Simulation of Facial Palsy Using Cycle GAN with Skip-Layer Excitation Module and Self-Supervised Discriminator

Takato Sakai[1,*], Masataka Seo[2], Naoki Matsushiro[3], and Yen-Wei Chen[1]

[1] Graduate School of Information and Engineering Ritsumeikan University, Shiga, Japan; Email: chen@is.ritsumei.ac.jp (Y.W.C.)
[2] Osaka Institute of Technology, Osaka, Japan; Email: masataka@oit.ac.jp (M.S.)
[3] Osaka Police Hospital, Osaka, Japan; Email: ent@opr.gr.jp (N.M.)
*Correspondence: is0410se@ed.ritsumei.ac.jp (T.S.)

***Abstract*—The Yanagihara method is used to evaluate facial nerve palsy based on visual examinations by physicians. Examples of scored images are important for educational purposes and as references, however, due to patient privacy concern, actual facial images of real patients cannot be used for educational purposes. In this paper, we propose a solution to this problem by generating facial images of a virtual patient with facial nerve palsy, that can be shared and utilized by physicians. To reproduce the patient facial expression in a public face image, we propose a method to generate a swapped face image using the improved Cycle Generative Adversarial Networks (Cycle GAN) with a skip-layer excitation module and a self-supervised discriminator. Experimental results demonstrate that the proposed model can generate more coherent swapped faces that are similar to the public face identity and patient facial expressions. The proposed method also improves the quality of generated swapped face images while keeping them identical to the source (genuine) face image.**

***Keywords*—facial nerve palsy, deep learning, Generative Adversarial Networks (GAN), Faceswap, few-shot image generation**

I. Introduction

Facial nerve palsy is a disease that paralyzes the nerves of the facial muscles, preventing voluntary movement of certain parts of the face. According to study, there are 20–30 cases of facial nerve palsy for every 100,000 people [1]. The diagnosis of facial nerve palsy necessitates adequate treatment depending on the degree of paralysis as it has the chance for long-term damage. In Japan, the most common method used to assess the degree of paralysis is the 40-point Yanagihara method [2]. In this method, the paralysis score is assessed by the expression of facial muscles based on 10 different facial expressions, including asymmetry when at rest and nine other facial expressions. Fig. 1 shows the 10 facial expressions used for evaluation in the Yanagihara method. However, this evaluation method is based on visual examination, results vary from doctor to doctor, especially for young doctors [3, 4]. To solve this problem, it is desirable to develop educational resources that might act as a specific indicator to unify evaluation standards among doctors. However, due to privacy concerns regarding patient person data, actual (genuine) facial images of patients cannot be used in academic conferences or medical training.

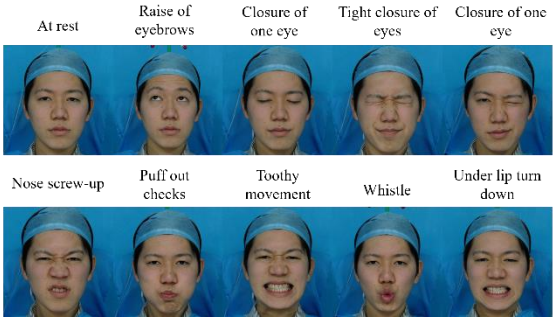

Figure 1. The 10 facial expressions of the Yanagihara method.

To solve this privacy problem, we employed Conditional Generative Adversarial Networks (CGAN) to generate virtual facial images that are similar to the expressions of patients with facial palsy in our previous works [5, 6]. In our previous methods, we extracted facial landmark points from the patient and used them as facial expression information to generate virtual facial nerve palsy patient faces. Due to the lack of patient data, the results were unsatisfactory and inconsistent with the patient's identity and facial expression. To this end, in this paper, we propose an improved Cycle GAN with a skip-layer excitation module (SLE module) and a self-supervised discriminator to generate a virtual facial image with accurate expression.

II. Related Work

In this study, we focus on the task of translating a public facial expression to a facial nerve palsy facial expression

Manuscript received December 3, 2022; revised March 30, 2023; accepted May 10, 2023.

(patients). We generate a corresponding image using only a small size of training data. Recently, some studies have focused on few-shot learning and translating facial expressions.

A. Star GAN

The Star GAN [7] method realizes multiple expression translation in a single model. The Star GAN employs a generator and a discriminator with two inputs, i.e., an image and a domain, to realize translation to multiple domains. However, this method can only translate to a specific domain. In contrast, the facial expressions reproduced by the proposed method are unique to each patient. Thus, the Star GAN method is not suitable for representing patient's facial expression in specific domains.

