Research and Analysis on Image Style Transfer Technologies

Runxin Yang

School of Automation and Software Engineering, Shanxi University, Taiyuan, Shanxi Province, 030000, China
yangrunxin979@gmail.com

Abstract. Image style transfer technologies seek to imbue original images with a desired artistic style while preserving their content structure at the same time. Early image style transfer algorithms mostly employed non-parametric methods to achieve style transfer. Recently, there are a series of breakthroughs have been achieved relying on classic Generative Adversarial Networks (GANs). However, current mainstream methods remain to face multiple challenges in distinct areas such as general scene coverage, high-resolution control, domain-specific detail preservation, etc. This study centers on several classical GAN-based style translation strategies including DRB-GAN, DualStyleGAN, WeditGAN and Multi-scale CycleGAN. The central technique and algorithm system of image style transfer field mainly consists of these four mainstream technologies. In modern's practices, these models show distinct limitations respectively in some aspects concerning the weak generalization of unknown styles, disappearance of non-facial details and troubles in real-time implementations. The future work conducted by scholars will seriously take challenges mentioned above into consideration and make progress in cross-modal fusion, detail augmentation designs and lightweight models. In summary, this review legibly demonstrates the different points on these models' style preservation level, migration resilience and circumstances adaptability. This essay points out some advice on how to select reasonable technology in a particular scene and optimization directions for later researches in image style translation fields.

Keywords: Deep Learning, Generative Adversarial Networks, CycleGAN, Image Style Transfer.

1 Introduction

Image style transfer, combining computer vision with provided digital art styles, means a progress of remapping the information of a target image with a style picture. The appearance of generative adversarial networks (GANs) has provided crucial support on style translation applications. Ever since the GANs were coined by Goodfellow in 2014, they have sparked extensive discussions among experts worldwide. The essential theory of GAN models is a two-way and zero-sum game between a generator (G) and a discriminator (D). It can fulfill the objective of generating realistic images by

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_87

mutually alternating training [1]. The significant value of GANs in image style transfer technique truly depends on achieving adversarial training progress through using two basic components including a G and a D. And this practice allows for exactly accurate control over content preservation and style replacement [2, 3]. The advancement of GANs in image style transfer has become one of the central issues under intensive academic scrutiny for its theoretical and practical value recently. Furthermore, image style transfer methods have been widely applied in artistic creation, cultural relics preservation and medical imaging fields [4, 5].

Nowadays, there are considerable high-quality studies and experiments to explore the applications of image style transfer techniques with GAN models. Xu et al. introduced DRB-GAN to complete collection style transfer tasks and respond to the distortion problems of texture and content, with the innovation of the Dynamic Residual Block (DRB) [6]. This model use style codes as shared parameters to bridge the two tasks by fusing dynamic convolution and AdaIN. Complemented by a dual-encoder (for fine-grained style feature extraction), an SW-LIN decoder (for spatial feature normalization), and a style collection discriminator (for reference-based style alignment), the model successfully resolves the performance imbalance, artifact generation, and weak genre capture issues. Quantitative and qualitative evaluations verify that it outperforms SOTA methods in both fooling rate and human-perceived quality. Yang et al. proposed DualStyleGAN to address the gaps in traditional StyleGAN for portrait style transfer, specifically designing a dual-path architecture: an intrinsic style path dedicated to governing facial content and an extrinsic style path focused on modeling artistic styles [7]. This design, combined with a progressive fine-tuning strategy and supervision from facial destylization, successfully mitigates traditional StyleGAN's flaws and then further achieves precise transfer of color and structural styles. The disadvantages have been overcome include insufficient fine-grained control over structural styles (e.g., exaggerated deformations in caricatures) and mode collapse under few-shot training. Duan et al. presented WeditGAN to address the inherent limitations of few-shot style transfer, especially targeting over-fitting and diversity imbalance [8]. The major innovation is its latent space relocation idea. It learns a constant offset (Δ w) and relocate latent distributions across domains. In addition, it also integrates editing intensity fine-tuning strategy and perpendicular regularization to improve translation stability. The main goal of the WeditGAN's framework is to provide an alternative in few-shot settings and effectively tackle sample memorization problem. It solely trains 0.04% of StyleGAN's parameter. Liang and Yan developed a multi-scale CycleGAN model based on CycleGAN and introduced multi-scale network architecture and cross-scale constraints [9]. This method has three novelties: a multi-scale generator/discriminator to maintain fine-grained details, a hierarchical adversarial loss to counterpoise features on texture and structure, and a specific self-attention semantic matching channel to improve semantic consistency. This targeted design overcomes the underlying problems of detail loss, training instability, and style-content imbalance, achieving more accurate and stable illustration style transfer compared to conventional CycleGAN [10].

