Image translation between human face and *wayang orang* using U-GAT-IT

Ciara Nurdenara, Wikky Fawwaz Al Maki
Department of Informatics, School of Computing, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received May 27, 2022
Revised Mar 19, 2024
Accepted Mar 31, 2024

Keywords:

Dual contrastive learning generative adversarial network
Image translation
U-GAT-IT
Unpaired dataset
Wayang orang

ABSTRACT

The people's puppets (*wayang orang*) performance typically requires approximately one hour for the performers to assume the role of a *wayang orang*, as this duration is necessary to apply makeup and select suitable attire. One potential solution to this issue involves the creation of a computerized simulation that replicates the process of putting makeup and traditional clothing on the face and head of the *wayang orang* performer. The completion of this work can be achieved through the utilization of image translation techniques. The objective of this study is to employ the unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation (U-GAT-IT) technique to convert human faces into *wayang orang* representations. The study utilizes an unpaired dataset comprising 1216 training data samples and 240 testing data samples. The primary objective of this study is to effectively preserve both the background picture and the facial identification component inside the given input image. This study utilizes quantitative assessment methods, specifically kernel inception distance (KID), Frèchet inception distance (FID), and inception score (IS), to evaluate the quality of the generated output image produced by the generator. Experimental results demonstrated that U-GAT-IT outperforms dual contrastive learning generative adversarial network (DCLGAN) in terms of the metrics IS, FID, and KID.

Corresponding Author:

Wikky Fawwaz Al Maki
Department of Informatics, School of Computing, Telkom University
Telekomunikasi street, No. 1, Terusan Buah Batu, Dayeuhkolot, Bandung, Indonesia
Email: wikkyfawwaz@telkomuniversity.ac.id

1. INTRODUCTION

In this modern era, only a few people know traditional Indonesian culture, especially people's puppets (*wayang orang*). The *wayang orang* players took about an hour to be dressed in a proper form of *wayang orang* since it takes time to have makeup before *wayang orang* performances. Moreover, it is challenging to find the costume. Only a few people that had an experience became a *wayang orang*.

Headwear and make-up are important elements of *wayang orang* costumes. In *wayang orang* headwear, use a crown called irah iran. The *wayang orang* makeup is used on the eyebrows, lips, forehead, and sideburns. Image translation can help everyone to see themselves as *wayang orang*. Image to image translation generally aims to change the style or characteristics of the image from one domain to another one. Image translation can be performed by implementing the generative adversarial networks (GANs) architecture. GANs are neural networks used for unsupervised learning that has many different types of GAN implementations. There are several examples of GAN applications, i.e., video prediction, translation of a text into an image, and image to image translation. GANs [1], [2] have been widely used in representation learning [3]–[5], image generation [6], [7], and image editing [8]. Also, in image generation applications, such as image inpainting [9], text2image [10], and future prediction [11], GAN is employed and has demonstrated impressive results.

Research on the translation of wayang and human images has not existed before. However, the development of research similar to this research on image translation has been initiated by Gatys *et al.* [12]. The research is focused on art. The proposed algorithms are able to perform translation tasks from four different datasets, namely selfie2anime, horse2zebra, cat2dog, photo2portrait, and photo2vangogh [13]–[16].

The image-to-image translation can be performed by implementing unsupervised as well as supervised learning methods. In unsupervised learning, image pairs are independent (unrelated images). The unsupervised learning method aims to translate between different domains by using unlabeled images without establishing a link between images and minimizing the cost of labeled data. The approach that uses unsupervised learning is CycleGAN [16], unsupervised image-to-image translation (UNIT) [17], and DualGAN [18].