B. Multi cGAN

Previously, we have proposed the multi-conditional GAN (MC-GAN) [5, 6], which uses two images, the public face and the image with lines connecting facial landmark points extracted from the patient, to generate a public face image with the same facial expression as the patient. The MC-GAN model extracts 68 landmark points from the patient and connects them with lines for each facial organ using this image as facial expression information. Thus, it is possible to reproduce the patient's unique facial expressions. However, the MC-GAN model frequently generates an unnatural face due to the differences in face shape between the patient and the public face. The model also generates faces that do not have the same facial expressions as the patient due to the lack of patient data.

C. Faceswap

Faceswap [8] is a face manipulation method that swaps the face identity between two people using deep learning models. The swapped face has the source face and attributes (e.g., pose, expression, lighting, and background) from the target image. Here, two autoencoders are employed to generate the swapped images. In this method, all information about the face, e.g., shape, facial expression, and identity, is quantified by the model to generate images. Thus, this image generation process is not affected by differences in face shape and can generate natural faces even when there is a difference between public face and patient in facial shape. However, the Faceswap model cannot generate high-fidelity swapped faces with small amounts of training data. In a previous study [9], we proposed an improved version of the Faceswap model that can generate swapped faces with the same identity as the public face by introducing two discriminators. The first discriminator determines if the input image is a real or fake image, and the second discriminator, i.e., the identity discriminator, determines whether the input pair of images show the same or different people. These discriminators are trained using adversarial learning to the mock the generator model. However, that model could not generate swapped faces with the same facial expression as the patient.

D. Cycle GAN

Cycle GAN [10] is a generative model that translates the style of an image. Previously introduced image-to-image translation models, such as Pix2Pix [11], require paired images as training data. However, Cycle GAN [10] does not require paired images for model training. Cycle GAN implements two generators and two discriminators for translating the style between A and B. The image style is translated from A to B by one of the incorporated Generators, and from B to A by the other Generator. While one discriminator decides if the style A is there or not, the second discriminator decides on the style B. Using an adversarial learning strategy that enforces the translation between the two styles (A and B), the included Generator and Discriminator networks are trained. Additionally, a cycle consistency loss is introduced to ensure that the translated image must generate images that are identical to the original given image. This constrained (cycle consistency loss) helps in converting the given image from style A to style B and back again from style B to style A. This technique is used in our proposed Faceswap to generate high-fidelity and consistent swapped facial nerve palsy expression images.

E. Fast GAN

Liu *et al.* proposed the lightweight Fast GAN structure [12] to handle the few-shot image generation task with minimum computing cost. Fast GAN [12] includes a slip-layer channel-wise excitation module (SLE module) and a self-supervised discriminator that has been trained as a feature encoder. This model converges from scratch with just a few hours of training on a single GPU, and it demonstrates consistent performance, even when training is performed with fewer than 100 training samples. Thus, the proposed method incorporates the Fast GAN [12] model structure to improve the performance of our model.

III. Proposed Method

In this study, we propose an improved Cycle GAN with the SLE module and self-supervised discriminator to generate a virtual facial image with accurate facial expression as the patient. The proposed method is an improved version of Faceswap [8]. Fig. 2 shows a conceptual diagram of Faceswap. In the learning phase, we first implement autoencoders G_A and G_B, which share the encoder component. Here G_A learns the public face reconstruction, and G_B learns the patient face reconstruction. In the swapping phase, the trained models are used to generate the swapped face. Note that we only use the trained encoder and trained G_A decoder as the swapping generator to swap the patient's face to a public face. The patient face is input to the swapping generator, and the output image is the swapped face, i.e., the public face with the same facial expression as the patient. As a result, we can obtain a facial image of a virtual facial nerve palsy patient.

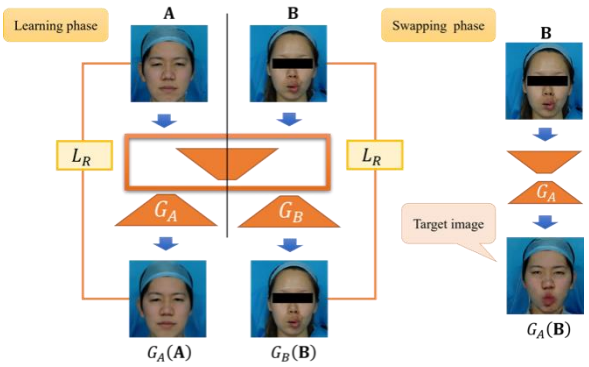

Figure 2. Conceptual diagram of Faceswap [8].

In this study, we introduce cycle consistency loss in addition to the previous research [9], and we attempt to generate high-fidelity swapped faces, especially for facial expressions. Fig. 3 shows the training processes of the previous model [9] and the proposed model. In addition, we introduce the SLE module to the generator and the self-supervised discriminator to facilitate stable learning. Fig. 4 shows the model structure of the generator in the previous model [9] and the generator with the SLE module in the proposed model. Fig. 5 shows the model structure of the discriminator in the previous method [9] and the self-supervised discriminator in the proposed method. In the following, we describe the Cycle GAN [10] and few-shot image generation methods, the SLE module, and the self-supervised discriminator.

