Knowledge Distillation Generative Adversarial Network for Image-to-Image Translation

Chayanon Sub-r-pa and Rung-Ching Chen *

Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan
Email: t5220317@gm.cyut.edu.tw (C.S.); crching@cyut.edu.tw (R.-C.C.)
*Corresponding author

***Abstract*—An Image-to-Image (I2I) translation technique is a method that transforms an image from one domain to another by mapping one domain onto another. This technique involves two generators and two discriminators. Each generator can only translate one domain to another. This paper proposes a new approach called Knowledge Distillation Generative Adversarial Network (KD-GAN). The KD-GAN uses an image generated from Cycle-Consistent Generative Adversarial Networks (CycleGAN) as part of the target in training for a new generator. Our experiment involved translating between males and females in the CelebA dataset. We compared our model's results with the state-of-the-art using Fréchet Inception Distance (FID) and Kernel Inception Distance (KID). The experiment showed that while KD-GAN is not the best regarding FID and KID, the output image can better keep the skin tone and hairstyle from the input image than other methods.**

***Keywords*—Generative Adversarial Network (GAN), unpaired Image-to-Image (I2I) translation, Knowledge Distillation (KD), deep learning, cycle-consistency loss**

I. Introduction

Image to Image (I2I) [1] translation is an area within computer vision and artificial intelligence that involves transforming visual content from one form to another. Unlike traditional image processing tasks, I2I translation requires converting an input image from a source domain to an output image in a target domain while retaining important information and preserving the overall structure of the image.

The I2I technology differs from Generative Adversarial Networks (GANs) [2–4], requiring an input image to transform to another domain. I2I technology can be applied in various fields, such as image segmentation [5], sketch-to-image [6], and image colorization [7]. However, the basic I2I requires a pair of image datasets for input and output, which can be challenging in large-scale model training.

CycleGAN [8] is a type of GAN that utilizes the Cycle-Consistency technique to enable the model to learn how to translate images between different domains even when the datasets are unpaired. CycleGAN-based models use two generators and two discriminators to translate images between two domains. There are separate generators for translating from domain A to domain B ($G_{A\to B}$) and from domain B to domain A ($G_{B\to A}$), as well as discriminators for each domain.

To simplify I2I usage, KD-GAN combines the techniques of Knowledge Distillation (KD) [9] and GAN loss function [2]. KD is a commonly used method for transferring knowledge from one model (teacher model) to another (student model), where the output from a trained model is used as a soft target for a new model. In addition to using soft targets from $G_{A\to B}$ and $G_{B\to A}$, we proposed to train the KD-GAN with a new GAN loss function.

Our experiment aims to find the optimal settings for KD-GAN in male-to-female image transformation tasks. To achieve this goal, we utilized the CelebA dataset [10] for training and employed the state-of-the-art UNet Vision Transformer cycle-consistent GAN version 2 (UVCGANv2) [11] as the teacher model. Moreover, we enhanced the image translation results by applying post-processing with GAN Prior Embedded Network (GPEN) [12].

This paper introduces three key contributions: 1) A new learning method that combines unsupervised GAN loss with supervised Pixel-to-Pixel loss; 2) We present KD-GAN, a conditional GAN model capable of performing image-to-image translation; and 3) Analyze and discuss the post-processing techniques to enhance the translated results.

The rest of the paper is structured as follows: Section II reviews related work, and Section III presents a detailed explanation of our proposed method. Section IV outlines the experimental details, including the dataset and evaluation metrics. Section V discusses the results obtained for each metric and analyzes the image quality through visualization. Finally, Section VI presents the conclusion.

II. Related Work

The CycleGAN framework, as shown in Fig. 1, is utilized for unpaired I2I translation. It makes use of two generator-discriminator pairs. In CycleGAN, there are two domains, A and B. The generator $G_{A\to B}$ is responsible for converting images from domain A to resemble those from domain B. Simultaneously, discriminator D_B distinguishes

Manuscript received December 20, 2023; revised February 5, 2024; accepted March 26, 2024; published August 7, 2024.

images in domain B from those converted from domain A. The same process applies to the other translation directions, $G_{B\to A}$ and D_A. The discriminators are trained by backpropagating the loss in distinguishing between real and translated (fake) images. The CycleGAN is designed to learn the mapping without paired training examples.