Image translation has wide applications such as image enhancement, style transfer, season transfer, and object transfiguration [16]. Isola *et al.* [19] firstly introduced GAN-based image translation. In this connection, when the generated image from the generator is adapting the input image, the process is called image translation. Zhu *et al.* [16] implemented the CycleGAN to solve the problem of translation of horse images into zebra images and obtained the Frèchet inception distance (FID) score of 89.7. Kim *et al.* [14] evaluated the performance of the unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation (U-GAT-IT) method by using the kernel inception distance (KID) evaluation metric. U-GAT-IT translated using many kinds of datasets such as selfie2anime, horse2zebra, cat2dog, and photo2portrait. However, from those image result did not maintain the person identity from the image. Therefore, this paper focus on keeping the identity of the person. The image translation of selfies into anime by implementing the U-GAT-IT method reaches the KID score of 11.67. These experimental results indicate that U-GAT-IT has the potential ability in performing an image-to-image translation task.

This paper implements U-GAT-IT [14] to perform image translation tasks. This study aims to add makeup, crowns, and accessories from *wayang orang* while maintaining the input image's identity, pose, and background features. The human face and *wayang orang* image datasets have never been used on the U-GAT-IT model. This research will add new tasks to the U-GAT-IT model, image translation of human to *wayang orang*.

2. PROPOSED METHOD

In this study, we adopt the U-GAT-IT architecture [14]. U-GAT-IT architecture has a pair generator and pair of discriminators. The model will translate an image from the human face (source domain) into a *wayang orang* (target domain). Then, the model translates it back to the human face (source domain). However, this study will only discuss human faces to the *wayang orang* image translation. The design of the system used to translate images of human faces with *wayang orang* is shown in Figure 1.

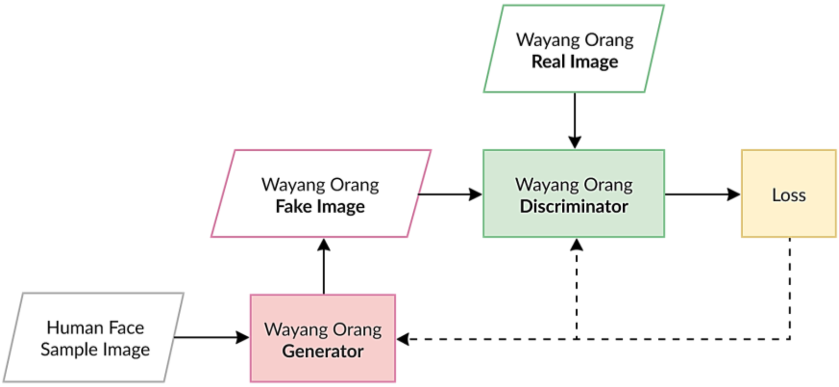

Figure 1. The flow of the image translation system

The flow of the system in Figure 1 begins with the human face sample image (source domain) as an input in the generator. The generator has a role in generating a *wayang orang* fake image (target domain).

Thenceforth, the discriminator will compare the *wayang orang* fake image with the *wayang orang* original image. Next, the loss function is calculated. Then, the loss function is updated to the generator network and also the discriminator. Adaptive moment estimation (ADAM) has a role in optimizing the model by updating the weights from the network in the training process. ADAM is relatively easy to configure where the configuration parameters work well in most cases. Hyperparameter configuration that is applied to the model is adversarial type GAN, ADAM optimizer, learning rate, beta1, beta2, identity weight, cycle weight, and class activation maps (CAM) weight with details can be seen in Table 1.

Table 1. Hyperparameter model

Optimizer	Beta 1	Beta 2	Identity weight	CAM weight	Cycle weight	Learning rate
ADAM	0,5	0,999	10	1000	10	0,0001

The generator of U-GAT-IT architecture can be seen in Figure 2. It consists of the encoder, auxiliary classifier, and decoder. The downsampling encoder has two convolution layers with two strides and one padding whereas the bottleneck encoder consists of four residual blocks. The auxiliary classifier section has the functionality to study the weights of the feature map from the origin domain using global max pooling and global average pooling. The last part is the decoder upsampling that consists of two convolution layers. Each encoder uses a normalization instance and AdaLIN [13] on the decoder. Furthermore, rectified linear units (ReLU) is used as an activation function in U-GAT-IT architecture.