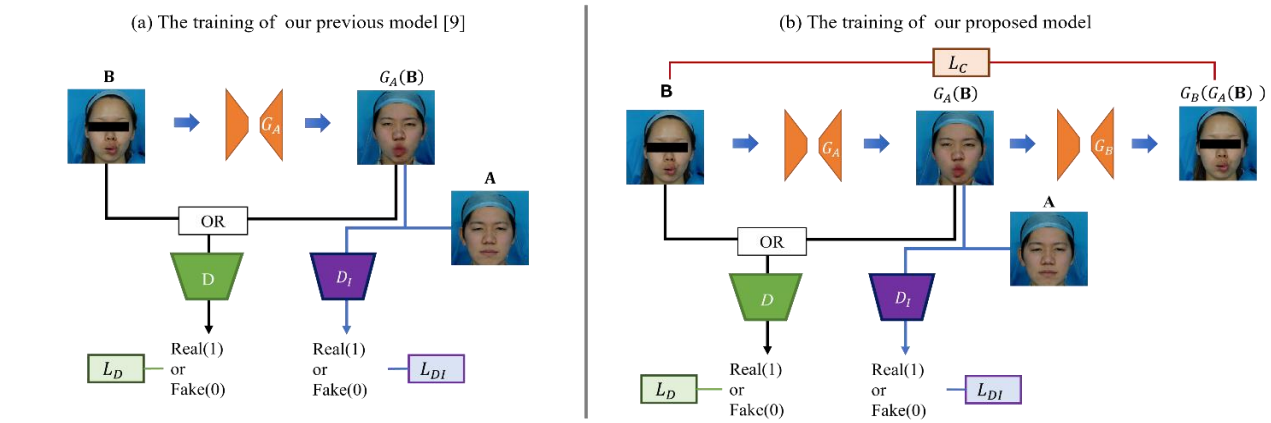

Figure 3. Training procedures of the (a) previous model [9] and (b) proposed model.

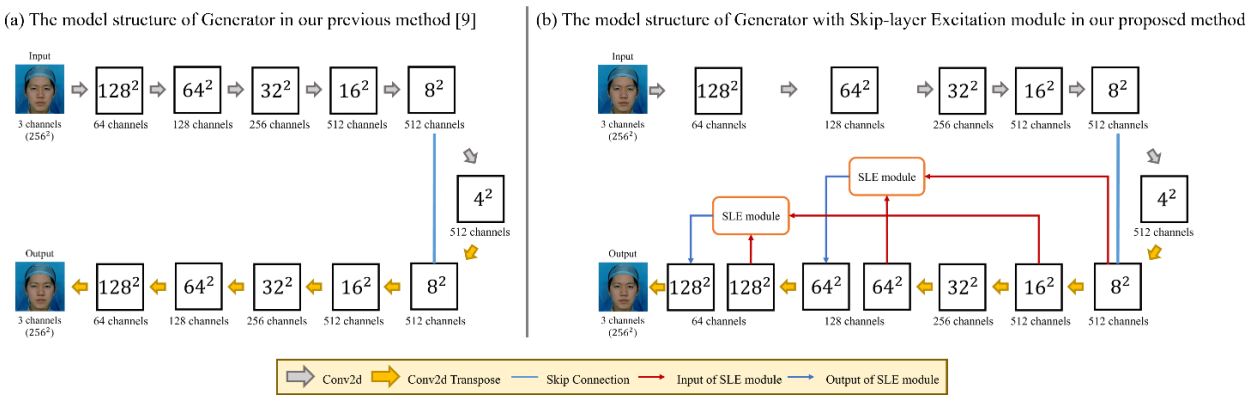

Figure 4. Model structure of (a) the generator in the previous method [9] and (b) the generator with the SLE module in the proposed method.