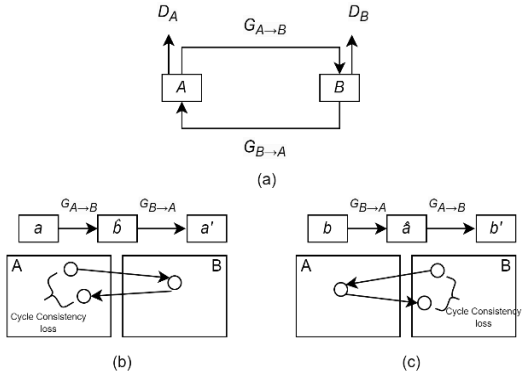

Fig. 1. (a) Overview of CycleGAN, (b) Learning to translate A to B, (c) Learning to translate B to A.

UVCGAN [13] is an upgraded version of CycleGAN that incorporates the generator's transformer architecture [14, 15] and new training methods, such as self-supervised pre-training. It has been proven that UVCGAN outperforms CycleGAN concerning FID and KID scores for male-to-female, cat-to-dog, and selfie-to-anime transformations.

UVCGANv2 [11] is the updated version of UVCGAN, which inherits the concept of UVCGAN but has improvements in the generator network architecture. UVCGANv2 enhances the design at the generator's bottleneck; the image is encoded as a sequence of tokens to be fed to the Transformer network. The result is an improvement in UVCGAN concerning the FID and KID benchmarks.

CycleGAN-based Image-to-Image (I2I) translation requires two generators to translate images between different domains. However, it is possible to have generator networks that can translate between more than two image domains. Image translation can be accomplished using technologies such as GAN. The basic GAN architecture includes a conditional GAN (cGCN) [16], which can generate a design image domain. However, it is important to note that while cGAN is an image generator, CycleGAN is an image translator, which feeds different inputs to the model. Both of these are used in unsupervised learning.

To combine the generators of CycleGAN, we utilize the KD technique [9]. KD is a universal approach to supervising the training of student networks by transferring the knowledge of already-trained teacher networks. The primary purpose of KD is to compress the model size by replicating the output of a group of models. Knowledge distillation has recently been widely explored and adopted in various applications such as image classification [17, 18], domain adaptation [19, 20], object detection [21, 22], semantic segmentation [23, 24], and GAN [25].

GANs [2–4] generate low-resolution images, which need to be upscaled and sharpened to meet the requirements of real-world applications. Super-resolution models [26] are commonly used to increase the size and sharpness of images. Super-resolution is a computer vision task that takes low-resolution input to produce an output image with higher resolution while maintaining the original content and structure.

Our research experiment focuses on male-to-female image translation, which is highly relevant to facial recognition research. Many studies have utilized super-resolution techniques in facial images, such as [12, 27]. In this paper, we have used the pre-trained model from reference [12] to sharpen the image through facial restoration, thereby improving the results.

III. Methods

A. Knowledge Distillation for GANs

KD [9] is a technique to transfer knowledge from one model to another. The most common use case is to transfer knowledge from a large model to a smaller one. Transfer knowledge is done because large models have a higher knowledge capacity (or parameters) that need to be utilized more, making them computationally expensive to evaluate. KD aims to transfer knowledge to smaller models without losing validity. Smaller models are less computationally expensive and can deploy on less powerful hardware like mobile or IoT devices.

The teacher model has been trained and provided knowledge. On the other hand, the student model receives the knowledge from the teacher model and enhances its performance. In supervised learning, the student model acquires knowledge from actual targets and the output from the teacher model, also called a soft target.

