# Icon Style Transformation Based on Generative Adversarial Networks

*Hongyi Yang*[1], *Xinyue Wang*[1], *Chengqi Xue*[1] *and Xiaoying Yang*[1],

[1] School of Mechanical Engeering, Southeast University
Nanjing, 211189, China

## ABSTRACT

The icon is an important part of the user interface, and they are a carrier between the user and the interface. In the process of icon design, designers need to take into account both its versatility and uniqueness, and an excellent icon is a perfect blend of function and aesthetics. In recent years, with the great success of generative adversarial networks in computer vision, it has become possible to assist designers in icon creation with the help of artificial intelligence technology. In this study, we constructed icon datasets containing 40,000 samples and improved the structure and loss function based on the MUINT to finally achieve the style conversion task between different styles of icons. The research results show that the improved model can effectively improve the quality and diversity of generated icons. Meanwhile, a questionnaire survey of 34 people with icon design experience proves that our research results can assist designers to a certain extent in the related work. This study can be used as a basis for the intersection of deep generative model and icon design, and we conclude the paper with suggestions and prospects for future work.

**Keywords**: Icon design, Deep learning, Generating Adversarial Network, Image Transformation.

# INTRODUCTION

The icon is a key component of the user interface, a carrier between the user and the graphical interface. A good icon not only enables users to quickly identify and understand the meaning of the function it represents but also highlights the business characteristics of the brand to a certain extent. From different perspectives, icons can be classified into various categories: face icons, linear icons, text icons, and so on. Designers have to design different types of icons according to different application scenarios and customer needs (Smith and Waterman. 1981).

In recent years, Generative Adversarial Network (GAN) (Goodfellow et al. 2014) has achieved excellent results in tasks such as image generation (Bodla et al. 2018), image complementation (Satoshi et al. 2017), and image super-resolution reconstruction (Yeh et al. 2016). In particular, the success of GAN in image translation has not only attracted wide interest from academia but also achieved good commercial landing in mobile applications. And the style transfer task in image translation has been successfully applied to a variety of fields since it was proposed by Gatys et al (Gatys et al. 2016), such as font style transfer (Wen et al. 2021), and face animation stylization (Pranjal et al. 2021). Meanwhile, with the successive proposals of models such as Pix2Pix (Liu et al. 2017) and CycleGAN (Zhu et al. 2017), GANs have become the main framework for implementing style transfer tasks.

In this paper, we hope to use GAN's success in image translation to realize the style transfer between different icons to give designers more inspiration and improve their work efficiency. Moreover, this study constructs an icon dataset consisting of linear and facet icons with a total of 40,000 samples. We hope that the model can eventually realize the style conversion work between linear and facet icons, and this result can be applied to the following scenarios: abstracting the facet icons and getting the corresponding linear icons, or getting the face icons from following the linear icons, and the icons generated by the model can provide the complementary color schemes, etc.

In order to achieve the above goal, we improve the model based on the idea of MUINT (Xun et al. 2018), which has the advantage of solving the problem of generating sample diversity, which can get multiple converted images after inputting one image, and this idea is also more consistent with the actual design scenarios. Meanwhile, we introduce operations such as self-attention mechanism, spectral normalization, etc., based on the original model and add cycle consistency loss to the original loss function to improve the conversion quality of icons. This study is the first study of icon style transfer based on GANs, which can be used as a basis for future studies.

# RELATED WORK

GAN-based image translation models were initially mainly supervised models with two styles of one-to-one sample data types. Among them, Pix2Pix, proposed in 2017, first proposed a general image translation problem framework for the conversion problem of paired samples, where the original style images are fed into both the generator and the discriminator as "conditional labels" to control the conversion effect of the model. BicycleGAN (Zhu et al. 2017) was designed to address the Pix2PixHD (Wang et al. 2017) proposes a multi-scale generator and discriminator structure based on the Pix2Pix, and introduces a feature matching loss in the loss function to improve the stability of training. Finally, higher resolution images are generated. Furthermore, Vid2Vid (Wang et al. 2018) implemented a video-to-video high-resolution style translation task using optical flow and timing constraints based on the above models.

The biggest problem with supervised image translation models is that it is difficult to obtain a sufficient number of high-quality paired datasets in realistic situations, so most subsequent related models are unsupervised. One of them, CycleGAN, achieved the first conversion task between two image domains in data asymmetry using two sets of generator domain discriminators. The original and target domains are exchanged for StarGAN uses a single generator to achieve transformation in multiple data domains generates higher quality image samples (Yunjey et al. 2018). On the other hand, the authors of UNIT proposed the assumption that different data spaces share the same hidden space, thus converting the image translation problem into a problem of solving the hidden space and finally achieving an unsupervised inter-image style transformation task (Liu et al. 2017).