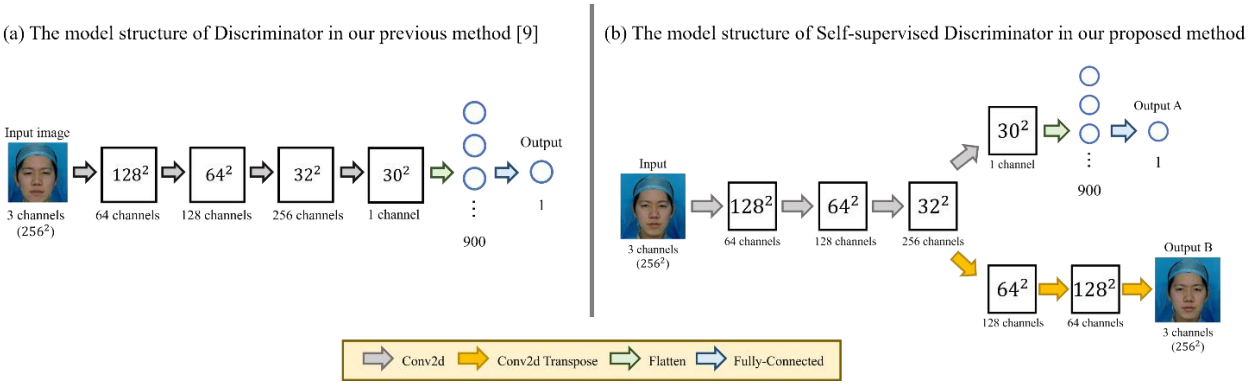

Figure 5. Model structure of (a) the discriminator in the previous method [9] and (b) the self-supervised discriminator in the proposed method.

A. Cycle GAN or Faceswap

In the proposed method, we impose an additional constraint on the previous model [9]. The model implements both a GAN structure and cycle consistency loss. Thus, we refer to the proposed model as the Cycle GAN [10].

In the proposed model, discriminator (D) is trained to discriminate the training image as real (1) and the image generated by the autoencoder as fake (0). By introducing D, we attempt to generate a swapped face that is clearer and more realistic.

The identity discriminator (D_I) is trained to determine whether the input pair of face images are the same person. Fig. 6 shows the model structure of the identity discriminator. Here, D_I outputs real (1) if they are the same person or fake (0) if they are different people. Note that D_I is trained using the same person pair (real) and another person pair (fake). In addition, D_I is trained to be adversarial with G_A, G_B Specifically, D_I tries to discriminate that swapped face G_A (**B**) and public face pair is fake. By introducing D_I, we attempt to generate a swapped face that has the same identity as the public face.

Cycle consistency loss is implemented in the proposed model. Here, if we use the swapped face G_A (**B**) generated by inputting the patient face into the G_A as the input to the G_B, the generated face G_B (G_A (**B**)) should be the original patient face. Using this property, the difference between the input image **B** and the G_B (G_A (**B**)) is added to the model's loss function. We also obtain the difference when we input public face **A**. By introducing this constraint, the swapped image G_A (**B**) is likely to have the same expression as the original patient face **B**, and we can generate a high-fidelity swapped face. Fig. 3(b) shows a conceptual diagram of the training process employed in the proposed model.

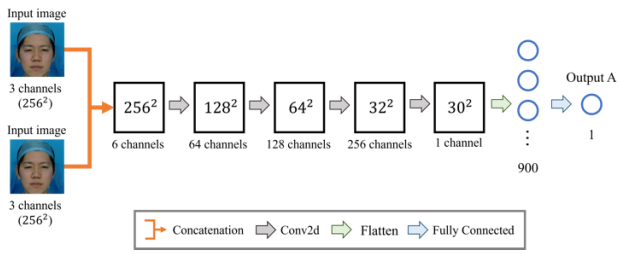

Figure 6. Model structure of identity discriminator.

We utilize adversarial training in the proposed model. First, we define the reconstruction loss L_R as follows:

$$L_R = \mathbb{E}_{\mathbf{A},\mathbf{B}\sim p_{data(\mathbf{A},\mathbf{B})}}[|\mathbf{A} - G_A(\mathbf{A})| + |\mathbf{B} - G_B(\mathbf{B})|] \quad (1)$$

where $\mathbb{E}$ represents expectation. Second, we train the proposed model to generate swapped faces that D and D_I discriminate as real. The adversarial loss with D is expressed as follows:

$$L_D = \mathbb{E}_{\mathbf{A},\mathbf{B}\sim p_{data(\mathbf{A},\mathbf{B})}}[\log|D(G_B(\mathbf{A}))| + \log|G_A(\mathbf{B})|] \quad (2)$$

where D (•) is the output of the discriminator and represents a real or fake value of 0 or 1, respectively, regarding the realism of the image. The adversarial loss with D_I is expressed as follows:

$$L_{DI} = \mathbb{E}_{\mathbf{A},\mathbf{B}\sim p_{data(\mathbf{A},\mathbf{B})}}[\log|D_I(\mathbf{A}, G_A(\mathbf{B}))| + \log|D_I(\mathbf{B}, G_B(\mathbf{A}))|] \quad (3)$$

where D_I (•,•) is the output of the identity discriminator and represents a real or fake value of 0 or 1, respectively, regarding the identity similarity of the images. The cycle consistency loss is expressed as follows:

$$L_C = \mathbb{E}_{\mathbf{A},\mathbf{B}\sim p_{data(\mathbf{A},\mathbf{B})}}[|\mathbf{A} - G_A(G_B(\mathbf{A}))| + |\mathbf{B} - G_B(G_A(\mathbf{B}))|] \quad (4)$$

Finally, the proposed model is trained with the sum of the above losses, which is expressed as follows:

$$L_G = \lambda L_R + L_D + L_{DI} + \lambda L_C \quad (5)$$

where λ is a hyperparameter that is set to 50.