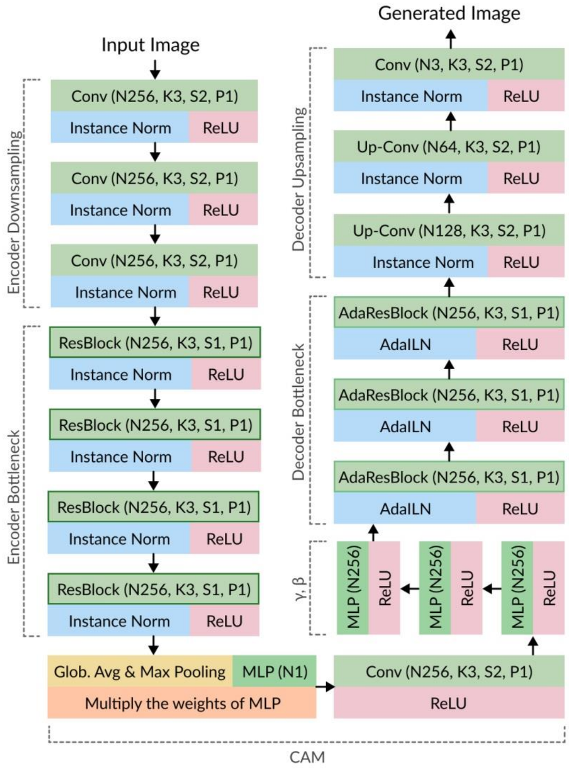

Figure 2. Generator architecture

Unlike the generator, the discriminator consists of the encoder, auxiliary classifier, and classifier. Each encoder uses instance normalization and activation function leaky ReLU. The auxiliary classifiers in the generator as well as in the discriminator are trained to distinguish the input image are generated or the target image. The discriminator uses patch-GAN [19] which classifies synthetic or original images with sizes 70×70 (local) and 286×286 (global). The flow of the discriminator network begins with the input of a sample image. Here, the discriminator network utilizes attention feature maps that use the weight of the encoded feature maps that the auxiliary classifier has trained. The discriminator uses spectral normalization (SN) [20] as a normalization function. The discriminator of U-GAT-IT architecture can be seen in Figure 3.