Along with the success of GAN in computer vision, it has gradually become possible to use artificial intelligence technology to assist designers in creation, and many designers are trying to use GAN for design creation (Li et al. 2019; Nauata et al. 2020). In graphic design, a logo and icon are both visual symbols, but a logo is more to carry a company's values. In contrast, an icon carries more superficial information with more robust functionality, and its purpose is to make users use or browse a product more. Unlike icon generation, logo generation does not require a high generated image quality because it does not need robust functionality and recognizability. The earliest research can be traced back to Sage et al. in 2017, whose main contribution was to construct the first large-scale logo dataset, LLD-logo (Sage et al. 2017), on which the logo generation task was based (Sage et al. 2017). Since then, Mino et al.'s research team constructed LoGAN (Mino et al. 2018) and LoGANv2 (Oeldorf et al. 2019), where LoGAN implemented the colour-conditional logo generation task and LoGANv2 implemented the logo generation task at high resolution. Inspired by the above logo generation-related tasks, researchers have gradually focused on icon generation tasks.

# METHODS

## Generative Adversarial Network, GAN

GAN consists of a generator and a discriminator. The two modules are trained alternately to generate samples comparable to real data eventually (see Figure 1). During the mutual adversarial process, the discriminator's goal is to determine as accurately as possible whether the input samples are from the real data or the generator; the generator's goal is to generate as many samples as possible that the discriminator cannot distinguish.

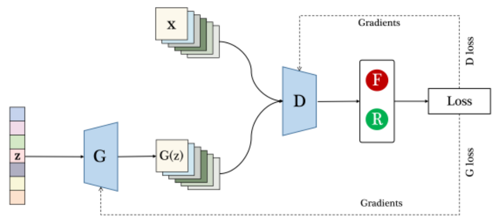



Figure 1. The basic framework of GAN.

## MUNIT

MUNIT can be regarded as an improvement on UNIT, which further assumes that the latent code can be decomposed into content code and style code based on the shared latent code, where the content code contains some shallow information of the image data, such as contours, edges, etc., while the style code (see Figure 2). The model assumes that the style code belongs to a Gaussian distribution. Therefore, for the samples in the source domain, fixing their content codes and changing their style codes can yield different images, and the combination of content codes and different style codes can generate multiple samples with the style of the target domain, which solves the problem of lack of style diversity in the previous image translation problems.

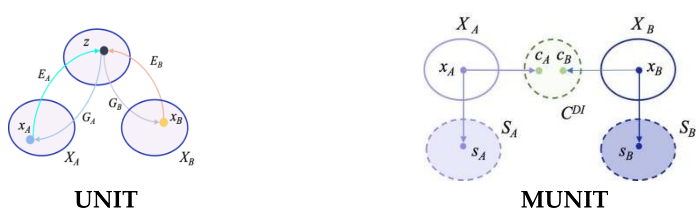



Figure 2. Comparison of the UNIT with the idea of MUNIT.

# EXPERIMENTS AND RESULTS

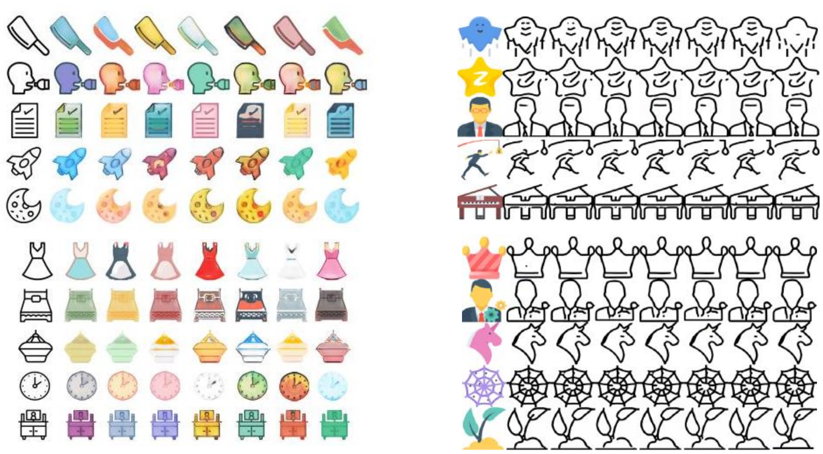
## Datasets And Train Detail

The data in the study were mainly obtained from Iconfont (www.iconfont.cn.) and Icons8 (www.icons8.cn.), which has an extensive resource of high-definition watermark-free icons. We selected a total of 40,000 linear and faceted icons and processed all the data into PNG format.