B. SLE Module and Self-Supervised Discriminator

Due to the lack of training data, leaning GANs is unstable, and the gradient vanishing problem occurs. Thus, we propose an improved model for few-shot image generation. Here, we implement the SLE module and the self-supervised discriminator [12] in the proposed model to realize stable learning and allow the model to efficiently utilize features in the image generation process.

The SLE module stabilizes deep model training for high-resolution image generation with only a small increase in computational costs. Fig. 7 shows the model structure of the SLE module, and Fig. 4(b) shows the model structure of the generator, which includes the SLE module. The SLE module has two unique designs. First, it implements skip connections at low computational cost by performing channel-wise multiplication with other level features compared to element-wise addition, e.g., a Residual Block. Second, the SLE module performs convolution such that low-level features are the same as the channels of the high-level features and multiplies them, as shown in Fig. 7. In existing GAN models, skip connections are only used within the same resolution. In contrast, the SLE module performs skip connections between resolutions with a much larger range (e.g., 8^2 and 64^2, and 16^2 and 128^2) because equal spatial dimensions are no longer required. As a result, the proposed model functions like a self-attention mechanism. Thus, the proposed model can generate images efficiently from the extracted feature maps, and the gradient vanishing problem is avoided by connecting the high-level layer's gradient to the low-level layer.

The self-supervised discriminator is employed to reconstruct the input image. Fig. 5(b) shows the model structure of the self-supervised discriminator. This discriminator performs real/fake discrimination and image reconstruction when a real image is input, and only real/fake discrimination when a fake image is input. Note that the feature map extracted by this discriminator can be used for real/fake discrimination and image reconstruction. Thus, we expect the feature map to contain all relevant image information. As a result, we can prevent the feature

maps extracted by D from becoming sparse, and we can stabilize the learning for GAN.

The loss function of the self-supervised discriminator is expressed as follows:

$$L_{SSD} = \mathbb{E}_{\mathbf{A},\mathbf{B}\sim p_{data(\mathbf{A},\mathbf{B})}}\Big[\log|D(\mathbf{A})| + \log|D(\mathbf{B})| + \log\big|1 - D\big(G_B(\mathbf{A})\big)\big| + \log|1 - G_A(\mathbf{B})| + \big|\mathbf{A} - \mathcal{G}\big(\mathbf{f}(\mathbf{A})\big)\big| + \Big|\big|\mathbf{B} - \mathcal{G}\big(\mathbf{f}(\mathbf{B})\big)\big|\Big|\Big] \tag{6}$$

where $\mathbf{f}(\bullet)$ is the intermediate feature maps from the discriminator, and function $\mathcal{G}(\bullet)$ contains the processing on $\mathbf{f}$ and the decoder.

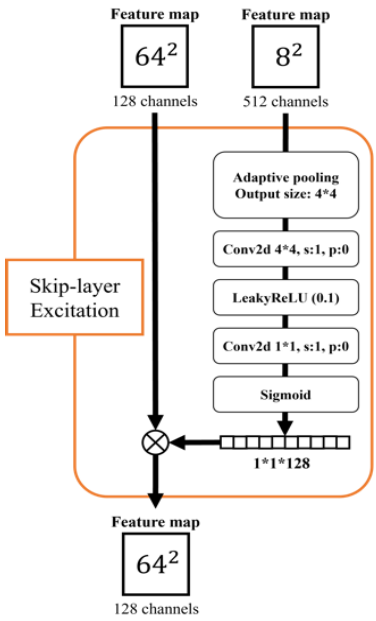

Figure 7. Model structure of the SLE module.

IV. Experiment

A. Implementation

In our experiment, we used facial expression images for two people (one female patient, and one male patient) provided by the Osaka Police Hospital. The dataset contained 2 (patients) × 10 facial expressions, with 60 images for each facial expression; thus, the dataset contained a total of 1,200 original images. The dataset was cropped around the face landmark points and normalized to pixel values of −1 to 1. Here, all images were resized to 256 × 256 pixels. We also augmented the dataset (Section IV-B). Finally, we used a total of 1,200 images. The proposed models were trained using this dataset, the number of epochs was set to 450, and the batch size was set to 30. In addition, the Adam optimizer was employed, where attenuation parameter beta1 was set to 0.5, parameter beta2 was set to 0.999, and the leaning rate alpha was set to 0.0002. Here, we used Python as the programming language, and TensorFlow was employed as the framework.