KD is designed for supervised learning, but GANs are unsupervised learning models that rely on loss functions for their generator and discriminator. It is necessary to introduce new loss functions to apply KD and soft targets to GANs and make the model semi-unsupervised.

$$Loss_{disc} = \mathbb{E}_X l_{gan}(D(x), 1) + \mathbb{E}_Z l_{gan}(D(G(z)), 0) \quad (1)$$

$$Loss_{gen} = \mathbb{E}_X l_{gan} D(G(z), 1) \quad (2)$$

Eqs. (1) and (2) show the original loss function of GANs, where l_{gan} presents a classification loss function (L2, cross-entropy, Wasserstein [3], etc.) In L_{disc} the first term represents the expected probability the discriminator assigns to real data. The second term means the expected log probability assigned by the discriminator to fake data generated by the generator. The main objective of GAN's loss function is to optimize the balance between the discriminator's ability to distinguish real and fake data and the generator's ability to produce realistic data.

Our proposed model, KD-GAN, involves training a model using both a soft target and GAN's loss. The soft target directs the output image, while GAN's loss guides the model in producing a realistic image.

B. KD-GAN Loss Function

We used the soft target obtained from UVCGANv2 as teacher models. Our L_{soft} function is defined to measure the difference between results from teacher models and our model. To calculate the $Loss_{soft}$, we use mean square error as described in Eq. (3), where n represents the number of samples, Y represents the pixel value of the generated image using trained UVCGANv2, and $\widehat{Y}_i$ represents the pixel value of the newly generated image.

$$Loss_{soft} = \sum_i^n \frac{|\widehat{Y}_i - Y_i|}{n} \quad (3)$$

It is not advisable to rely solely on $Loss_{soft}$ when training a new generator as it may result in the generator producing an image that is too similar to the teaching model. Since the teaching model still does not create the real image, a modified discriminator is used to output zero for the real image. The new discriminator output helps the training generator to minimize the output of the discriminator in the same way as a GAN loss.

$$Loss_{disc} = \mathbb{E}_C \, l_{gan} D(x, C, 1) + \mathbb{E}_{\acute{C}} \, l_{gan} D\left(G_{C \to \acute{C}}(x), \acute{C}, 0\right) \quad (4)$$

$$Loss_{GAN} = \mathbb{E}_{\acute{C}} \, l_{gan} D\left(G(x), \acute{C}, 1\right) \quad (5)$$

Eqs. (4) and (5) represent the updated discriminator and generator loss, where C is the image's label and $\acute{C}$ is the target translation label. The first term indicates the expected probability distribution of the discriminator to assign to real data. The second term means the expected probability distribution of the discriminator to give to the data generated by UVCGANv2 in different image domains.

The modification aims to create a discriminator that does not consider UVCGANv2-generated data as real data. This modification $Loss_{GAN}$ has the objective to produce an image that is similar to the real image and has the same direction as the teacher model.

In addition, we have incorporated the identity loss function employed in CycleGAN to ensure that the image is not altered when both the source and target label translations are identical. The $Loss_{idt}$, described by Eq. (6), ensures that the input and output data are alike, where l_{reg} is can be any regression loss function (L1 or L2, etc.)

$$Loss_{idt} = \mathbb{E}_C l_{reg}(G_{\acute{c} \to c}(x), x) \quad (6)$$

Eq. (7) defines the loss function during the KD-GAN training process. The equation optimizes the model's performance by adjusting the importance of three hyperparameters: λ_{GAN}, λ_{soft}, and λ_{idt}. Each of these hyperparameters has a specific objective in the training process. By adjusting their values, certain objectives can be prioritized over others, allowing for fine-tuning of the model's performance.

$$Loss_{KDGAN} = \lambda_{GAN} \text{Loss}_{GAN} + \lambda_{soft} \text{Loss}_{soft} + \lambda_{idt} \text{Loss}_{idt} \quad (7)$$

C. KD-GAN Network Architecture

1) Generator

Different network architectures can be used as the generator to transfer knowledge from multiple teacher models to a single student model. UVCGANv2 is based on the U-Net [28] architecture with modifications to the transformer [14] in the backbone. Therefore, the generator used should be based on U-Net and transformer.