Although these models have made remarkable progress in style translation tasks based on GANs, they still need to overcome some inevitable challenges. This paper

will conduct a series of comprehensive studies and in-depth analysis on these four mainstream technologies mentioned above. Through experimental results on classic datasets, and with the help of some evaluation metrics, the paper will analyze the advantages and drawbacks of distinct models and propose future prospects.

2 Mainstream Technologies

2.1 DRB-GAN

Style Code. This code can convert the visual features (color, brushstrokes) of stylized images into numerical strings. This string of numbers serves as a shared parameter, guiding the migration between individual styles and group styles. For example, individual style codes (such as a single Picasso work) and aggregate average style codes (such as a collection of Picasso works) can be generated using the same logic.

Dynamic Residual Block (DRB). DRB receives content image structural features (such as the position of facial features) and adjusts its own parameters according to the style code (e.g., optimize the brush stroke filtering method according to the Van Gogh style). The final output is a structural feature with style attributes.

SW-LIN decoder is used to avoid false textures (such as blurred facial edges and color banding) when generating high-resolution images. Optimization is achieved by adjusting features within a local window, thereby ensuring that the produced image is both sharp and clear and fits the target style.

The proposed approach employs a two-layer feature extraction architecture of VGG network and learnable small network to extract general visual features and style detail features from specific style images. After alignment work, a series of standard style code can be generated. This method firstly extracts structural characteristics from the content image, and then inputs the extracted structural features into a dynamic residual block (DRB). These style codes guide the residual block to properly adapt to its internal parameters, ultimately fulfilling precise fusion of both structural and style traits. The fused stylized structural features are input into the SW-LIN decoder to generate a high-resolution stylized image that effectively avoids false textures. The discriminator uses a single style map and a style ensemble map as reference benchmarks to judge the style authenticity of the generated image, thus indirectly encouraging the generator to optimize the style consistency and visual clarity of the output image.

Its training mainly relies on the Place365 dataset (624,777 content images) and the WikiArt dataset (11 art sets). The deception rate is 0.573, which is better than AST's 0.450, and the human perception score is 72.2% for content and style. Its benefits include comprehensive task coverage, high efficiency, and excellent visual quality. Its drawbacks contain complex network structure and weak generalization to non-natural styles.

2.2 DualStyleGAN

Dual-style path decoupling mechanism. This paper designs a dual-style path architecture, decoupling content control and style control into two independent processing pipelines. The Intrinsic Path focuses on preserving core facial information, such as the individual's facial contours and features, ensuring the uniqueness and recognizability of the person's identity after style transfer operations. The Extrinsic Path specifically extracts and applies target style features, such as exaggerated large eyes in a cartoon style or dark outlines in a comic book style, securing accurate style transfer without altering the people's identity attributes.

Progressive fine-tuning training strategy. To avoid training instability or feature learning bias caused by directly learning complex style transfer tasks in the early stages of training, this paper adopts a progressive fine-tuning strategy. The training process can be divided into three stages: the first stage focuses on basic color transfer tasks, such as converting the color system of real photos to the target artistic style (e.g., high-saturation colors in cartoon style); the second stage advances to structural transfer tasks, such as adjusting the facial structure of real faces to conform to the target style (e.g., a round face shape in cartoon style); the third stage finally completes high-definition transfer training, gaining high-fidelity stylized output at 1024×1024 resolution.

Facial destylization assisted training. To solve the essential issue of facial feature distortion during style transfer, this paper introduces facial destylization techniques to construct auxiliary training data. By generating a paired dataset of "artistic portrait → real face" (e.g., inversely converting cartoon-style portraits into corresponding real face photos), this dataset serves as a reference benchmark for model training, providing the model with a supervisory signal of style transfer boundary. This helps the model learn reasonable style mapping relationships, ensuring that key facial features (such as the position of facial features and facial proportions) are not distorted or deformed during style transfer.

The DualStyleGAN model logic is as follows. Firstly, cartoon/comic portraits are converted into corresponding real faces employing facial destylization techniques, constructing a style map-content map pairing dataset to provide a basis for supervised training. Next, a dual-path fusion architecture is adopted. The inner path receives the latent code of the real face to accurately control the core content information of the person, while the outer path inputs the latent code of the cartoon portrait and achieves style feature regulation by adjusting the facial structure in the coarse layer and optimizing the color style in the fine layer. The outputs of the two paths are fused to generate a styled portrait feature that has both content consistency and style recognition. Finally, the fused feature is input into the StyleGAN2 synthesis network to produce a high-definition stylized portrait with a resolution of 1024×1024. At the same time, the discriminator performs dual supervision on the generated result—both judging whether the portrait conforms to the target style (such as cartoon style) and making sure that the

main identity information of the person is not lost, thereby achieving high-fidelity and high-resolution facial style transfer.