B. Data Augmentation

The experimental dataset was augmented by applying left-to-right flipping, random cropping, and noise to the images. Note that these data augmentation techniques were applied at the beginning of each epoch to address the lack of data. Specifically, left-to-right flipping was applied at a 50% probability, random cropping was applied with magnification based on a normal distribution, and three cannels noises based on normal distribution at a probability of 80%. Fig. 8 illustrates the data augmentation process. Note that the identity, structure and nature of the original data are not substantially changed by these operations in this case. Therefore, we believe that the outcomes of the Cycle GAN are unaffected by these data augmentation techniques. As a result, the proposed model was trained on a dataset with many variations.

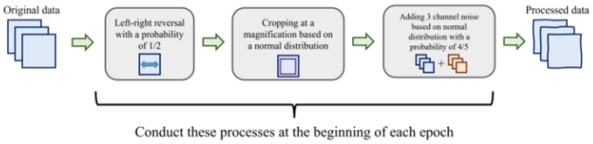

Figure 8. Data augmentation process.

C. Experimental Results

As shown in Fig. 9, we compared the proposed model with the original Faceswap model, our previous model [9] (i.e., the model with D and D_I), and the model with D, D_I, and cycle loss. Figure 9 shows (from left-to-right), the public face, the input image (i.e., the patient's face), the results of the original model, the results of the model with D, D_I, the results of the model with D, D_I, and cycle loss, and the results of the proposed model. Figs. 10–12 show the values of the generator loss and discriminator loss of the proposed model. As shown in Fig. 9, the proposed model generated the highest fidelity swapped images that have same identity as the public face and the same facial expression as the patient. Note that the original Faceswap model generated images that cannot be identified as the corresponding public face. In addition, our previous model and the model with D and D_I generated swapped faces that have the same identity as the public face but different facial expressions. The proposed model and the model with D, D_I, and cycle loss generated swapped faces that have mostly the same facial expression as the patient for other models. In addition, the proposed model generated swapped faces with similar facial expressions as the patient more than the model with D, D_I, and cycle loss because the proposed model efficiently extracted feature maps even with a small amount of training data due to the implementation of the SLE module and self-supervised discriminator. Fig. 13 shows the simulation results for all facial expressions in the Yanagihara method. As can be seen, the proposed model generated swapped faces with the same expressions as the patient, including the degree of paralysis, in all facial expressions. In terms of the patient face, we could not accurately observe whether the facial expression was the same or not because the patient face must be blurred. Thus, we simulated facial

expressions between public faces using the proposed model. Fig. 14 shows the simulation results for all facial expressions of the Yanagihara method. As can be seen, the proposed model can mostly simulate the facial expressions. However, the details of the facial expressions in the swapped face are not entirely the same as those of the patient face, e.g., the eyes and teeth (Fig. 13: closure of one eye and toothy movement), and artifacts are evident in some areas of the swapped face (Fig. 14: raised eyebrows).

Two user studies were also conducted to evaluate the performance of the proposed model. Here, the users evaluated (1) whether the simulation results have the same facial expression as the input image and (2) whether the simulation results have the same identity as the public face. In each study unit, the patient face (i.e., the input image), the public face, and four swapped faces generated by the original Faceswap model, the model with D and D_I [9], the model with D, D_I, and cycle loss, and the proposed model were shown to the participants. We asked the participants to evaluate the swapped faces on a scale of 1–5 (5: completely the same; 1: completely different) for the above two points. Finally, we collected answers from 20 participants. Table I shows the mean and standard deviation results of the user studies. As shown, the proposed model largely surpassed other models in in terms of both facial expression and identity.

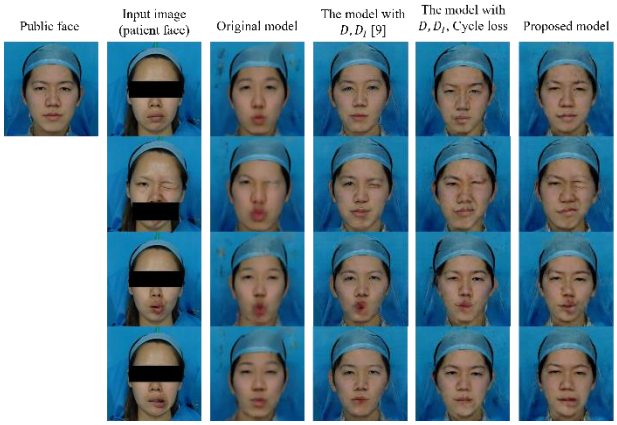

Figure 9. Comparison of results obtained by the original Faceswap model, the model with D and D_I [9], the model with D, D_I, and cycle loss, and the proposed model. There are four facial expressions from the Yanagihara method: (top to bottom): at rest, tight closure of eyes, whistle, and under lip turn down.

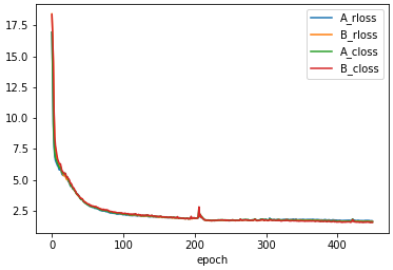

Figure 10. Values of generator reconstruction loss (A_rloss and B_rloss) and cycle loss (A_closs and B_closs). For example, A_rloss represents the reconstruction loss of G_A.