In KD-GAN, we also use the U-Net architecture, but we require the target translation label to be inputted into the model. Our generator uses U-Net with positional encoding.

The network architecture of KD-GAN is based on U-Net and Positional Embedding. The label is encoded using a transformer at the end of each down-sampling and up-sampling block. Overall, the KDGAN network has more parameters and is larger than the UVCGANv2 generator. However, the larger network has the advantage of having more capacity to learn from the teacher model, which allows it to train for translation for more than two domains.

2) Discriminator

The discriminator is an essential component of a GAN, distinguishing between genuine data and data generated by the generator network. We have used the standard GAN [2] setup as the basis for our discriminator architecture.

In the KD-GAN model, we have modified the discriminator to handle 256×256 pixel images by increasing the number of layers and channels. Although the discriminator architecture is similar to the original GAN, it has been refined to be suitable for large image sizes.

In summary, our proposed KD-GAN framework is presented in Fig. 2, which provides an overview of the KD-GAN architecture.

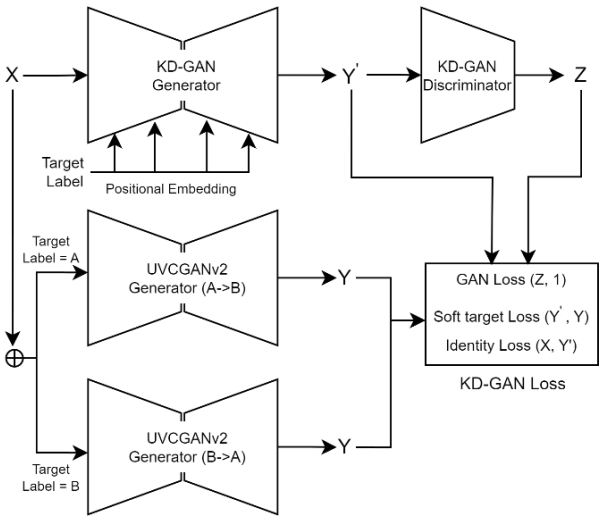

Fig. 2. Overview of KD-GAN framework.

D. Post-Processing

Our experiment focuses on translating images between the domains (male and female) and improving the realism of translated images. While GANs can generate images up

to 256×256 pixels using UVCGANv2 models, they still lack the detail required for many real-world applications, where many researchers have overcome this issue using super-resolution methods to improve low-resolution images in scale, resolution, and level of detail.

To enhance the detail and realism of our translated images, we use a GPEN [×1] state-of-the-art Face Restoration. By applying this pre-trained model to our translated images, we can improve their level of detail and make them more realistic. Improving the facial image's realism is crucial because our image-to-image translation process primarily focuses on changing its domain.

IV. Experimental

A. Dataset

To compare the effectiveness of KD-GAN with other advanced models for translating male-to-female facial features, we conducted an experiment using the CelebA dataset [10]. Our experiment used a version of the dataset that included 30,000 facial images and was also used to train UVCGANv2. The dataset was divided into two subsets: the training and validation sets. The training set consisted of 17,943 images of females and 10,057 images of males, while the validation set had an equal number of images of both genders, with 1,000 images in total. All images were resized to 256×256 pixels for consistency.

B. Training Detail

A commonly used technique for cycle consistency training involves using two generators. However, a single generator using a standard GAN training approach with soft targets is possible in our proposed KD-GAN training. In this approach, the discriminator processes the input data, which generates an image. This generated image is then compared with the soft target image, and the discriminator provides feedback to the generator. This feedback helps the generator produce more accurate and realistic images during training. The generator is trained with the discriminator loss and soft-target, aiming to generate real images in the same direction as the output from the teacher model.

We trained the KD-GAN model on a training set and then evaluated its performance on a validation set. As KD-GAN is a semi-supervised learning model, we trained it similarly to supervised learning. To optimize the model, we used the Adam optimizer with β values of (0.5, 0.99) and set the learning rate to 1×10^{-4} at the beginning. At the end of each epoch, we calculated the loss with a validation set; if the loss was not reduced compared to the previous epoch, we decreased the learning rate by 50% until it reached 1×10^{-6}. We considered the best model with the lowest loss value during evaluation and used it to evaluate the result.