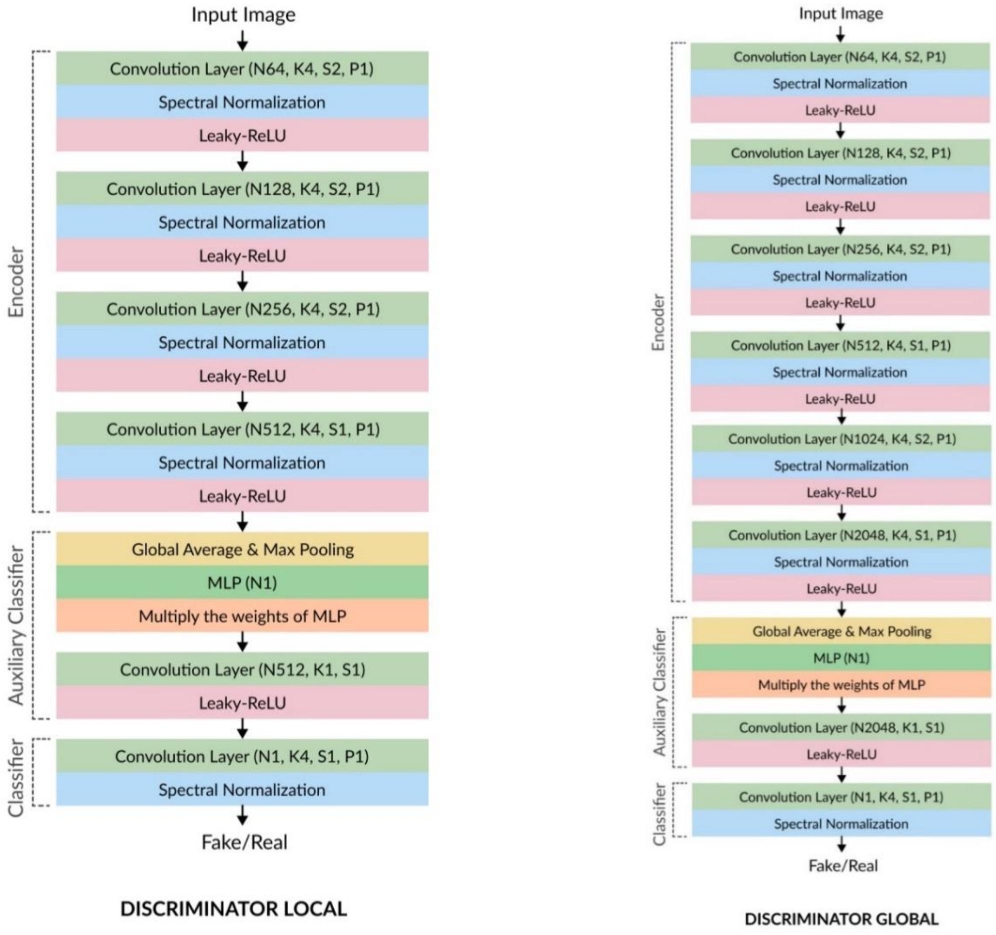
Input Image
Encoder
Convolution Layer (N64, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N128, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N256, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N512, K4, S1, P1)
Spectral Normalization
Leaky-ReLU
Auxiliary Classifier
Global Average & Max Pooling
MLP (N1)
Multiply the weights of MLP
Convolution Layer (N512, K1, S1)
Leaky-ReLU
Classifier
Convolution Layer (N1, K4, S1, P1)
Spectral Normalization
Fake/Real
DISCRIMINATOR LOCAL
Input Image
Encoder
Convolution Layer (N64, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N128, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N256, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N512, K4, S1, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N1024, K4, S2, P1)
Spectral Normalization
Leaky-ReLU
Convolution Layer (N2048, K4, S1, P1)
Spectral Normalization
Leaky-ReLU
Auxiliary Classifier
Global Average & Max Pooling
MLP (N1)
Multiply the weights of MLP
Convolution Layer (N2048, K1, S1)
Leaky-ReLU
Classifier
Convolution Layer (N1, K4, S1, P1)
Spectral Normalization
Fake/Real
DISCRIMINATOR GLOBAL

Figure 3. Discriminator architecture

3. RESEARCH METHOD

In our study, we used an unpaired dataset including human face images (source domain) and *wayang orang* images (target domain). The *wayang orang* image dataset was collected from Instagram by scraping image method. Meanwhile, the face images are using the dataset CelebA [21] mixed with Indonesian faces which can be seen in Figure 4. These images are merged so that the image dataset is varied and unbiased. The image used in this study is limited to the male gender only. The sample image can be seen in the image on Figure 4. The more the number of images and the more varied the image from the image, the better effect the model will obtain.

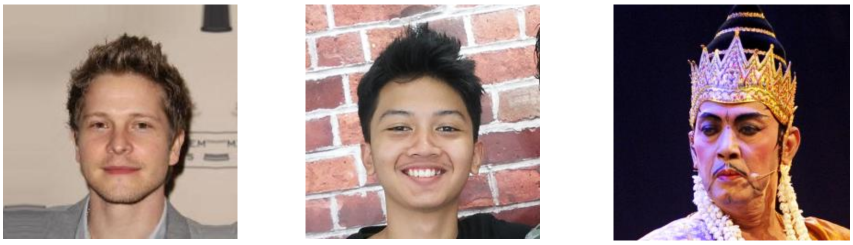

Figure 4. Sample images

The dataset that has been collected is divided into training data and testing data. The distribution of facial image datasets on training data and test data is achieved by using the Pareto principle with a ratio of 80:20. Therefore, the number of training data and test data are 976 and 240 images, respectively. Before performing the training phase, it is important to perform facial alignment on all images by rotating the image based on the key points of the face. In this study, the facial alignment process was carried out manually by equalizing the points on the eyes and mouth. Facial alignment aims to avoid changes in the face shape of the original image by maintaining the facial structure of the original image in the training process. Next, we apply image resizing, i.e., the input images are cropped to obtain square shape images.