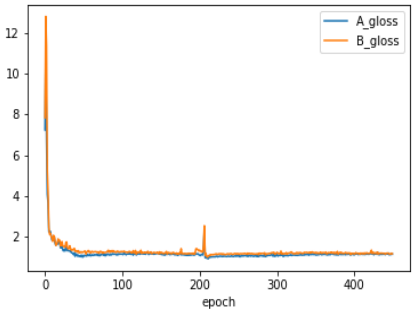

Figure 11. Values of generator adversarial loss (A_gloss is the adversarial loss of G_A, and B_gloss is the adversarial loss of G_B).

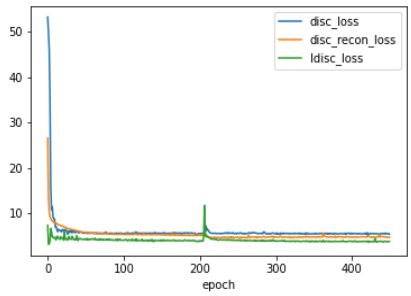

Figure 12. Values of discriminator loss (disc_loss is the self-supervised discriminator real/fake loss, disc_recon_loss is the self-supervised discriminator reconstruction loss, and Idisc_loss is the identity discriminator loss)

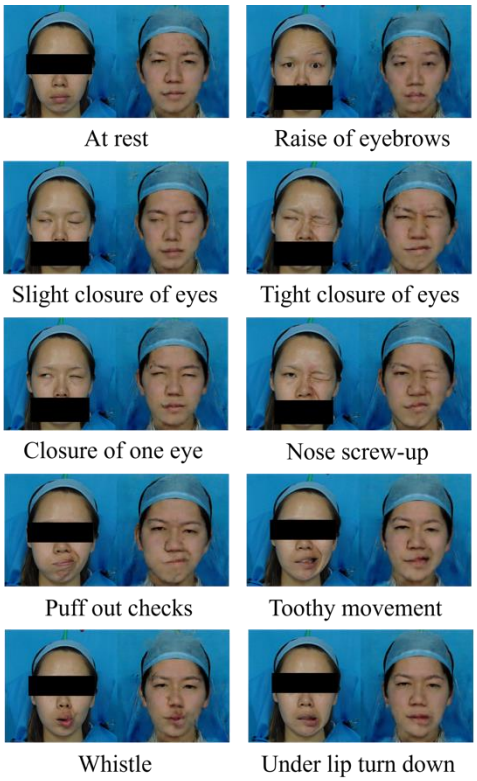

Figure 13. Simulation results for all facial expressions in the Yanagihara method. For each facial expression, the left image is the patient face, and the right image is the simulation result of the proposed model

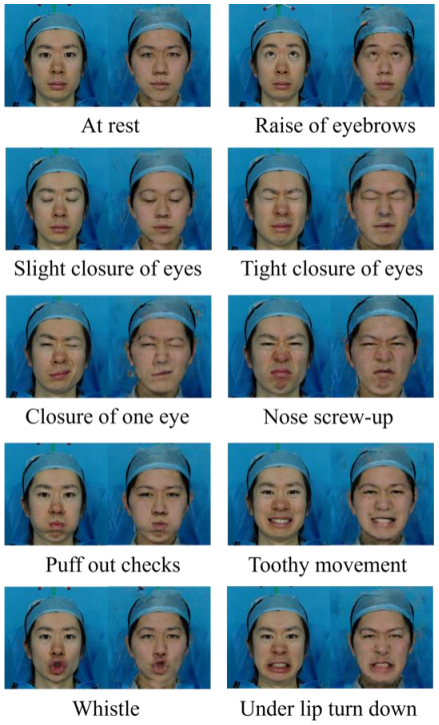

Figure 14. Simulation results for all facial expressions in the Yanagihara method between two public faces. For each facial expression, the left image is one public face, and the right image is the simulation result of the proposed model.