Hyperparameter tuning is essential while conducting experiments as it can significantly affect the results. However, our research indicates no significant difference in outcomes when λ_{GAN} is less than 0.5. However, our model fails to translate when λ_{GAN} is greater than or equal to 0.5. It is important to note that the results presented in this paper are obtained with specific hyperparameter values of $\lambda_{GAN} = 0.1$, $\lambda_{soft} = 0.8$, and $\lambda_{idt} = 0.1$.

C. Evaluation Metric

In the context of unpaired image-to-image translation, it is customary to employ metrics like Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) to evaluate the similarity between the generated images and the real ones. FID and KID serve as benchmarks to determine the quality of the generated images. This benchmark confirms the quality of the generated images and provides a means of comparison with baseline and state-of-the-art models.

1) Fréchet inception distance

The FID [29] is a measure used to assess the quality of images produced by GANs. FID uses the Fréchet distance, a statistical method for determining the similarity between two probability distributions. In the case of GANs, these distributions represent the features extracted from real and generated images using a pre-trained Inception-v3 [30] neural network.

The mean and covariance matrix of the feature representations for both real and generated samples are determined to calculate FID. Then, the Fréchet distance is computed between the multivariate Gaussian distributions defined by these statistics, yielding a singular FID score. A lower FID score indicates a higher degree of similarity between the distributions, implying that the generated images closely match the characteristics of real images.

$$FID = \|\mu_1 - \mu_2\|^2 + Tr(\sigma_1 + \sigma_2 - 2\sqrt{\sigma_1 \times \sigma_2}) \quad (8)$$

Eq. (8) provides a detailed explanation of FID. The terms μ_1 and μ_2 represent the average features in the real and generated images. Similarly, σ_1 and σ_2 represent the covariance matrix of the feature vectors for the real and generated images. The expression $\|\mu_1 - \mu_2\|^2$ refers to the squared sum of the differences between the two mean vectors, while the operation *Tr* denotes a trace in linear algebra. Lower FID scores indicate that the statistics of the two sets of images are more similar. A score of 0.0 implies that the two groups of images are identical.

2) Kernel inception distance

KID [29] is a metric widely used to evaluate the performance of GANs in generating high-quality and diverse images. KID is calculated by extracting features from real and generated images using a pre-trained Inception model. These features are then used to compute the Gram matrix, which captures the statistical properties of the images.

The KID value is then obtained by measuring the distance between the Gram matrices of the real and generated images. By considering the statistical properties of the images, KID provides a reliable measure of the similarity between the generated and real datasets. Therefore, it is an essential tool for assessing the quality and diversity of the images generated by GANs.

$$\text{KID} = \text{MMD}(f(x), f(x'))^2 \quad (9)$$

Eq. (9) provides the calculation details for the Kernel Inception Distance (KID), where the function *f* uses a pre-trained Inception-v3 model [29], *x* represents a real image, and *x'* represents a generated image. The Maximum Mean Discrepancy (MMD) function calculates KID's mean and standard deviation over several subsets.

FID and KID differ in how they measure the similarity between the feature representations of real and generated images. FID calculates the distance between the feature representations, while KID focuses on the statistical properties of the images captured by the Gram matrix. Unlike FID, which considers the mean and covariance of feature representations, KID only examines the covariance matrix. As a result, KID offers a unique perspective on evaluating GANs by emphasizing the statistical properties of the images rather than just the feature representations.

V. Result Analysis and Discussion

As part of our experiment, we comprehensively compared various models, including the baseline and state-of-the-art ones. We evaluated the effectiveness of our method against these models, which comprised UVCGAN [13] and UVCGANv2 [11] for both male-to-female and female-to-male comparisons. It's important to mention that our KD-GAN was trained with smaller datasets than those used in UVCGAN and UVCGANv2. To compare the results with state-of-the-art, we used the trained weights of UVCGAN and UVCGANv2 and evaluated them with our version of the validation set. In our evaluation, results in FID/KID of UVCGAN and UVCGANv2 differ from those in the UVCGANv2 paper.

