, inspired by AlexNet, many studies on CNNs have been actively conducted, the layers of CNN models have been deepened, new model structures have been studied, and models with satisfactory performance are continuously emerging.

Privacy is one of the important issues in the information society that enables technology and services. Communication, multimedia, biometrics, big data,

cloud computing, data mining, internet, social, etc. each can potentially be a means of invasion of privacy. The main reason for the increasing privacy concerns in today's network society is the advent of new technologies such as street view, social network, and big data. And as the scale of surveillance increases, the number of CCTVs is steadily increasing, and as a result, privacy is becoming more and more vulnerable [8]. Faces are the main identifier for multimedia content that needs to be de-identified for privacy reasons. Facial recognition data has traditionally been collected, processed, and stored by dedicated recognition systems, but advances in mobile and camera technology today make it easier to capture and share facial recognition data online. Numerous face images are being uploaded to various social media platforms. These images can be easily used to identify and match people by inferring sensitive information about individuals or profiling selected users [9].

Advances in AI require large amounts of high-quality training data. The rapid increase in Internet use in recent years has resulted in the amassing of a considerable amount of usable data. In addition, a large amount of data, collected through autonomous vehicles, data collection vehicles, CCTV, and Street View, can be used as high-quality learning data. However, images collected from various sources, obtaining by using cameras, include personal information, such as faces and car license plates. Currently, there are restrictions on data collection and distribution under the Personal Information Protection Act, hindering the development of related industries. In the case of Street View, faces are blurred and obscured to de-identify personal information. As a result, unlike the original video, it is visually unnatural, and its value as AI learning data is diminished. For example, AI cannot train an object called 'person' on an image where a person's face is blurred. Therefore, it cannot be used as essential learning data for autonomous vehicles or CCTV control systems.

Figure 1 shows an example of anonymizing personal information in Street View. Because a person's face is classified as personal information, the faces in the image have been blurred. As shown in Fig. 1, the value of collected data as training data is lost when faces are blurred. In addition, because the de-identification of large-capacity big data is performed manually, significant time and financial costs are

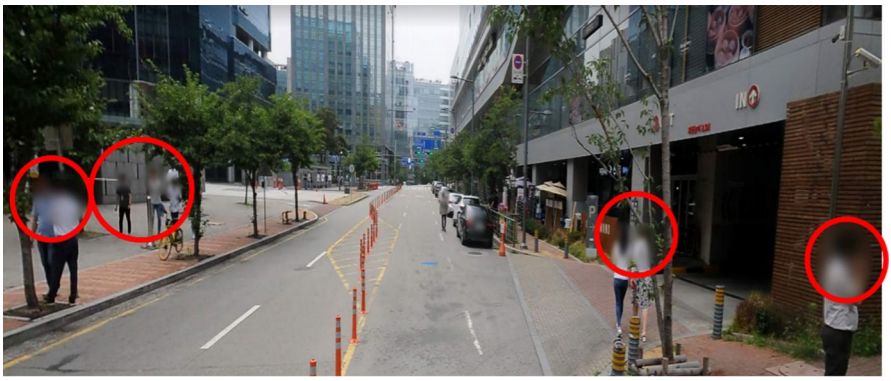

Fig. 1 Example showing anonymization of personal information

associated with this task. Furthermore, work efficiency is reduced and mistakes can occur. Street views can provide high-quality training data for autonomous vehicle research. However, because the faces are blurred, the image is no longer a human subject and cannot be used as training data.

To solve these problems, we aim to de-identify personal information while maintaining the value of the image data collected as data for AI training. This can be accomplished by generating a nonexistent face with a different appearance while maintaining the basic structure of the human face, i.e., the shape of the face, positions of the eyes and mouth, and angle of the face. Generating a face in this manner creates a natural image. Because the classification is maintained and only the appearance changes, it can be used as data for training various AI. Therefore, a generative adversarial network (GAN) model was applied to generate a face with a different appearance, while maintaining the basic structure and angle of the human face.

2 Related work

2.1 Face de-identification

The removal of sensitive and private information contained in videos and images is a challenging task. Currently, research on the de-identification of facial information in images is ongoing, and in general, all the sensitive information is removed by changing the original image. However, it is difficult to remove all the personal information using this method.

Until recently, the face de-identification technology of computer vision was able to de-identify faces, which are personal information in images, using approaches such as intuitive black box, blurring, pixelation, and masking. The black box method is a method of replacing a black or white square box in the face area after face recognition and face location check in the image. Blur is a simple way to use a Gaussian filter to soften the faces in an image. Pixelation is to reduce the resolution of a face area [8]. These methods are commonly used owing to their simplicity and ease of use. However, although sensitive personal information has been successfully removed, useful information from data is also removed, resulting in a loss of their value as data for AI learning. In addition, deep learning CNN-based recognition models can successfully identify faces with high accuracy from de-identified data using these techniques, leaving the risk of privacy exposure [10, 11]. This approach also contributes to privacy, but makes all other information contained in the image somewhat useless. Therefore, there is a need for a technology that can balance the usefulness of biometric information and privacy [9].

To solve this problem, Newton et al. [12] proposed the k-Same algorithm, which is the first privacy protection activation algorithm. It determines the similarity between faces based on a distance metric and generates a new face by averaging the image components, which can be either the original image pixels (k-Same-Pixel) or eigenvectors (k-Same-Eigen). Although this procedure theoretically limits the recognition performance to 1/k, the resulting image sometimes contains ghost-like objects

owing to minor alignment errors [13]. Since then, various methods of the k-Same algorithm have been proposed to naturally improve de-identified face images [14–18]. However, these methods have limitations in that the level of privacy protection is lowered when there is an image sharing similar biometric characteristics, and the de-identified face may be similar to the original image or seem unnatural.

With the introduction of GANs [19, 20], active research is being conducted in the field of image generation, which has demonstrated various possibilities. GANs allow for the creation of new images that are indistinguishable from the actual data distribution. Therefore, this is an effective method for facial recognition.

Chen et al. [21] proposed the privacy-preserving representation learning variational GAN (PPRL-VGAN) that uses variational autoencoders (VAE) with GAN, and argued that their approach balances privacy and data usefulness. This method can control de-identification using a condition vector for ID. However, because this control vector is a one-hot encoded vector, the scope of de-identification using it is limited.

Ren et al. [22] used GANs to de-identify privacy-sensitive information (i.e., human faces) in video data. They trained a face modifier to remove privacy-sensitive information, while the action detector attempted to maximize the spatial action detection performance. Thus, de-identification was performed by hiding personal information and modifying the pixel level of the video frame.

Li et al. [23] proposed the use of hostile disturbances to protect the sensitive information in images from humans and AI. In the privacy-preserving attribute selection (PPAS) algorithm, de-identification was performed by manipulating the attributes such that the distribution of facial attributes was close to the actual distribution. Wang et al. [24] introduced a discriminator and de-identified it by manipulating the attributes using a two-step method.

DeepPrivacy [25] de-identifies faces from the original by directly removing identifiable regions of the face and generating new faces based on sparse pose estimation. In addition, they introduced FDF (Flicker Diverse Faces), a diverse dataset of human faces, including unconventional poses, occluded faces, and vast variability in backgrounds. The FDF data set consists of 1.47 million human faces with a minimum resolution of 128×128 with face key points and bounding box annotations for each face. And the proposed model is a conditional GAN that generates images based on the surrounding of face and sparse pose information. The 7 key points of the face are used to describe the pose of the face. And to reduce the parameters of the network, the pose information is preprocessed into a one-hot encoded image. As a learning method, progressive growth training techniques are applied. However, this solution can generate unrealistic images with high occlusions and irregular poses.

Zhongzheng Ren et al. [22] applied adversarial training in which two competing systems fight, and the result was to perform pixel-level corrections to de-identify human faces with minimal impact on motion detection performance. There is a problem that the face generated using the method can identify a person, and since this method cannot control the generation process, all IDs are mapped to the same fake ID.

Sun et al. [26] proposed a head inpainting obfuscation technique. They proposed the generation of facial landmarks and conditioned head inpainting in the image

context for the smooth hypothesis of rational head poses. As a method of generating a de-identified face by modifying the landmark of the face, which is different from the face in a real image and often results in unnatural-looking results.