After applying the proposed method, we evaluate the method by using three metrics. In this research, we employ inception score (IS), FID, and KID to evaluate our model. Salimans *et al.* [22] firstly introduced IS. IS is an alternative evaluation of human annotators in image quality, especially for the generated image of the GAN model. The value of the IS is based on various images (each *wayang orang* image is different from one another), and each generated image looks identical to the image target. The limit value of IS in the range [1, ∞]. The larger the IS value, the better the image quality. The IS is given in (1),

$$IS = exp(E_{w\sim z}H_{K\,L}(p(q|w)||p(q))) \tag{1}$$

where $w\sim z$ indicates that w is an image obtained by performing sampling to z, $H_{KL}(s||t)$ is the KL-divergence between the distributions s and t, $p(q||w)$ is the conditional class distribution, and $p(q)$ is the marginal class distribution.

FID can be used to evaluate image quality produced by the generator and to evaluate the performance of GAN. The data distribution is modeled by using multivariate Gaussian distribution with mean μ dan covariance Σ. Heusel *et al.* [23] evaluation using FID is better than that using the IS since FID evaluates the synthesized image data set compared to the original image set from the target domain. A smaller FID score represents a better image quality which can be seen from the generated image, i.e., the image is more similar to the target image. The FID score between the generated image *g* and target image *t* is calculated in (2),

$$FID\ (t,g) = \left\|\mu_t - \mu_g\right\|_2^2 + tr(\textstyle\sum_t + \sum_g - 2(\sum_t^{1/2} \sum_g \sum_t^{1/2})^{\frac{1}{2}}) \tag{2}$$

Where,
t =target image,
g =generated image,
μ=mean from image feature,
tr=trace of a matrix, and
Σ =covariance matrix from a vector of image feature.

KID evaluation [24] is similar to FID. FID and KID use a two-sample test variance in the studied "perceptual" feature space, the Inception pool3 space, to assess distributional fit. The difference between FID and KID is that KID calculates the maximum mean discrepancy (MMD) square between inception representations. KID has an advantage compared to FID, i.e., KID is an unbiased estimator.

4. RESULTS AND DISCUSSION

In this section, we provide our experimental results. We demonstrate the generated images obtained by employing our proposed model. Also, we provide the evaluation metrics and discussion based on our experimental results.

4.1. Image generation results

We employed five types of input images shown in Figure 5(a) to investigate the capability of dual contrastive learning generative adversarial network (DCLGAN) and U-GAT-IT in this task. The types of images used in our experiments are a regular man, a man wearing sunglasses, and a man wearing a mask. Also, we employed images that contain multiple persons. By employing the proposed model, we successfully generated *wayang orang* images from the dataset. Figures 5(b) and (c) show the images obtained from the testing phase after implementing DCLGAN and U-GAT-IT, respectively. These output images are obtained from the model by employing various sample pictures. In this connection, we investigate the generator output according to different sample pictures.

The generator offers the capability to generate a diverse range of images. One of the primary difficulties encountered in the process of converting a human face image into a *wayang orang* image lies in the task of modifying the image's style while simultaneously preserving the facial identity and background of the original domain image. The experimental results show that DCLGAN changes image texture and color to yellowish in the whole image, while in U-GAT-IT, the color is almost similar to that of the original image. U-GAT-IT and DCLGAN can generate a variety of images and prevent mode collapse on regular GAN issues. In the images generated by U-GAT-IT, the shape of the crown has been formed. Also, the makeup can be transferred properly without changing the content (identity and background) of the original image. In the case of images that contain a face wearing a mask, lips are not detected so no lips makeup is transferred. From Figure 5, it is clear that the U-GAT-IT model can also translate more than one face in a single image. Also, U-GAT-IT generates better images than DCLGAN does. This fact can be seen by comparing images in Figures 5(b) and (c). When DCLGAN is applied to perform the image-to-image translation tasks, the generated *wayang orang* images are not realistic enough. In addition, the blur artifact dominates the *wayang orang* images as shown in Figure 5(b).