Our analysis involved a detailed assessment of the performance of each model, which included a comparison of their features, strengths, and weaknesses. Our experiment's results provided valuable insights into the performance of these models and helped us identify the most effective one for our specific use case.

A. Quantitative Analysis

In this experiment, we used three different methods to generate images and measured their performance using FID and KID metrics. The methods were UVCGAN, an updated version of CycleGAN; UVCGANv2, the latest iteration of UVCGAN; and our proposed KD-GAN, a technique that distills knowledge from trained UVCGANv2. Moreover, we include the result when applying post-processing to each model to see the effectiveness of post-processing.

TABLE I. FID and KID Scores

Model	Male-to-Female		Female-to-Male	
	FID	KID	FID	KID
UVCGAN	32.583	0.018	45.762	0.024
UVCGANv2	34.175	0.023	39.992	0.024
KD-GAN	62.631	0.040	54.295	0.035
UVCGAN+GPEN	27.666	0.012	37.648	0.017
UVCGANv2+GPEN	24.357	0.009	31.883	0.012
KD-GAN+GPEN	35.124	0.018	57.172	0.038

Table I displays the FID and KID results of our experiment. It indicates that UVCGAN and UVCGANv2 demonstrate better performance compared to KD-GAN. Furthermore, we found that UVCGANv2 has slightly surpassed UVCGAN. Upon analyzing the results, we discovered that KD-GAN has an issue that causes high FID and KID scores due to the sharpness of the image in the jawline and ear. Fig. 3 (upper) shows the results that cause higher FID and KID in KD-GAN.

Fig. 3. Male-to-female translated using KD-GAN and KD-GAN with post-processing (Upper male to female, lower female to male).

KD-GAN significantly improves male-to-female image translation tasks when used with post-processing techniques. The scores of KD-GAN with post-processing are similar to those of UVCGANv2 without post-processing. The improvement is because image blur can be corrected in post-processing.

However, in female-to-male, the score of KD-GAN is not improved while using post-processing. From the observation, we found that some translations with a hairline cover partial of the face remain in the output as a transparency overlay. The examples are shown in Fig. 3 (lower), where the transparency overlay cannot be removed by using post-processing.

The analysis of FID and KID measurements suggests that the KD-GAN algorithm can be further optimized to improve the realism of generated images. Although some issues can be resolved through post-processing, we recommend incorporating a sharpness loss function [31] to enhance the final image quality without relying on post-processing.

B. Qualitative Analysis

Our visual representations are used to evaluate the quality of image translation models. Figs. 4 and 5 demonstrate the results of image translations from male to female and female to male using our method and state-of-the-art models.

The translated images generated by UVCGAN, UVCGANv2, and KD-GAN successfully depicted the intended image translation. However, upon closer examination, it was observed that some of the images produced by UVCGAN (v1 and v2) exhibited over-modification of the facial features, destroying the original image structure, such as aging and hairstyle.

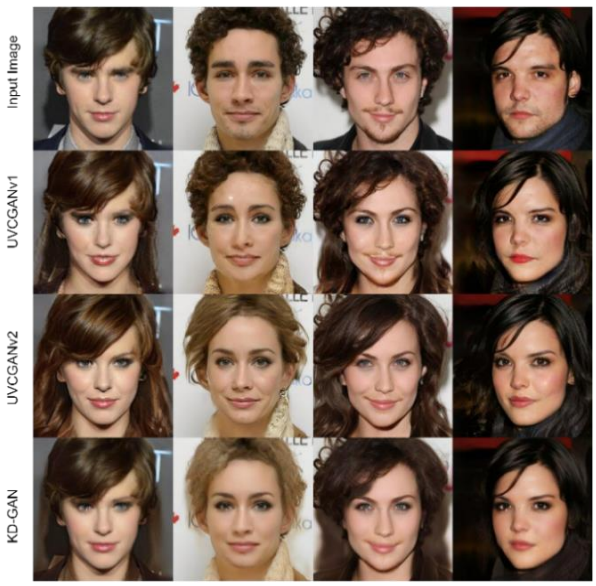

Fig. 4. Sample translation for male-to-female.