Maximov et al. [27] proposed the conditional identity anonymization generative adversarial network (CIAGAN), an image and video de-identification model based on conditional generative adversarial networks. CIAGAN harnesses the power of generative adversarial networks to create photorealistic images. They proposed a new identity discriminator to train CIAGAN to control the identity generation process and ensure anonymization. In this proposed method, CIAGAN de-identifies a face based on the desired ID from a face image. It also uses facial landmarks and identifies the one-hot vector to remove the identifying characteristics of human faces. The method also provides full control over the de-identification process, ensuring both anonymity and diversity.

IdentityDP, proposed by Wen et al. [28], is a face de-identification method that combines a data-based deep neural network (DNN) with a differential privacy (DP) mechanism. It consists of three steps. The first step is to extract the entangled high-level identity attribute representation and train the network to reconstruct the original face. The second step generates a perturbed identity representation based on the novel Laplace ε-IdentityDP mechanism. The third step is to generate an anonymous face from the perturbed identity representation and original attribute representation through a frozen network.

Cho et al. [29] proposed a face de-identification method based on a deep generative model that separates the encoder's output vector into identity-relevant parts and the remainder representing face attributes. It solves the face attribute-related vectors from the latent vectors and efficiently transforms the identity-related vectors into new identities. However, the output of the proposed method is degraded compared to the actual image, and discontinuity of the plane boundary appears.

Zhan et al. [30] proposed a model that can de-identify while maintaining as much as possible the facial attribute information of the original video, such as facial expression, gender, hairstyle, and whether or not glasses are worn. Multi-attribute retention uses shallow face attribute information and deep face attribute information and adopts different processing strategies according to face attribute information based on the uniqueness of the face attribute information.

Kuang et al. [31] proposed a de-identification generative adversarial network (DeldGAN) for face de-identification. DeldGAN anonymizes a face by seamlessly replacing a given face image with another synthetic but realistic image. This article consists of two steps. First, we obfuscate the original ID by anonymizing the input face. And the anonymized face is synthesized using a designed de-identification generator. However, there are potential risks of background attacks and often unnatural consequences.

Zhang et al. [32] proposed a Domain Embedded Multi-Model Generative Adversarial Network (DEGAN) for Face Inpainting. Face Inpainting extracts features from images to restore and generate hidden and invisible parts. In this paper, they aim to generate images similar to reality using features extracted from image data by applying GAN. A face region domain (face mask, face part, landmark image) is inserted into the latent variable space as a guide by a variable autoencoder. Only

facial features are placed in the latent variable space. It then combines the latent variables contained in the face region into a generator for inpainting the face. Finally, the global discriminator and the patch discriminator determine whether the finally generated distribution is close to the actual distribution.

Li et al. [33] proposed a solution called SF-GAN (Secret Face Generative Adversarial Network) that does not lose face attribute information. The proposed model effectively de-identifies faces while maintaining as much information as possible on the facial properties of the original image, such as facial expressions, gender, hairstyle, and whether glasses are worn. They use a face segmentation network to retain thin facial attribute information to preserve facial attributes, and feature extraction to retain deep facial attribute information. However, the result produced is similar to the original face and is unnatural.

2.2 Generative adversarial network (GAN)

GAN [19] is a training method that was first introduced by Ian Goodfellow in his 2014 paper. Based on the continuous development of the DNN model, various studies on the GAN model are being actively conducted and they are showing interesting results and diverse possibilities. GAN is a deep learning model that is mainly used to generate images similar to datasets. Classification models work by extracting and classifying features from data. The output of a typical machine learning model is the prediction of an object or interval prediction of a continuous random variable. Its purpose is not to generate a shape, but to find the highest probability of it. In contrast, GAN generates data. In other words, the shape of the data is generated. The shape of the data implies the distribution or variance; generating a data distribution does not generate a function that finds a result but rather generates the data in its actual shape.

The characteristic of GAN is that it is adversarial. Opposing objective functions between the generator and discriminator networks lead to an adversarial relationship. The generator (G) generates new data, and the discriminator (D) compares the generated image with the sample data. The discriminator can easily identify if the first data generated by the generator is fake; however, with an increase in training and data generation, the difficulty for the discriminator to make the distinction also increases. In this case, the adversary arises because the goal of the generator is to identify the generated image as a real image, whereas the goal of the discriminator is to identify a fake image as fake. That is, learning converges when the generator learns how to generate a near-realistic image from a random noise vector and the discriminator cannot distinguish between the real image and the image generated by the generator. In GAN, the generator and discriminator structures are in the form of neural networks. A simple GAN model structure is illustrated in Fig. 2.

As mentioned previously, GANs have adversarial principles. The generator and discriminator compete with each other to make the fake images appear real. GAN learning trains a discriminator with a real image, which is the training data, to find the distribution of the training data, and uses a random noise vector as an input to the generator to generate a fake image. It initially produces very crude images, but as

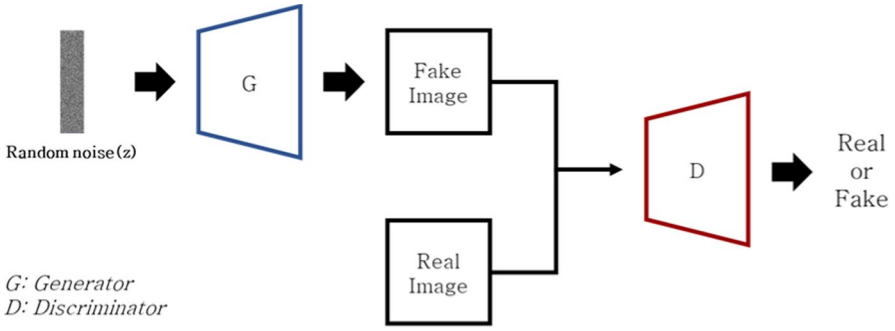

Fig. 2 GAN model structure diagram

the distribution approaches the distribution of the training data, it produces images that are similar to real ones. Finally, the generated image (i.e., the fake image) is distinguished from the real image by the discriminator and the generator learns to generate a more realistic fake image through error backpropagation. If this method is repeated, the discriminator and generator recognize each other as hostile, leading to the learn of each. Consequently, the generator can generate a fake image that is similar to the real image; thus, the discriminator cannot differentiate between the real and fake images. That is, GAN generators compete with each other by trying to lower the discriminant probability, whereas the discriminator tries to increase the discriminant probability.

$$\min_{G}\max_{D} V(D,G) = E_{x\sim p_{\text{data}}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log\left(1 - D(G(z))\right)\right]. \tag{1}$$

Equation (1) represents the objective function of the GAN. This function is learned using $V(D, G)$ in the min–max problem method, as shown in Eq. (1). $X\sim P_{\text{data}}(x)$ is the probability distribution data for a real image and $z\sim P_z(z)$ is an image sampled from noise using a Gaussian distribution. In general, z is commonly referred to as a latent vector, meaning a vector in dimensionally compressed latent space. $D(x)$ is the discriminator and a value between 0 and 1 indicates the true probability. $D(x)$ outputs 1 if the image is real, and 0 if it is fake. The discriminator $D(G(z))$ of the second term has a value of 1 if the data generated by G and $G(z)$ are determined to be genuine, and a value of 0 if they are determined to be fake. Training D to maximize $V(D, G)$ is the process by which the discriminator learns to correctly classify real and fake image. Minimizing $V(D, G)$ is irrelevant, as it is omitted because there is no G in the first term. For the second term to be minimal, $D(G(z))$ must be close to 1. This means that the discriminator trained the generator to generate a fake image that is sufficiently realistic to be identified as a real image. Thus, learning a discriminator to maximize $V(D, G)$ and learning a generator to minimize $V(D, G)$ is the min–max problem method.

Conditional GAN (cGAN) [34] is a method of applying certain conditions to the original GAN. For example, in MNIST, which recognizes handwritten digits to

generate the desired number you can add a label (one-hot vector) that corresponds to the class of the number. However, this is possible in many forms other than class labels.