TABLE I. USER STUDY RESULTS (MEAN AND STANDARD DEVIATION)

	Original model	Model with D and D_I [9]	Model with D, D_I, and cycle loss	Proposed model
Facial expression	2.42/0.81	3.05/0.76	3.89/0.64	**4.68/0.46**
Identity	2.11/0.71	3.11/0.91	4.00/0.73	**4.47/0.50**

D. Discussion

Our investigation demonstrates that the proposed model generates highly realistic swapped faces, and the results shows improved accuracy for generating images of facial nerve palsy expressions that are also the same as the given input image. In our experiments, we discovered that the use of cycle loss is useful to enhance the quality of facial expression, generating more coherent and clearer images. Additionally, the outcomes demonstrate the usefulness of the SLE module and Self-Supervised Discriminator network in stabilizing the GAN training and enhancing the images generated. The SLE module enables the network to generate enhanced images using different resolutions by weighting the important channels, while the Self-Supervised Discriminator facilitates the network to extract key facial expression features from the entire image. However, in some cases, the patient's facial expression was only partially retained, which remains a major challenge to solve. This is because the public facial dataset does not contain any facial images of persons having the same facial expressions as the patient's paralyzed expression. We intend to use deformation region vectors of facial expressions using the techniques like optical flow to address these issues and increase the accuracy of swapping facial expressions. Some results show a few artifacts and less clarity in the face shape in comparison to a real face. The lack of the training data is the primary reason for these problems. A possible solution could be the use of state-of-the-art Super Resolution frameworks to enhance the visual quality of the generated images. Our proposed method has not encountered the overfitting issue despite the limited size of training data. The convergence of loss functions is shown in Figs. 10–12. These figures illustrate the utilization of Generator Reconstruction losses and Cycle Consistency loss are useful in converging the framework well. In our future work, we will increase the training dataset in order to enhance the performance of the model.

V. CONCLUSION

In this paper, we have proposed an improved version of the Cycle GAN with an SLE module and a self-supervised discriminator. The proposed model can generate high-fidelity swapped faces that have the same identity as the public face and the same facial expression as the patient. Cycle consistency loss is considered in the proposed model to keep the expression of the swapped face the same as that of the input face. In addition, the SLE module is implemented to allow the generator to efficiently extract feature maps and generate high-fidelity swapped faces. The self-supervised discriminator is implemented to prevent mode collapse and stabilize model learning. The experimental results demonstrate that the proposed model can generate higher fidelity swapped faces than the compared models. Finally, the swapped faces generated by the proposed model can be used to represent virtual facial paralysis patients that can be used in related academic conferences and medical training. Although, in this paper, we have performed face expression swapping between Japanese faces, our proposed model can be combined with other GAN-based models (such as Style GAN [13]) to perform face translation for other races, using available Japanese facial expression data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Takato Sakai, Masataka Seo, Naoki Matsushiro and Yen-Wei Chen conducted the research; Takato, Masataka Seo and Yen-Wei Chen developed the algorithm and system; Naoki Matsushiro collected dataset and provided supervision from the medical view point. Takato Sakai wrote the paper; Yen-Wei Chen and Masataka Seo revised the paper. All authors had approved the final version.

REFERENCES

[1] K. K. Adour, *et al.*, "The true nature of Bell's palsy: Analysis of 1000 consecutive patients," *Laryngoscope*, vol. 88, pp. 787–801, 1978.

[2] N. Yanagihara, *et al.*, "A study on the criteria for determining the degree of facial neuropathy," *Journal of the Japanese Society of Otolaryngology*, vol. 80, no. 8, pp. 799–805, 1997.
[3] N. Matsushiro, "Differences in the evaluation of facial palsy (Yanagihara grading system) among 52 ENT doctors at Osaka University," *Facial Nerve Research*, vol. 29, pp. 60–62, 2010.
[4] N. Matsushiro, "Differences in the evaluation of facial palsy (Yanagihara grading system) among 9 facial palsy specialists and 47 general ENT doctors in Japan: A 9 University collaborative investigation," *Facial Nerve Research*, vol. 29, pp. 63–65, 2010.
[5] S. Yaotome, M. Seo, N. Matsushiro, and Y.-W. Chen, "Simulation of facial palsy using conditional generative adversarial networks," in *Proc. of 2020 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, 2020, pp. 579–582.
[6] T. Sakai, M. Seo, N. Matsushiro, and Y.-W. Chen, "Simulation of facial palsy using conditional generative adversarial networks and face shape normalization," in *Proc. of 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, 2021, pp. 793–797.
[7] Y. Choi, *et al.*, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8789–8797, 2018.
[8] Faceswap. [Online]. Available: https://faceswap.dev/
[9] T. Sakai, M. Seo, N. Matsushiro, and Y.-W. Chen, "Simulation of facial palsy using an improved GAN with two discriminators," in *Proc. of 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*, 2022.
[10] J-Y Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. 2017 IEEE International Conference on Computer Vision (ICCV2017)*, pp. 2223–2232, 2017.
[11] P. Isola, *et al.*, "Image-to-image translation with conditional adversarial networks," in P*roc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017)*, pp. 1125–1134, 2017.
[12] B. Liu, *et al.* "Towards faster and stabilized gan training for high-fidelity few-shot image synthesis," in *Proc. of International Conference on Learning Representations (ICLR)*, 2021.
[13] T Karras, *et al.*, "A style-based generator architecture for generative adversarial networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition 2019 (CVPR2019)*, 2019, pp. 4401–4410.