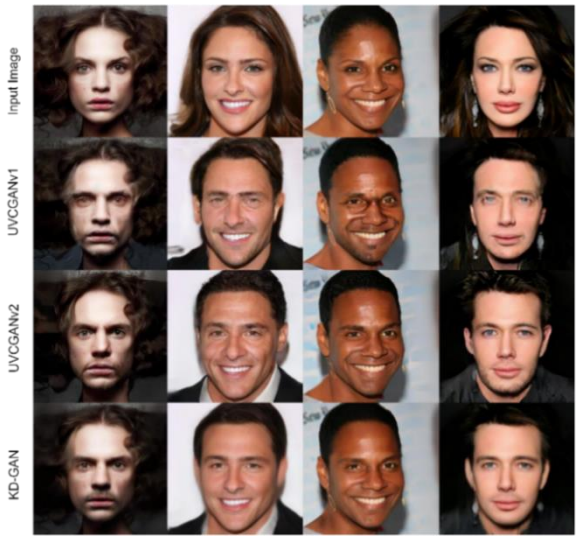

Fig. 5. Sample translation for female-to-male.

For male-to-female transformations, the primary changes to the input image were removing facial hair and smoothing the skin. All the models could convert the image successfully; however, UVCGAN and UVCGANv2 models tended to produce unrealistic images by over-modifying the hairstyles. On the other hand, KD-GAN generated a more realistic hairstyle that maintained the original image's structure.

All models successfully translated images from female to male, but UVCGAN produced unrealistic images, and UVCGANv2 aged the input image's skin. On the other hand, KD-GAN can maintain the skin tone from the original image.

According to the experiment, Fig. 6 demonstrates that post-processing using GPEN [12] can enhance the outcome of image translation between male and female domains. However, it's important to note that GPEN is customized for facial images, and for other image-to-image tasks, individual task post-processing needs to be investigated.

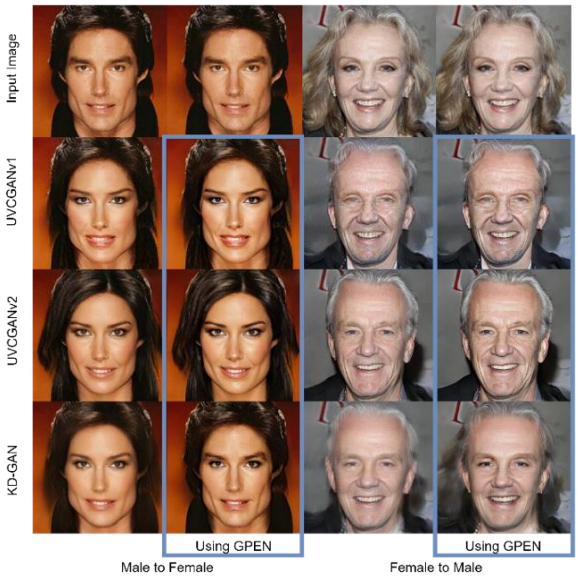

Fig. 6. Sample translation with and without post-processing.

To summarize, the experiment results show that KD-GAN's learning method has the potential for I2I translation by incorporating assistance from the discriminator and teacher models. The discriminator model assists in optimizing the image translation process by identifying differences between the generated and real images, enabling the generator to produce more accurate and consistent images. The teacher models provide guidelines for image output. However, the results still have sharpness issues that require post-processing.

VI. Conclusion

Our research paper presented a new learning method and model called KD-GAN. This method combines the KD technique with GANs to translate images between different domains. KD-GAN is a conditional generator that uses the target domain as input and the original image to produce the output for the respective image domain. We used the CelebA dataset for male-to-female and female-to-male translations to test KD-GAN and used the UVCGANv2 models as the teachers for training. Our experiment showed that KD-GAN can translate images successfully between male and female domains. However, KD-GAN still faces image quality issues, which can be resolved by post-processing in male-to-female tasks. Further investigation is needed for female-to-male tasks.