Figure 3 shows the structural diagram of the cGAN model. The cGAN model comprises addition of the condition c to the generator and discriminator of the GAN model. Equation (2) is the objective function of the cGAN.

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}\left[\log D(x|c)\right] + E_{z\sim p_z(z)}\left[\log\left(1 - D(G(z|c))\right)\right]. \tag{2}$$

Since the introduction of the GAN learning method, several studies have been conducted. Among them, image-to-image translation [35] uses GAN, which deals with interdomain changes in the images. In other words, it is a type of generative model that maps input and output images using an image dataset. This field converts black-and-white photographs into color photographs, and daytime landscapes into nighttime landscapes. The existing image translation problem was solved by applying a CNN. However, it is difficult to design effective losses when training a CNN. For example, if a commonly used CNN is trained to reduce the Euclidean distance between the target and prediction result, the model outputs a blurry image. This is because CNN solves the problem based on loss. Rather than finding the optimal answer for each pixel, it attempts to reduce the loss in all the pixels of the image. This is because predictions are made using safe values rather than accurate values. GANs have been used to address this problem. Phillip Lsola is the author of Pix-2Pix, a paper presented at UC Berkeley's 2017 Computer Vision and Pattern Recognition (CVPR) [36]. It detailed converting black-and-white images to color images and approaching the conversion through several CNN-based techniques. Because color images can be easily collected and paired, black-and-white images can be easily generated; consequently, the author collected paired data and applied GAN-based learning to the problem. The generator uses an image as the input and outputs a new image. The discriminator uses two inputs: real and generated images in pairs. When

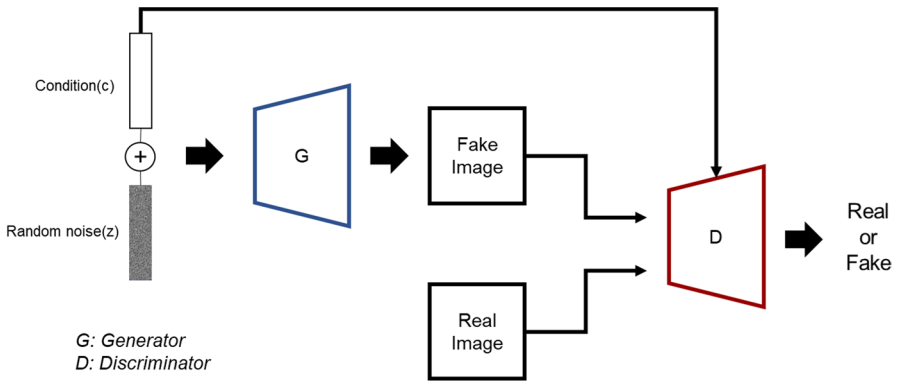

Fig. 3 cGAN model structure

two inputs are provided to the discriminator simultaneously, it compares the pairs to determine whether the image is fake or real and learns from this process. For image translation problems, the generator has a symmetric encoder–decoder structure, such as semantic segmentation, and upsamples the downsampled feature map to the decoder via encoder pooling. The discriminator was based on the PatchGAN [37]. The input image was decomposed into $N \times N$ resolution patches to determine whether each patch was real or fake. The output of the discriminator is the average of the true values of all patches. It understands the details of an image by learning to identify patches, rather than the entire image.

Figure 4 shows an image–image transition model structure wherein the source image is converted to the same style as the target image. For example, if a black-and-white photograph becomes the source image and a color photograph becomes the target image, learning progresses, and inserting a black-and-white photograph converts it into a color photograph. Similarly, general landscape paintings can be converted into Van Gogh paintings.

There are three main types of GANs currently in use: GAN, cGAN, and image-to-image translation. These models are compared using a human face image. Image generation and translation models using GANs have advanced significantly and many studies on the generation of nonexistent fake faces are in progress. Boundary equilibrium generative adversarial networks (BEGAN) [38] and style-based generator architecture for generative adversarial networks (StyleGAN) [39], which show satisfactory performance in face generation models, have been presented. However, these face generation models are GAN structures based on random noise vectors as inputs. The facial images generated by these face-generating models are high-quality images that cannot be compared to those of real people. However, the disadvantage of face generation models that use random noise vectors as input is that they cannot generate the desired face angles and shapes because the input values are noisy. Because this study aims to generate a nonexistent face while maintaining the angle and shape of a human face in a real image, pre- and post-processing is required to apply the above method.

There are two main image-to-image translation models: Pix2Pix [36] and CycleGAN [40]. In face generation and image-to-image translation, paired and nonpaired datasets are required to train the two models. However, it is difficult to construct a

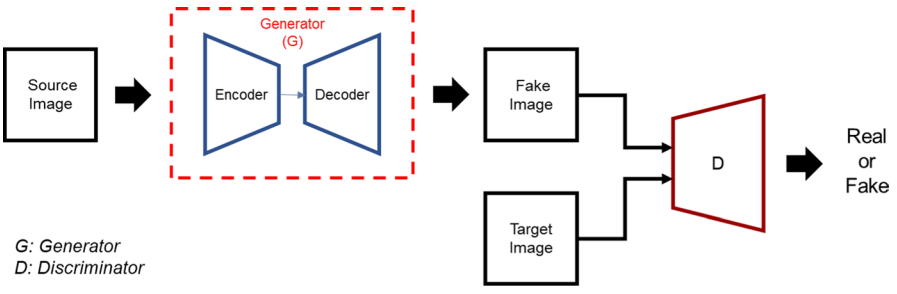

Fig. 4 Structure of the image–image translation model

human face image. As image-to-image translation tends to retain its original shape, it produces the same image as the input image.

Figure 5 shows a representative GAN-based face generation and transformation model. (a) Presents faces generated using the original GAN. (a) Generates a fake face image similar to the real image by reflecting the distribution of the training image through the GAN from the random noise vector. Therefore, the angles and shape of the face cannot be generated as desired and the face is generated randomly. (b) Shows the conditional GAN (cGAN) [34], which is a method of applying a specific condition to the basic GAN. Conditions can also take many forms besides the vectors. In (b), a fake face is outputted by learning the sunglass label as a condition. Accordingly, faces wearing sunglasses are generated. However, because (b) also generates a face by inputting a random noise vector and condition, the angles and shape of the face cannot be generated as desired and a conditional random face is generated. (c) Presents an image-to-image translation