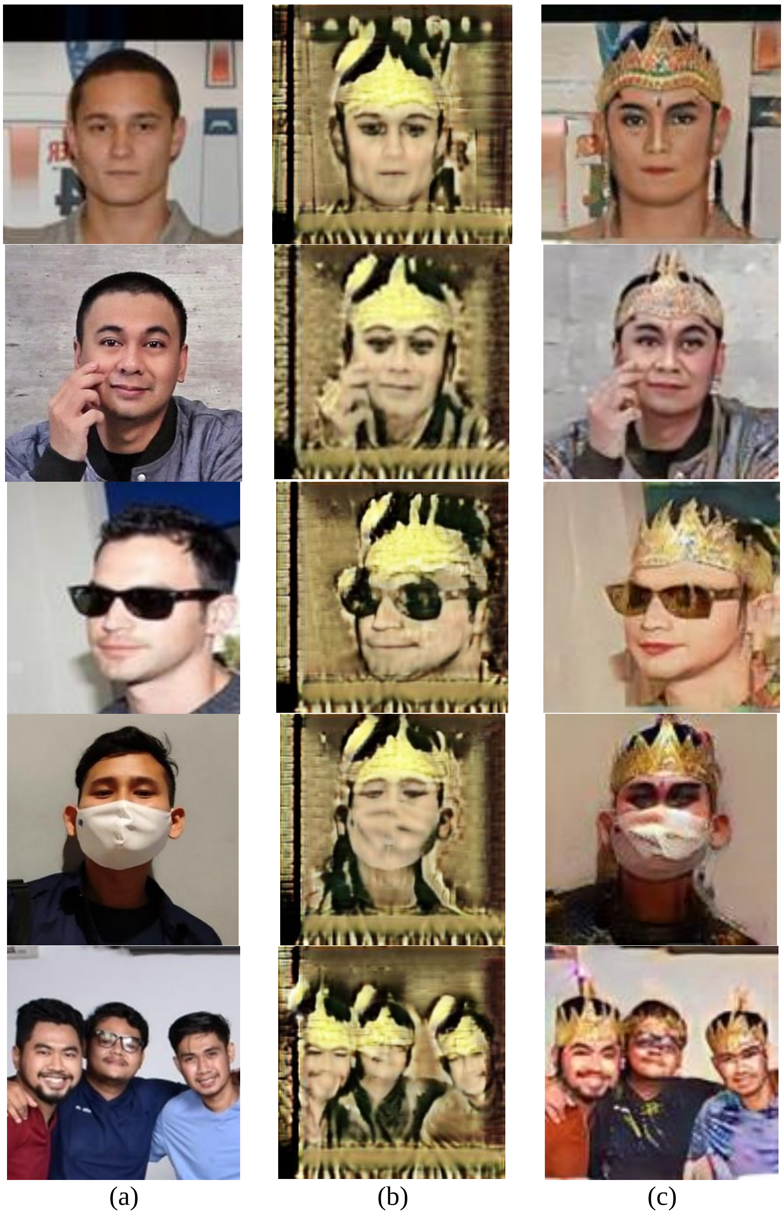

Figure 5 Comparing original images and generated images from both model according to different sample pictures: (a) input images, (b) generated images using DCLGAN, and (c) generated images using U-GAT-IT

4.2. Evaluation metrics

The testing process is carried out to find out the performance of the model. One of the important factors in evaluating image quality is based on how similar and realistic the resulting image is. Although blurry images can still look realistic, image sharpness is important to consider. Another important factor is that the generator must produce a variety of images. The challenge in translating facial images into *wayang orang* is to change the style of an image while maintaining the identity of the face and background of the original image.