In our future work, we aim to analyze the factors that can significantly impact our results, such as feature normalization [32]. Our objective is to gain a better understanding of how hyperparameters, generator network architecture, datasets, loss functions, and objective functions interact with each other.

Conflict of Interest

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Rung-Ching Chen conducted the research; Chayanon Sub-r-pa analyzed the data; Chayanon Sub-r-pa wrote the paper; both authors had approved the final version.

FUNDING

This paper is supported by the National Science and Technology Council, Taiwan. The Nos are NSTC-112-2221-E-324-003-MY3 and NSTC-112-2221-E-324-011.

REFERENCES

[1] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
[2] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. the International Conference on Machine Learning*, 2017, pp. 214–223.
[4] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
[5] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.
[6] W. Chen and J. Hays, "Sketchygan: Towards diverse and realistic sketch to image synthesis," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9416–9425.
[7] S. Huang *et al.*, "A fully-automatic image colorization scheme using improved CycleGAN with skip connections," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26465–26492, 2021.
[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in in *Proc. the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
[9] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789–1819, 2021.
[10] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.
[11] D. Torbunov *et al.*, "UVCGAN v2: An improved cycle-consistent GAN for unpaired image-to-image translation," arXiv preprint, arXiv:2303.16280, 2023.
[12] T. Yang, P. Ren, X. Xie, and L. Zhang, "GAN prior embedded network for blind face restoration in the wild," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 672–681.
[13] D. Torbunov *et al.*, "UvcGAN: U-net vision transformer cycle-consistent GAN for unpaired image-to-image translation," in *Proc. the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 702–712.
[14] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[15] K. Han *et al.*, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2022.
[16] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. the 2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 2089–2093.
[17] X. Zhu, S. Gong *et al.*, "Knowledge distillation by on-the-fly native ensemble," *Advances in Neural Information Processing Systems*, vol. 31, 2018.
[18] F. Zhan *et al.*, "Unbalanced feature transport for exemplar-based image translation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15028–15038.
[19] L. T. Nguyen-Meidine, A. Belal, M. Kiran, J. Dolz, L.-A. Blais-Morin, and E. Granger, "Unsupervised multi-target domain adaptation through knowledge distillation," in *Proc. the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1339–1347.
[20] S. Gupta, J. Hoffman, and J. Malik, "Cross modal distillation for supervision transfer," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2827–2836.
[21] G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficient object detection models with knowledge distillation," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[22] R. Cheng, B. Wu, P. Zhang, P. Vajda, and J. E. Gonzalez, "Data-efficient language-supervised zero-shot learning with self-distillation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3119–3124.
[23] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2604–2613.
[24] C. Yang, H. Zhou, Z. An, X. Jiang, Y. Xu, and Q. Zhang, "Cross-image relational knowledge distillation for semantic segmentation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12319–12328.
[25] K. Cui, Y. Yu, F. Zhan, S. Liao, S. Lu, and E. P. Xing, "Kd-dlgan: Data limited image generation via knowledge distillation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3872–3882.
[26] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.
[27] R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE Transactions on Image Processing*, vol. 29, pp. 6251–6263, 2020.
[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. the 18th International Conference Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*, Munich, Germany, 2015, pp. 234–241.
[29] 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," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[30] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. the AAAI Conference on Artificial Intelligence*, 2017, vol. 31, no. 1.
[31] S. Butte, H. Wang, M. Xian, and A. Vakanski, "Sharp-GAN: Sharpness loss regularized GAN for histopathology image synthesis," in *Proc. the 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022, pp. 1–5.
[32] W.-C. Cheng, H.-C. Hsiao, and D.-W. Lee, "Face recognition system with feature normalization," *International Journal of Applied Science and Engineering*, vol. 18, no. 1, pp. 1–9, Mar. 2021.