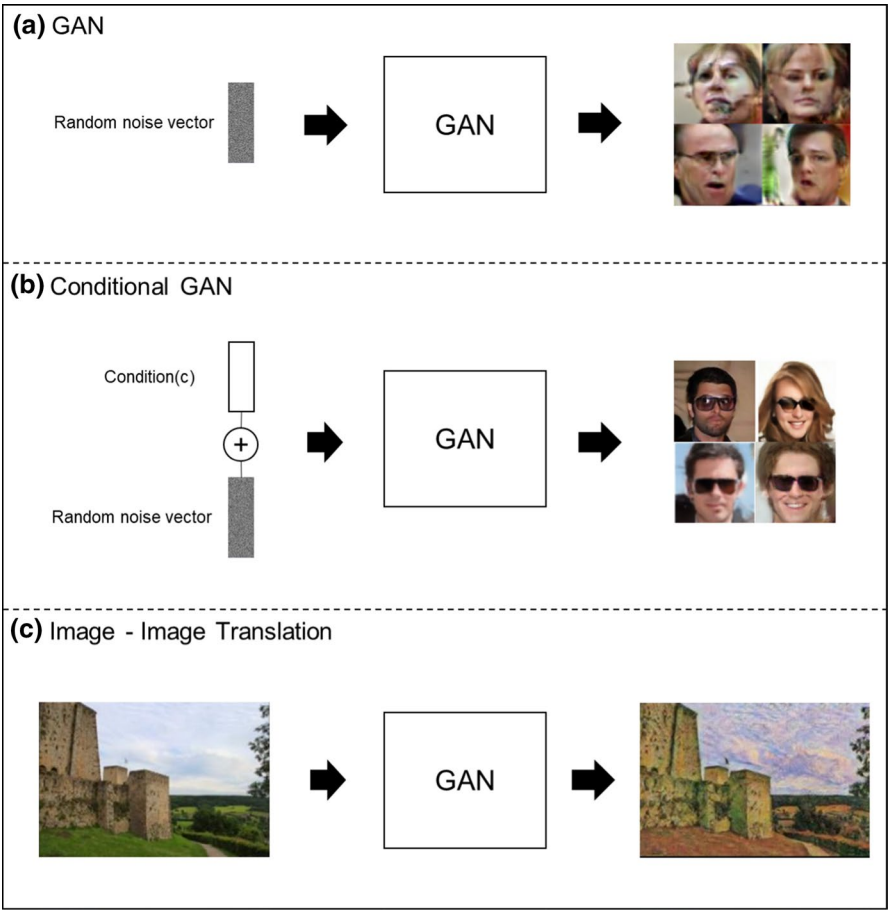

Fig. 5 Comparison of GAN-based face generation models

model that takes a real facial image as the input and translates the style of the image. With regard to the input face, the output preserves the face shape and angles, as well as the impression, so that the face does not change. (a) And (b) generate nonexistent fake faces. However, because the angles and shape of the face are randomly generated, this is unsuitable for the generation of a fake face that maintains the angle and shape of the real face. The shape and angles of the real face are maintained in (c). However, because face information is maintained without change, it is not suitable for this study because the target face information cannot be de-identified. Therefore, as shown in Fig. 6, we propose a landmark-based (L)-GAN that generates a fake face by sufficiently altering the image to ensure privacy while maintaining the angles and shape of the original face.

Our method is based on the structure of image translation and it is used as a condition for the L-GAN model by detecting and extracting facial landmarks [41] that can determine the angles and shape of the face from the real face image. Unlike the one-dimensional conditional vector of the existing cGAN, the basic structure of the face (position and direction of the eyes, nose, mouth, etc.) is conditional using two-dimensional spatial characteristics. Then, for the formation of a fake face, random noise images and facial landmark images were concatenated and input into the L-GAN. As shown in Fig. 6, it is possible to generate faces with altered appearances while maintaining the basic facial structure of the landmark image. This method is different from the deepfake [42] techniques. Deepfake is a facial image synthesis technology that uses a GAN to synthesize the face of another person onto the human face in the original image. The method proposed in this study aims to generate fake faces by de-identifying facial information based on landmarks rather than real human faces. In this paper, we propose a facial L-GAN model to effectively de-identify facial information.

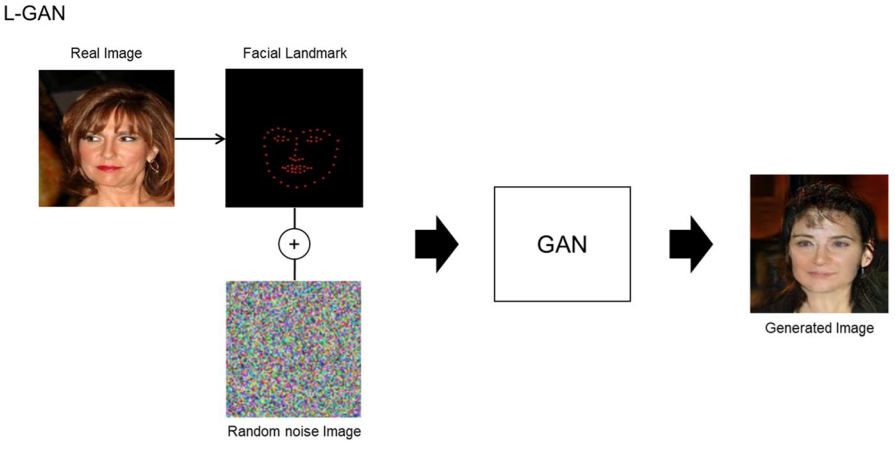

Fig. 6 L-GAN concept diagram

3 System design and implementation

The fake face generation system using the facial landmark image proposed in this study consists of two steps. Figure 7 shows a conceptual diagram of the proposed system. The real face image is input into the facial landmark detector to detect human facial landmarks and the coordinates are extracted to visualize the landmark coordinates as a blank image of the same size as the input. Subsequently, the facial landmark image and real face image are input into the L-GAN model and it is trained to generate a fake face.

3.1 Facial landmark detection and extraction

Facial landmarks [41] are also known as facial feature points. Facial feature point detection tracks facial features and is primarily used for face alignment [43]. This technology predicts the locations of facial landmark feature points by extracting and learning features from numerous faces. Facial features, including the jaw line, eyes, nose, eyebrows, and chin are the subjects of the analysis.

Facial landmarks include the eyes, nose, mouth, eyebrows, and chin, which determine the appearance of a person's face. Facial landmarks are useful because they are key features of many facial analysis methods and algorithms. Examples of facial landmarks include face recognition, face alignment, facial expression recognition, face deformation, drowsiness detection, eye-blink detection, and head angle estimation. Facial landmarks localize feature points representing the unique locations of facial components such as eyes, nose, mouth, eyebrows, mouth, and jawline, as shown in Fig. 8. The facial landmark detection process detects faces in images and the feature points of the face within the detected area. Face recognition uses Dlib [44], which can be utilized by applying various machine learning algorithms. In Dlib, the histogram of oriented gradients (HOG) [45] function can be utilized for face detection or a trained CNN can be applied. In addition to face detection, it can be used to detect a wide variety of objects. Next, a landmark feature point of the face is detected within the detected facial region. Facial feature points are detected through a facial landmark detector included in Dlib. The facial landmark detector extracts features from numerous faces and learns to predict the location of feature points. Face detection is performed before facial landmark detection. After detecting a face, it is easier to identify components within the facial area (Table 1).

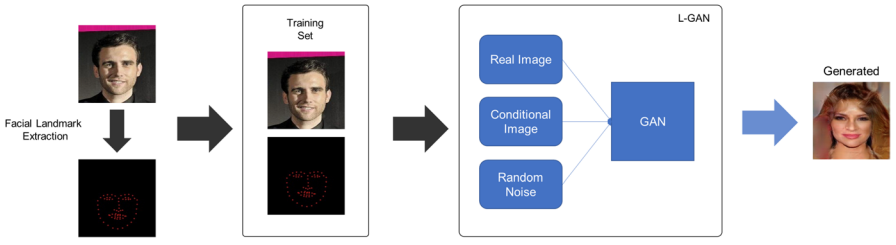

Fig. 7 Overall system structure

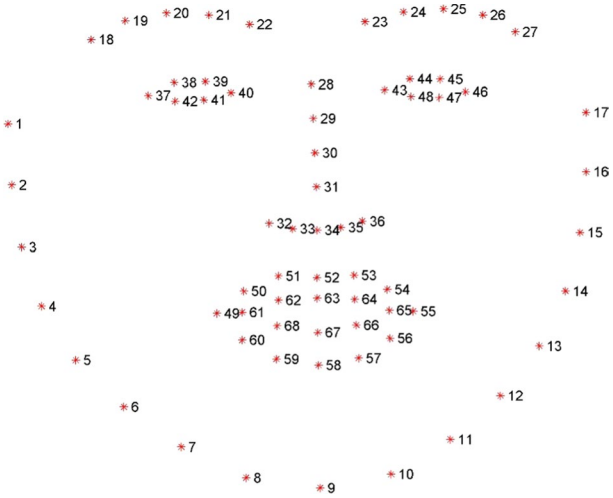

Fig. 8 Sixty-eight facial landmark coordinates visualization

Table 1 Facial elements with specific coordinates

Face components	Landmark coordinates
Jawline	1–17
Right eyebrow	18–22
Left eyebrow	23–27
Nose	28–36
Right eye	37–42
Left eye	43–48
Mouth outline	49–61
Mouth inner	62–68

As shown in Fig. 9, the facial landmark algorithm collected 68 coordinates from the facial image. If the landmarks of the face are detected accurately, it is possible to estimate the rotation angle of the head and the shape and basic structure of the face. In addition, the facial landmark detection comprises detection of features representing the components of the face and the basic structures and angles of the eyes, mouth, and nose are used as conditions for maintaining the shape of the face image generated through the L-GAN. That is, it maintains the shape and angle of the original image using the landmark. In the existing cGAN, the condition was a one-dimensional vector; however, in this study, a two-dimensional matrix representing the spatial features of the face was adopted.