We performed the training on RTX 3090, where the optimizer was selected as Adam optimizer, the parameters used were 0.5 and 0.999, and the batch size was set to 1. The learning rate was adopted from the TTUR method (Heusel et al. 2017), and the learning rates of the generator and discriminator were set to 0.0002 and 0.0004, respectively; the sampled from the target domain.

## Results

Figure 3 shows the icon style conversion results of the model. The conversion results in converting linear icons to face icons with good diversity and can generate various exciting colour fills and combinations. In contrast, in converting face icons to linear icons, the conversion results are more single and straightforward, and the model tends to learn the contour information of the face icons more but fails to generate some detailed texture structures in the original icons.



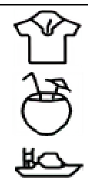
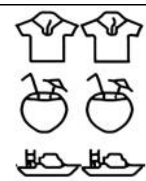
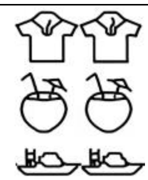
(**a**) (**b**)

Figure 3. (a) Face icon is converted to linear; (b) Linear icon is converted to face.

We conducted multiple comparison experiments in the same environment. In the conversion of linear icons to facet icons, our model has higher generation quality, and also improve the conversion results compared with the original MUNIT (see Table

1). In converting face icons to linear icons, the model does not show significant improvement compared to Baselines.

Table 1: Comparison with Baselines' icon style transfer results.

| | **Iuput Image** | **CycleGAN** | **UNIT** | **MUNIT** | **Ours** |
|---|---|---|---|---|---|
| Linear→Face |  |  |  |  |  |
| Face→Linear |  |  |  |  |  |

Our model has a clearer texture structure, especially in converting linear icons to face icons, and the improved model has a significant improvement. In addition, since we can set the number of samples encoded from the target domain style, our conversion results have multiple outcomes (see Table 2).

Table 2: Quantitative evaluation according to IS and FID.

| | **IS** | **FID** |
|---|---|---|
| CycleGAN | 13.67 | 157.67 |
| UNIT | 19.09 | 176.34 |
| MUNIT | 19.76 | 137.03 |
| Ours | 22.69 | 131.77 |

The quantitative evaluation results can prove the above subjective visual perception, and we compared the conversion results with IS (Salimans et al. 2016) and FID (Szegedy et al. 2016). As shown in Table 2, CycleGAN and UNIT have lower IS and higher FID scores, and the model generates lower quality and diversity results in the task. In contrast, both MUNIT and our improved model have a better perceptual evaluation of human behaviour and FID scores, and our improved model has some improvement effect compared with the original model.

## User Study

We conducted a relevant study with 34 subjects to test whether the icon style transfer study can assist in design tasks to some extent. We will present subjects with 20 sets of icons, each consisting of the original image and five transformed icons, and the interconversion between the two sets of icons will be performed separately, and the subjects will be asked to evaluate them (see Figure 4). Afterwards, we asked the participants to evaluate the converted icons based on a 5-point scale, and the criteria were whether the icons were helpful to their daily design tasks. The study results are shown in Figure 5, where the 5 points are "1=very low, 2=slightly low, 3=medium, 4=slightly high, 5=very high.".

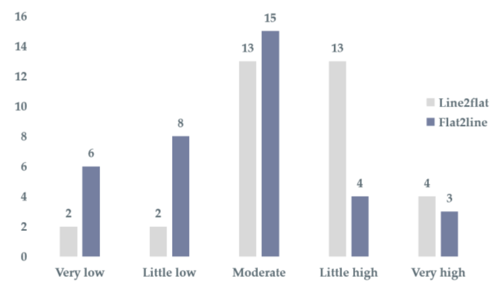



Figure 4. Results of the survey on the usefulness of generating icons for design work.

From the research, the style transfer from linear icons to facet icons was better evaluated, and most of the participants thought that it could effectively provide a variety of color schemes and give them more design inspiration to a certain extent. In contrast, the result of switching from face icons to linear icons was considered to be of limited significance.