Evaluating GAN is not enough by using only the loss function. Therefore, to measure the model's performance and the resulting image, the FID method was used as a quantitative evaluation using test data of 240 images. In this connection, the image that is used as the test data is never used in the training process.

The U-GAT-IT model is compared to the DCLGAN [25] model. These two models have a similar ability, i.e., unsupervised learning image-to-image translation. Moreover, DCLGAN and U-GAT-IT are able to translate various types of images. In this research, the DCLGAN model is treated similarly to the U-GAT-IT model, i.e., on training iteration and hyperparameters model configuration.

In our experiments, we compare our proposed model with another architecture. Since there has not existed any study about the image to image translation to generate *wayang orang* images, we could not compare our results with the results provided by the other researchers. Therefore, we compared the U-GAT-IT architecture and the DCLGAN architectures to the same dataset to demonstrate the performance of our proposed model. Table 2 shows the performance comparison between the U-GAT-IT and the DCLGAN in our study. Based on the results of visualization and evaluation, the U-GAT-IT model produces better results than those of the DCLGAN. From Table 2, it is clear that U-GAT-IT achieves the value of IS 2.414, FID 0.924, and KID 4.357 on DCLGAN. Here, the bigger the IS score indicates the better the output images.

Table 2. Architecture performance comparison

Model	IS	FID x 100	KID x 100
U-GAT-IT	2.414	0.924	4.357
DCLGAN	1.670	2.189	22.13

5. CONCLUSION

The process of this system begins with the collection of datasets. In the preprocessing process, it is better to perform facial alignment before conducting the training process. Based on the results and analysis of the experiments that have been carried out, the best output images are images of men where makeup and costumes are successfully added to their faces and heads, respectively. In addition, the resulting image retains the background as well as the facial identity of the original image. From the experimental results, it is clear that the U-GAT-IT model generates better-quality images. Also, the U-GAT-IT can produce images more similar to the *wayang orang* than the DCLGAN can. It is proven by performing an evaluation using the IS. The value of U-GAT-IT is 2.414 which is higher than DCLGAN. Moreover, FID and KID scores on the U-GAT-IT model have a smaller value than DCLGAN. FID and KID scores on U-GAT-IT are 0.924 and 4.357. Meanwhile, FID and KID scores on DCLGAN are 2.189 and 22.13. These results indicate that U-GAT-IT can perform well in translating human faces into *wayang orang*. There are some recommendations for future work. It is suggested to modify the architecture of U-GAT-IT to investigate the influence of various GAN architectures involved in U-GAT-IT. Another GAN architecture involved in U-GAT-IT may provide better results in generated images, IS, KID, and FID. In future works, we will improve the *wayang orang* images generated by the U-GAT-IT by investigating the best values of its parameters. Also, we will modify the U-GAT-IT architecture to improve the experimental results. Some GAN architectures are interesting to investigate as the image translation methods in generating *wayang orang* images.

ACKNOWLEDGEMENTS

Authors would like to express our sincere thanks to Wayang Kautaman that has provided the *wayang orang* images collection.

REFERENCES

[1] I. J. Goodfellow *et al.*, “Generative adversarial net,” in *Advances in Neural Information Processing Systems,* Red Hook, New York: Curran Associates, Inc., 2014.

[2] J. Zhao, M. Mathieu, and Y. LeCun, “Energy-based generative adversarial network,” in *5th International Conference on Learning Representations,* 2017, pp. 1–17.

[3] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016, pp. 1–16.

[4] T. Hinz, M. Fisher, O. Wang, and S. Wermter, “Improved techniques for training single-image GANs,” in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV),* IEEE, Jan. 2021, pp. 1299–1308. doi: 10.1109/WACV48630.2021.00134.

[5] M. Mathieu, J. Zhao, P. Sprechmann, A. Ramesh, and Y. Le Cun, “Disentangling factors of variation in deep representations using adversarial training,” in *Advances in Neural Information Processing Systems,* Red Hook, New York: Curran Associates, Inc., 2016, pp. 5047–5055.