In this study, OpenCV [46] and Dlib were used for facial landmark detection and extraction, and the facial landmarks were extracted, as shown in Fig. 10.

When a face image is inputted, face detection is performed after converting the image to grayscale to suppress noise. Face detection is performed because it is easier

Fig. 9 Facial landmark example

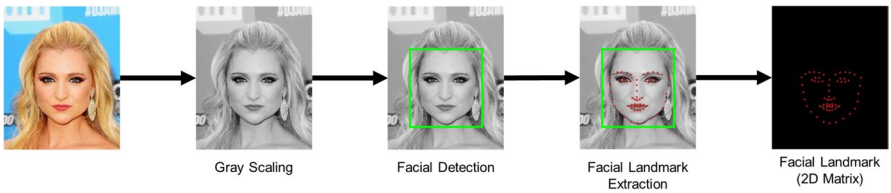

Fig. 10 Facial landmark detection and extraction process

to find components in the face after finding the detected face region. The facial feature points are then detected and predicted using a pretrained AI system. Finally, only landmark coordinates of the detected face are extracted. The landmark coordinates are expressed in a blank image of the same size as the input image, as a two-dimensional matrix, and are visualized. That is, the condition of the L-GAN is composed of an image with two-dimensional spatial characteristics, unlike the one-dimensional condition vector of the existing cGAN, and contains information such as the basic structure and angle of the face.

3.2 L-GAN model

In this study, we propose the L-GAN model for effective de-identification of facial information. We describe the L-GAN model that generates an image of a face while maintaining the angle and shape of the faces contained within the previously processed data. Figure 11 shows the structure of the L-GAN model.

The preprocessed face landmark image and the real face image are each composed of pairs that are input into the L-GAN. A random noise image (z) and conditional face landmark image (L) are input into the generator to generate a new image. The discriminator receives the real image (Y), image generated by the generator ($G(L, z)$), and conditional image (L).

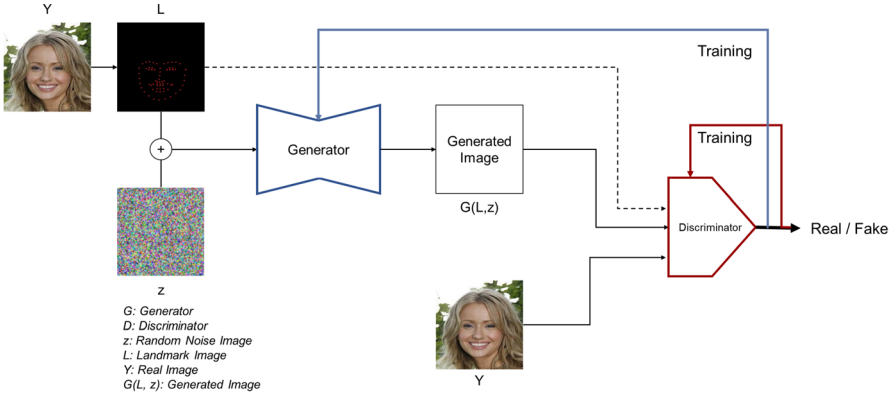

Fig. 11 L-GAN model structure

Then, it determines whether the $G(L, z)$ image is real or fake. In this process, the discriminator makes a decision by considering the conditional image, which is effective in generating an image based on a landmark. Through this process, the image relationship is learned; thus, an image that is similar to the real image can be generated from the conditional image. In addition, because $G(L, z)$ is generated from L and z, the basic structure and angle (based on the landmark) are maintained to generate a face with an altered appearance.

$$\min_G \max_D V(D, G) = E_{y,l}\left[\log D(y|l)\right] + E_{z,l}\left[\log\left(1 - D(G(z|l))\right)\right]. \tag{3}$$

Equation (3) is the objective function of the L-GAN. The objective function of the L-GAN is learned by solving the min–max problem using $V(D, G)$. In the first term, $D(y|l)$ is a discriminant that outputs a value between 0 and 1, which implies that the probability of receiving the condition and real data is true. $D(y|l)$ outputs a value close to 1 if the data are true, and close to 0 if the data are fake. The discriminator $D(G(z|l))$ of the second term has a value close to 1 when the data $G(z|l)$ generated by generator G is indistinguishable from the real data. It has a value close to 0 if it is determined to be fake. D is a learning process to maximize $V(D, G)$ that forces the discriminator to classify data as either real or fake. Minimizing $V(D, G)$ means that the first term is omitted because it is unrelated to G; thus, the second term must be minimized. For the second term to be minimal, $D(G(z|l))$ must be close to 1. This means that the generator is trained such that the discriminator can generate a fake image that is perfect enough to classify the fake image as a real image. Therefore, learning the discriminator in the direction of maximizing $V(D,G)$, and learning the generator in the direction of minimizing $V(D,G)$, is the min–max problem learning method.

3.2.1 Generator and discriminator model

Because the input is essentially an image in the generator model, the model has an encoder–decoder structure. However, in the basic encoder–decoder, dimensions are first reduced and then expanded. Accordingly, the features of the image received by the encoder are expressed as a compressed vector and the decoder generates a result of the desired size; however, the features of the input image are largely lost in this process. Therefore, in this study, we use a U-Net structure that can transmit detailed information of the input image to the decoder by adding a symmetric skip connection to the encoder–decoder to maintain the face landmark information. In addition, using a skip connection has the advantage of solving the vanishing gradient problem during the backpropagation process. Figure 12 shows the encoder–decoder and U-Net structures. The discriminator applies the PatchGAN structure. The traditional GAN discriminator looks at the entire image to identify real and fake patches, whereas PatchGAN divides the image into areas of $N \times N$ size to determine whether each patch is real or fake. This approach is superior because there are fewer parameters to consider and is consequently faster.

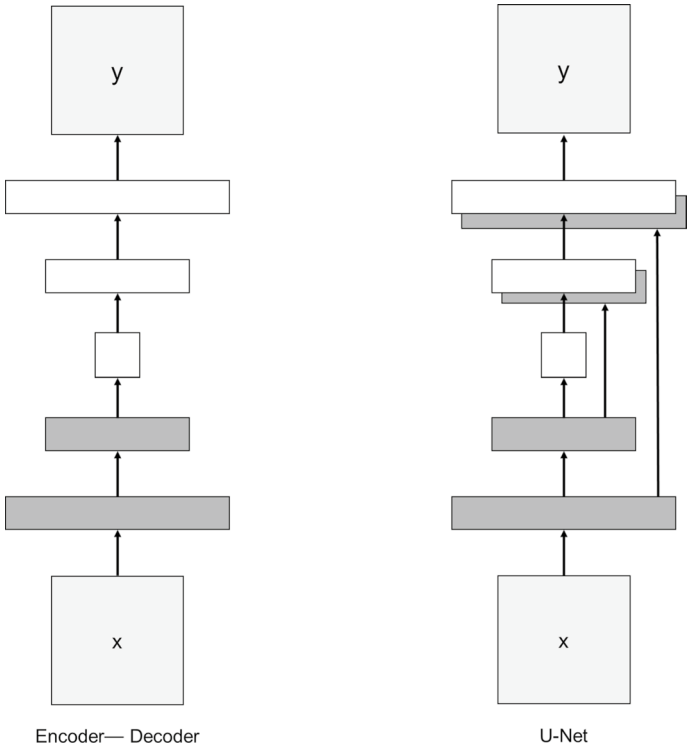

Fig. 12 Encoder–decoder and U-Net architecture

4 Experiment and analysis

4.1 Datasets

In this study, we trained a model and tested its performance using the CelebA dataset [47]. CelebA contains data released through International Conference on Computer Vision (ICCV) in 2015 and is the most commonly used dataset for face generation using GAN. It consists of approximately 230,000 celebrity face images. Forty binary labels were used for each face. Information about characteristics such as youthfulness, masculinity, hair color, hairstyle, and jewelry was tagged. However, this study did not use binary labels for these features or images alone. The training and test sets consisted of 190,000 and 12,000 pieces, respectively. Figure 13 shows some images from the CelebA dataset.