# CONCLUSIONS

Icon design is a task that designers often encounter daily, and it is always an exciting research question how it can be combined with cutting-edge technologies related to artificial intelligence to assist in design. In this study, we implemented the style conversion task between different icons based on MUNIT and constructed a large scale icon datasets consisting of linear and faceted icons for the first time. The results show that the model we constructed can perform the icon conversion task better, and the construction idea based on the MUNIT model makes the generated icon results more diverse, which is also more in line with the needs of the actual design task. The objective and subjective evaluation metrics show that the improved model has a better generation effect. Especially for converting linear icons to face icons, our model can make the converted face icons have a more detailed texture and contour with less noise and more decadent colour combinations. The subjective evaluations conducted in the study also showed that our study could assist designers to a certain extent in the relevant tasks.

# REFERENCES

Bodla, N., Hua, G., & Chellappa, R. (September 8-14, 2018). Semi-supervised fused GAN for conditional image generation. 15th European Conference on Computer Vision, Munich, Germany.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (December, 2014). Generative adversarial nets. 27th International Conference on Neural Information Processing Systems, Cambridge, MA, USA (2672-2680).

Gatys, L.A., Ecker, A.S., Bethge, (2016). M. Image style transfer using convolutional neural networks. In: CVPR.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (December 4-9, 2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA.

Liu, T. Breuel, and J. Kautz. (2017). Unsupervised image-to-image translation networks. Proc. NIPS.

Liu, M.Y., Breuel, T., Kautz, J. (2017). Unsupervised image-to-image translation networks. In: NIPS.

Li, J. N.; Yang, J., Hertzmann, A., Zhang, J. M; Xu, T. F. (2019). LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators. ICLR.

Mino, A.and Spanakis, G. (2018). LoGAN: Generating Logos with a Generative Adversarial Neural Network Conditioned on color. arXiv: 1810.10395v1.

Nauata, N., Chang, K. H., Cheng, C. Y., Mori, G., Furukawa, Y. (2020). House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation. arXiv:2003.06988.

Oeldorf, C. and Spanakis, G. (2019). LoGANv2: Conditional Style-Based Logo Generation with Generative Adversarial Networks. arXiv preprint arXiv:1909.09974v1.

Pranjal Singh Rajput, Kanya Satis, Sonnya Dellarosa, Wenxuan Huang, Obinna Agba. (2021). cGANs for Cartoon to Real-life Images. arXiv:2101.09793.

Smith, T.F., Waterman, M.S. (1981). Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197

Satoshi Iizuka; Edgar Simo-Serra; Hiroshi Ishikawa. (2017). Globally and Locally Consistent Image Completion. ACM Transaction on Graphics.

Sage, E. Agustsson et al. (2017). Lld-large logo dataset-version 0.1. Available online: https://data.vision.ee.ethz.ch/cvl/lld.

Sage, E. Agustsson, R. Timofte, and L. Van Gool. Logo synthesis and manipulation with clustered generative adversarial networks. Unpublished.

Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. (2016). Improved techniques for training GANs. in Advances in Neural Information Processing Systems, pp. 2234–2242.

Szegedy, C., Vanhoucke, V., Ioffe, S.; Shlens, J. (2016). Rethinking the Inception Architecture for Computer Vision. In CVPR: 2818-2826.

Wen, Q., Li, S., Han, B.F., Yuan, Y. (2021). ZiGAN: Fine-grained Chinese Calligraphy Font Generation via a Few-shot Style Transfer Approach. arXiv:2108.03596.

Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, Andrew, Kautz, J., Catanzaro, B. (2017). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. arXiv:1711.11585.

Wang, T.C., Liu, M.Y., Zhu, J.Y., Liu, G., Tao, A., Kautz, J., Catanzaro, B. (2018). Video-to-video synthesis. In NeurIPS.

Xun H., Ming Y. L., Serge B., and Jan K. (2018). Multimodal unsupervised image-to-image translation. ECCV, pp.172–189.

Yeh, R. A., Chen, C., Lim, T. Y. (2016). Semantic Image Inpainting with Deep Generative Models. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Yunjey C., Minje C., Munyoung K., Jung-Woo Ha, Sunghun Kim, Jaegul Choo. (2018). StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. In CVPR.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (October 22–29, 2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. International Conference on Computer Vision, Venice, Italy.

Zhu, J.Y. ; Zhang, R., Pathak, D. et al. (2017). Toward Multimodal Image-to-Image Translation. In NIPS.