[6] E. Denton, S. Chintala, A. Szlam, and R. Fergus, “Deep generative image models using a laplacian pyramid of adversarial networks,”

in *Advances in Neural Information Processing Systems*, vol. 2015-Janua, Red Hook, New York: Curran Associates, Inc., 2015, pp. 1486–1494.

[7] J. Li, J. Jia, and D. Xu, "Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks," in *2018 37th Chinese Control Conference (CCC)*, IEEE, Jul. 2018, pp. 9159–9163. doi: 10.23919/ChiCC.2018.8482813.

[8] J. Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Computer Vision – ECCV 2016*, Cham: Springer, 2016, pp. 597–613. doi: 10.1007/978-3-319-46454-1_36.

[9] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016, pp. 2536–2544. doi: 10.1109/CVPR.2016.278.

[10] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *33rd International Conference on Machine Learning, ICML 2016*, 2016, pp. 1060–1069.

[11] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016, pp. 1–14.

[12] L. Gatys, A. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," *Journal of Vision*, vol. 16, no. 12, pp. 1–16, 2016, doi: 10.1167/16.12.326.

[13] A. Gokaslan, V. Ramanujan, D. Ritchie, K. I. Kim, and J. Tompkin, "Improving shape deformation in unsupervised image-to image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 649–665.

[14] J. Kim, M. Kim, H. Kang, and K. H. Lee, "U-GAT-IT: Unsupervised Generative Attentional Networks With Adaptive Layer-Instance Normalization for Image-To-Image Translation," in *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[15] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *34th International Conference on Machine Learning, ICML 2017*, 2017, pp. 2941–2949.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.

[17] Ming-Yu Liu, Thomas Breuel, and Jan Kautz, "Unsupervised Image-to-Image Translation Networks," in *Advances in Neural Information Processing Systems*, Red Hook, New York: Curran Associates, Inc., 2017, pp. 700–708.

[18] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017, pp. 2868–2876. doi: 10.1109/ICCV.2017.310.

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632.

[20] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–26.

[21] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," in *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Dec. 2015, pp. 3730–3738. doi: 10.1109/ICCV.2015.425.

[22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Advances in Neural Information Processing Systems*, Red Hook, New York: Curran Associates, Inc., 2016, pp. 2234–2242.

[23] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Advances in Neural Information Processing Systems*, Red Hook, New York: Curran Associates, Inc., 2017, pp. 6627–6638. doi: 10.18034/ajase.v8i1.9.

[24] M. Binkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–36.

[25] J. Han, M. Shoeiby, L. Petersson, and M. A. Armin, "Dual Contrastive Learning for Unsupervised Image-to-Image Translation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2021, pp. 746–755. doi: 10.1109/CVPRW53098.2021.00084.

BIOGRAPHIES OF AUTHORS

Ciara Nurdenara was a student in Department of Informatics, School of Computing, Telkom University. She joined a study group at the High-Performance Computing (HPC) Lab and the RPL GDC Lab during her studies. She has been a member of the computer vision research group in Multimedia Lab for 2 years. Moreover, she has one year experience as data scientist. Her research interests include digital image processing, artificial intelligence, and computer vision. She can be contacted at email: ciarand@student.telkomuniversity.ac.id.

Wikky Fawwaz Al Maki received the Dr. Eng. degree in Integrated Science and Engineering from Ritsumeikan University, Japan in 2009. He also received his B.Eng. degree from University of Indonesia in 2004 and M.Eng. degree from Ritsumeikan University in 2007. He is currently a lecturer at School of Computing, Telkom University, Bandung, Indonesia. He is also the head of Multimedia Research Lab at Telkom University. His research areas are digital image processing, signal processing, stochastic systems, artificial intelligence, pattern recognition, and computer vision. He is affiliated with Indonesia Data Scientist Society as member. He has served as invited reviewer for several international conferences and journals. He can be contacted at email: wikkyfawwaz@telkomuniversity.ac.id.