The dataset for the system training proposed in this study must contain pairs of facial landmark images and real images. Therefore, a training dataset was constructed by applying the facial landmark detection and extraction method to the CelebA dataset. As shown in Fig. 14, the landmark image of the face and the real image were paired and used as input data for the L-GAN model.

4.2 Training

In this study, the L-GAN model was trained by applying a U-net structure [48] generator and a PatchGAN structure discriminator. To avoid representing the same

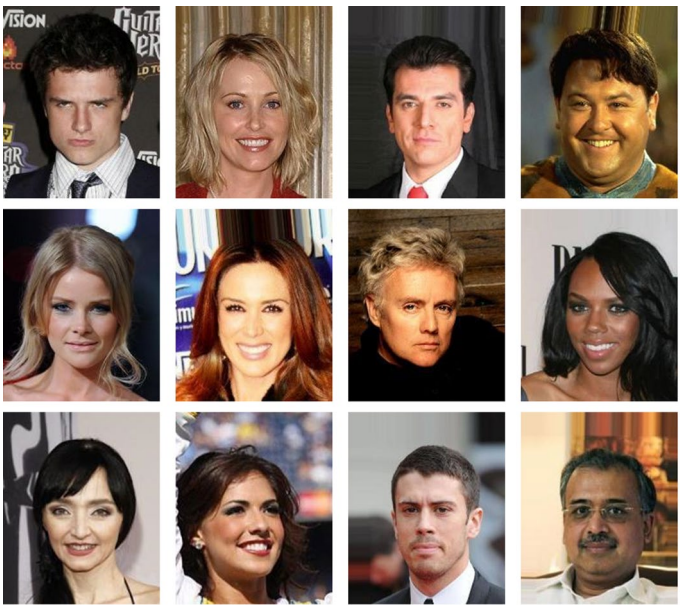

Fig. 13 CelebA dataset examples

Fig. 14 Example of the training dataset configuration

distribution as the training data, both the random noise image and landmark image were input into the generator. A paired training set was used to train the model. The model parameter values were trained by applying 200 epochs, batch size of 64, and learning rate of 0.0002.

The server hardware and software environment in which the model training was performed. High-performance hardware is required for GAN learning, and in this paper, three GPUs (RTX 3080) and AMD Ryzen 2950×CPU, were used for learning.

Figure 15 shows an image generated during the learning process. First, the image generated during third epoch maintains the facial landmark feature point input as a

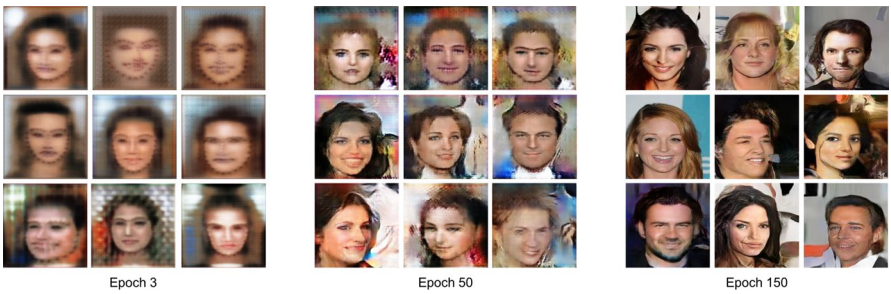

Fig. 15 Images generated during L-GAN training

condition and it can be confirmed that the face is generated according to the corresponding feature. In epoch 150, a fake face image is generated while maintaining the basic structure and angle of the landmark.

5 Results and analysis

The method proposed in this study extracts facial landmarks that can determine the basic structure and angle of a face from a real face image. Through the L-GAN model, we aimed to generate a nonexistent fake face that de-identifies facial information with the basic structure and angle of the real face.

To evaluate the performance of the proposed method, the number of feature points of facial landmarks was compared, and 36, 68, and 106 feature points were extracted, verified, and compared, respectively. Figure 16 shows a comparison of the generated images according to the number of facial landmarks. A total of 36 facial landmark feature points were detected and extracted for the jawline, eyebrows, and nose, providing sufficient information to determine the shape and angle of the face. However, if these 36 landmarks are used as a condition in the L-GAN and a fake face is generated, there is no information about the eyes, mouth, etc.; thus, the generated face image is unnatural and poorly expressed. All basic facial structures including the jawline, eyes, nose, mouth, and eyebrows, were detected from 68 landmark feature points. Similarly, 68 landmark feature points were generated by de-identifying the information on the basic structure and face angle. Finally, 164 landmark feature points were expressed in detail by collecting additional feature points. As a result, there was a disadvantage in that the generated face appeared the same as in the real image, and the goal of de-identification was not achieved.

Next, the images generated for the encoder–decoder and U-Net structures, which are the generator structures, were compared. When landmarks are extracted from real facial images and trained with a generator, the generated image loses some information of the basic structure of the face and becomes blurry. However, learning and experimenting with the generator of the U-Net structure can ensure that the basic structure and the angle of the face are well expressed and clear (Fig. 17).

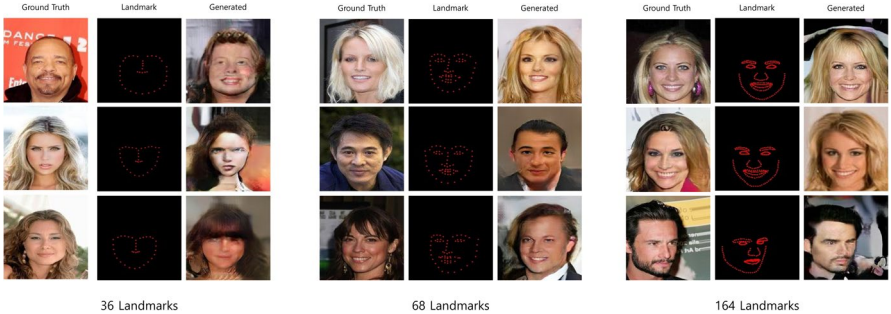

Fig. 16 Comparison of generated images based on the number of landmarks

Fig. 17 Comparison of generated images for generator structures

The dataset used in L-GAN was applied to the other models. The dataset consisted of landmark images extracted from real images and experiments were performed on the CycleGAN and Pix2Pix models. Figure 18 shows a comparison of the models applied to the dataset used in this study. The CycleGAN model is primarily used for style translation. In addition, learning was arbitrarily performed in the real and landmark image domains by applying the dataset in pairs. The result of CycleGAN is that the basic structure and angle of the face based on the landmark are not maintained, and the face cannot be generated properly. Pix2Pix is a model that is primarily used for image conversion and only landmark images, given as experimental conditions, are input into it. As a result, the basic structure and angle of the face based on the landmarks were lost. Furthermore, because no arbitrary image was input, a face similar in appearance to the original image was created and blurred. The last image was generated by the L-GAN proposed in this study. The basic structure and angle of the face were maintained to generate faces with different appearances.

As shown in Fig. 19, faces with various impressions and appearances are generated by de-identifying face information, while maintaining the basic structure and angle of the face according to landmark conditions.

Figure 20 shows the results of the images generated according to various landmark angles. New faces were generated according to the angle and shape of the faces expressed in the landmark.

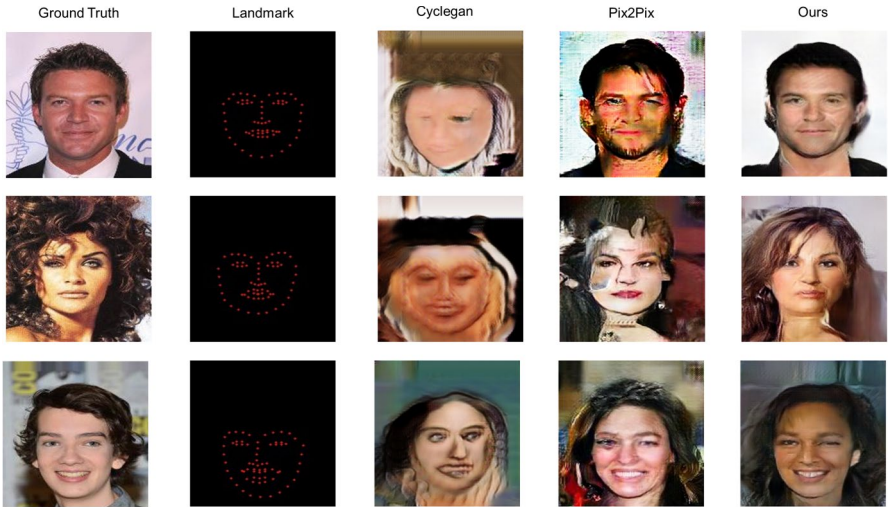

Fig. 18 Comparison of image generation results of multiple models with facial landmark dataset applied

Fig. 19 L-GAN face image generation result

An experiment was conducted on the detection rate, which is an important function that the face de-identification method should have. We want the detector to have a high detection rate for new faces generated by the trained L-GAN system. We compared it with traditional de-identification methods.

Pixelation is a method of clustering pixels in a face region in a color space and then replacing the clusters with an average value. We set the cluster to 8×8 and 16×16 and proceeded with the experiment. Blur conducted an experiment by applying the method proposed by Ryoo et al. [22] by downsampling the face region to a low resolution and then upsampling again. When downsampling the face region

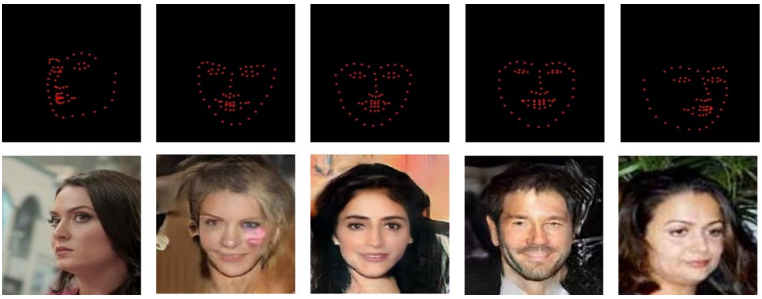

Fig. 20 Image results generated according to landmark angle

to a low resolution, the experiment was conducted with 7×7 and 19×19 settings. A mask is a method of erasing by masking the entire face. We masked the face by filling the entire area with black. Figure 21 shows various face de-identification methods. From left to right, the actual image, pixelization (4×4, 8×8), blur (7×7, 19×19), mask, and the face image generated by our method are in order.

We evaluated the above traditional de-identification method and the detection accuracy of de-identified faces with our method. The face detector was evaluated using the classical method Dlib [44] and the deep learning-based Dual Shot Face Detector (DSFD) [49] model. To evaluate the performance of the detectors, we use the percentage of detected faces. Table 2 shows the face detection accuracy of the existing de-identification method and our method. In Table 2, the face recognition rates of de-identified images generated by L-GAN are shown in bold. In de-identified images generated by L-GAN, face detection rates were 97.6% in Dlib and 97.2% in DSFD. The blurring method has a low detection rate and cannot detect faces at all in pixelated and masked images.

Since L-GAN generates a face based on the facial landmarks in the original image, the structure and shape of the original face are the same. Therefore, there

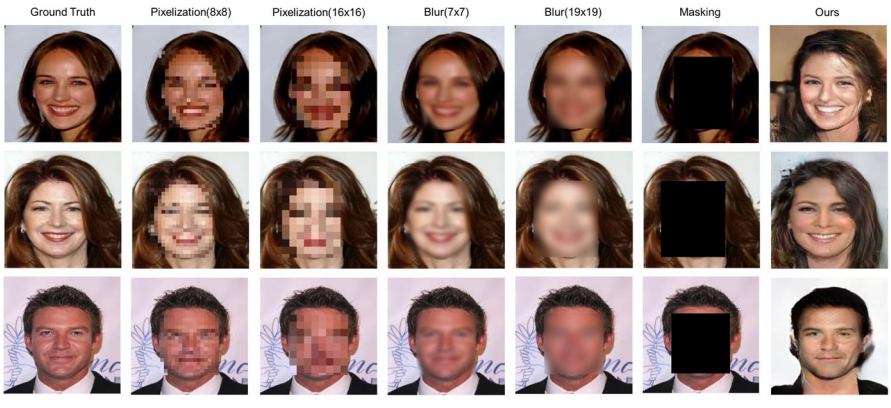

Fig. 21 Comparison with traditional de-identification methods

Table 2 Results of detection accuracy of de-identified faces

Models	Dlib	DSFD
Original	99.5	99.8
Pixelization (8×8)	2.2	2.5
Pixelization (16×16)	0.2	0.1
Blur (7×7)	90.2	39.3
Blur (19×19)	54.1	0.2
Mask	0.0	0.0
Ours	**97.6**	**97.2**

is a feature that the facial expression of the face is also preserved. We compared the face generation results with DeepPrivacy, which is a representative method for de-identification. Figure 22 shows the results of DeepPrivacy and our model. The first column is the original face image. The second column is the face generated by DeepPrivacy and the third column is the result of the L-GAN. Faces generated by DeepPrivacy differ slightly from the actual face structure and shape. And you can see that the facial expression is different from the original face. On the other hand, it can be confirmed that the image generated by the L-GAN has the same structure, shape, and facial expression of the face as the original image and has been de-identified.

An experiment was conducted to check whether the face generated through L-GAN was de-identified. Figure 23 shows the results of identifying an individual using a face recognizer on the generated face image. Testing was performed using four IDs. In the case of an actual image, an individual's ID is recognized and in the case of a generated image, it is detected as 'unknown.' Here,

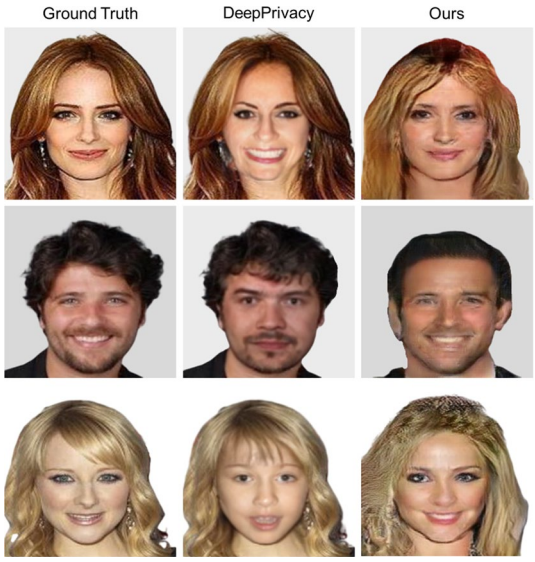

Fig. 22 Comparison of our method with DeepPrivacy

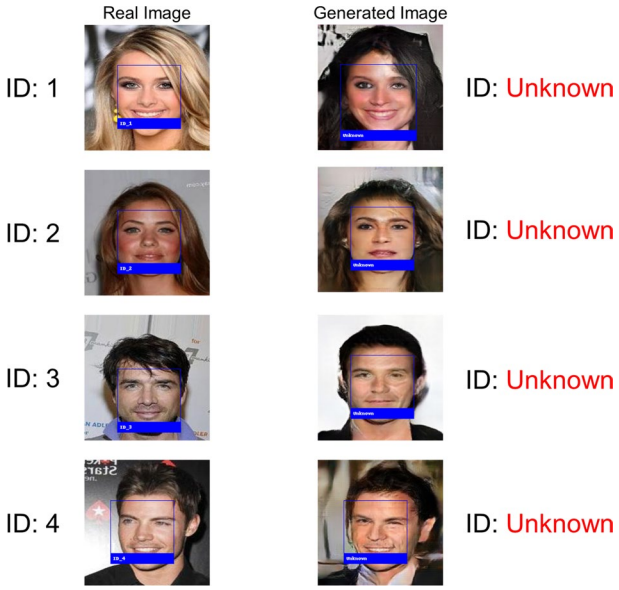

Fig. 23 Results of identifying an individual using real and generated faces

'unknown' is a case in which the softmax result value of all recognized IDs is less than 0.6. Therefore, it was confirmed that the face generated through L-GAN was de-identified.

6 Conclusion

In this paper, we proposed an L-GAN model that generates fake faces by de-identifying real face information while maintaining the basic structure and angle of real faces in images. The existing face generation GAN model generates a face by receiving an arbitrary noise vector; thus, it is difficult to generate the desired face angle and shape. The method proposed in this study detects, extracts, and visualizes facial landmarks to maintain the basic structure and angle of a real human face. Unlike the one-dimensional condition vector of the existing cGAN, a nonexistent fake face of a different shape is generated while maintaining the basic structure and angle of the face under the landmark image condition with two-dimensional spatial characteristics.

In this study, the number of feature points (36, 68, and 106) of facial landmarks was compared to evaluate the performance of the proposed method. Tests and comparisons were performed after learning according to the number of each feature point and the best performance was obtained when the number of facial landmark feature points was 68. The structure of the L-GAN generator was compared with the encoder–decoder structure, which is the basic structure, and the U-Net structure was proposed herein. In the encoder–decoder structure, the detailed information of

the input image is often lost. Conversely, if the U-Net structure is applied, detailed information of the input image can be transmitted to the decoder through the skip connection; thus, the face is generated without the loss of feature points of the face landmark. Finally, the training dataset presented in this study was applied to other models and compared. The training results for CycleGAN and Pix2Pix were compared with the method proposed in this paper. The proposed method substantially maintained the basic facial structure and angle and created a de-identified nonexistent fake face.

The method proposed in this study is dependent on training data. In particular, it relies heavily on real face images. However, in this study, both the face and background were included in the real face image. The background is not clean because it is significantly influenced by the model when generating the face image. Therefore, good performance can be expected from future research if the area of the face is segmented and configured as training data, without including the background.

Acknowledgements This paper was supported by Konkuk University Researcher Fund in 2021.

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
2. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:1–13
3. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160(1):106
4. Fukushima K (1988) Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1(2):119–130
5. Ide H, Kurita T (2017) Improvement of learning for CNN with ReLU activation by sparse regularization. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 2684–2691.
6. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
7. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
8. Ribaric S, Ariyaeeinia A, Pavesic N (2016) De-identification for privacy protection in multimedia content: a survey. Signal Process Image Commun 47:131–151
9. Meden B, Rot P, Terhörst P, Damer N, Kuijper A, Scheirer WJ et al (2021) Privacy–enhancing face biometrics: a comprehensive survey. IEEE Trans Inf Forensics Secur 16:4147–4183
10. Oh SJ, Benenson R, Fritz M, Schiele B (2016) Faceless person recognition: privacy implications in social media. European Conference on Computer Vision. Springer, Cham, pp 19–35
11. McPherson R, Shokri R, Shmatikov V (2016) Defeating image obfuscation with deep learning. arXiv preprint arXiv:1609.00408
12. Newton EM, Sweeney L, Malin B (2005) Preserving privacy by de-identifying face images. IEEE Trans Knowl Data Eng 17(2):232–243
13. Gross R, Sweeney L, De la Torre F, Baker S (2006) Model-based face de-identification. In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). IEEE, pp. 161–161
14. Gross R, Airoldi E, Malin B, Sweeney L (2005) Integrating utility into face de-identification. International Workshop on Privacy Enhancing Technologies. Springer, Berlin, pp 227–242
15. Sweeney L (2002) k-anonymity: a model for protecting privacy. Int J Uncertain Fuzziness Knowl Based Syst 10(05):557–570
16. Cootes T, Edwards G, Taylor C (2001) Robust real-time periodic motion detection, analysis, and applications. IEEE Trans Patt Analy Mach Intell 23(6):681–685

17. Du L, Yi M, Blasch E, Ling H (2014) GARP-face: Balancing privacy protection and utility preservation in face de-identification. In: IEEE International Joint Conference on Biometrics. IEEE, pp. 1–8
18. Jourabloo A, Yin X, Liu X (2015) Attribute preserved face de-identification. In: 2015 International Conference on Biometrics (ICB). IEEE, pp. 278–285
19. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al (2014) Generative adversarial nets. In: Advances in neural information processing systems, vol 27, pp 2672–2680
20. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. Preprint at arXiv:1511.06434
21. Chen J, Konrad J, Ishwar P (2018) Vgan-based image representation learning for privacy-preserving facial expression recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1570–1579
22. Ren Z, Lee YJ, Ryoo MS (2018) Learning to anonymize faces for privacy preserving action detection. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 620–636
23. Li T, Lin L (2019) Anonymousnet: natural face de-identification with measurable privacy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops
24. Wang HP, Orekondy T, Fritz M (2021) Infoscrub: towards attribute privacy by targeted obfuscation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3281–3289
25. Hukkelås H, Mester R, Lindseth F (2019) Deepprivacy: a generative adversarial network for face anonymization. International symposium on visual computing. Springer, Cham, pp 565–578
26. Sun Q, Ma L, Oh SJ, Van Gool L, Schiele B, Fritz M (2018) Natural and effective obfuscation by head inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5050–5059
27. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802
28. Wen Y, Song L, Liu B, Ding M, Xie R (2021) Identitydp: Differential private identification protection for face images. Preprint at arXiv:2103.01745
29. Cho D, Lee JH, Suh IH (2020) CLEANIR: controllable attribute-preserving natural identity remover. Appl Sci 10(3):1120
30. Zhan F, Zhu H, Lu S (2019) Spatial fusion gan for image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3653–3662
31. Kuang Z, Liu H, Yu J, Tian A, Wang L, Fan J, Babaguchi N (2021) Effective de-identification generative adversarial network for face anonymization. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 3182–3191
32. Zhang X, Wang X, Shi C, Yan Z, Li X, Kong B, Mumtaz I (2022) De-gan: domain embedded gan for high quality face image inpainting. Pattern Recogn 124:108415
33. Li Y, Lu Q, Tao Q, Zhao X, Yu Y (2021) SF-GAN: face de-identification method without losing facial attribute information. IEEE Signal Process Lett 28:1345–1349
34. Mirza M, Osindero S (2014) Conditional generative adversarial nets. Preprint at arXiv:1411.1784
35. Murez Z, Kolouri S, Kriegman D, Ramamoorthi R, Kim K (2018) Image to image translation for domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4500–4509
36. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134
37. Demir U, Unal G (2018) Patch-based image inpainting with generative adversarial networks. Preprint at arXiv:1803.07422
38. Berthelot D, Schumm T, Metz L (2017) Began: Boundary equilibrium generative adversarial networks. Preprint at arXiv:1703.10717
39. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410
40. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232

41. Zhang Z, Luo P, Loy CC, Tang X (2014) Facial landmark detection by deep multi-task learning. European Conference on Computer Vision. Springer, Cham, pp 94–108
42. Nguyen TT, Nguyen QVH, Nguyen CM, Nguyen D, Nguyen DT, Nahavandi S (2019) Deep learning for deepfakes creation and detection: a survey. Preprint at arXiv:1909.11573.
43. Liu Y, Jourabloo A, Ren W, Liu X (2017) Dense face alignment. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1619–1628
44. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
45. Shu C, Ding X, Fang C (2011) Histogram of the oriented gradient for face recognition. Tsinghua Sci Technol 16(2):216–224
46. Culjak I, Abram D, Pribanic T, Dzapo H, Cifrek M (2012) A brief introduction to OpenCV. In: 2012 Proceedings of the 35th International Convention MIPRO. IEEE, pp. 1725–1730
47. Liu Z, Luo P, Wang X, Tang X (2018) Large-scale celebfaces attributes (celeba) dataset. Retrieved August 15(2018):11
48. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 234–241
49. Li J, Wang Y, Wang C, Tai Y, Qian J, Yang J, Huang F (2019) DSFD: dual shot face detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5060–5069

Authors and Affiliations

Sung-su Jang[1] · Cheol-jin Kim[1] · Seong-yeon Hwang[1] · Myung-jae Lee[1] · Young-guk Ha[1]

Sung-su Jang
pik1100@naver.com

Cheol-jin Kim
cjfwls1070@naver.com

Seong-yeon Hwang
wiw100@naver.com

Myung-jae Lee
dualespresso@naver.com

[1] Department of Computer Science and Engineering, Konkuk University, Neungdong-ro, Gwangjin-gu, Seoul 143-701, Korea