It uses a dataset of 317 cartoon images, 199 comic images, and 174 anime images, with user preference rating (0.83, far exceeding StarGANv2's 0.02) as the core metric. Its advantages are high-resolution generation and flexible style control. Besides, its disadvantages are that it is only suitable for portrait scenes, non-facial details are easily lost, and it is greatly affected by data bias (such as anime straight-out preferences).

2.3 WeditGAN

Latent space relocalization. WeditGAN models consider image style transfer as a spatial translation problem. The source domain latent space corresponds to the parameter space generated from content such as real faces. The target domain latent space accords with the parameter space generated from styles such as sketches. Both have similar structures but differ in spatial location. The model learns a fixed offset Δw to translate the source domain latent space to the target domain latent space, thereby generating a face in the target style.

Constant offset Δw design. The model learns only a single constant offset Δw, which represents solely 0.04% of the generator's total parameters. The remaining parameters are reused from a pre-trained generator, such as a real face generator. This design effectively mitigates the overfitting problem due to parameter redundancy in scenarios with few samples.

WeditGAN's model logic can be segmented into three major steps. At first, it uses StyleGAN2 as the basic generative model and pre-trains it on a large-scale real face dataset to enable it to generate high-resolution realistic face images. Secondly, the model utilizes simply up to 10 target style reference images (such as sketch-style images) and concentrates on training a single offset parameter Δw. The latent code corresponding to the real face is added to Δw and then input into the pre-trained synthesis network to generate an initial target style image. At the same time, a discriminator is introduced to evaluate the style realism of the generated image, and adversarial training is used to iteratively optimize Δw to its optimal value. Finally, after Δw is trained, for any input real face latent code, it is merely necessary to add it to the optimal Δw and then input it into the pre-trained synthesis network to generate the target style image. The entire process does not require additional training of other parameters, achieving efficient style transfer.

It adopts a target dataset of 10 images per group (e.g., sketches, babies), and uses FID (FFHQ→sketches 35.41, better than AdAM's 38.11) and LPIPS as metrics. Its strengths are strong adaptability with few samples, lightweight design, and a balance between diversity and fidelity. Its weaknesses include reliance on source-target domain correlation, weak generalization to unseen styles, and limited resolution (256×256).

2.4 Multi-scale CycleGAN

Multi-scale generation/discriminator. Source pictures are handled at three different resolutions respectively: low, medium, and high level. In fact, the low resolution (e.g., 64×64) intends to capture the overall style of the illustration (e.g., tone, large outlines). The medium resolution (e.g., 128×128) is responsible for extracting structural details (e.g., clothing lines). And the high resolution (e.g., 256×256) seizes textural details (e.g., brushstroke texture). The discriminator also corresponds to these three scales, determining whether the generated image at each scale resembles the illustration, avoiding loss of main details.

Depthwise separable convolution. This is a tool that makes the model lighter. Traditional convolution is computationally intensive and cannot run smoothly on edge devices (e.g., tablets). Depthwise separable convolution separates detail filtering and channel integration, reducing computation by 90% without affecting generation quality.

The logic of the Multi-scale CycleGAN model is as follows. To begin with, the core architecture adopts a dual-generator bidirectional mapping design. Generator G_XY is responsible for the forward style transfer from X-domain photos to Y-domain illustrations, while generator G_YX undertakes the reverse conversion from Y-domain illustrations to X-domain photos. Through a cyclical verification mechanism between pictures and illustration, it makes sure that the main content of the image cannot lost during the style translation process. Secondly, the method introduces a multi-scale supervision design, creating three generators with three kinds of distinct resolutions, each outputting illustration features at the corresponding scale. In the meantime, three matching discriminators are configured to determine the style authenticity of the features at their respective scales. Through multi-scale joint supervision, it is ensured that the generated results satisfy the requirements of overall style consistency while completely retaining detailed content. Additionally, the model deploys depthwise separable convolution instead of traditional convolution before. While producing 256×256 high-definition illustrations and maintaining rich details, it considerably decreases the computational complexity of the model, enabling it to work efficiently on edge devices such as tablets. Ultimately, it achieves a dual optimization of style translation effects and device deployment feasibility.

Researchers often choose the ArtBench dataset (32×32, 256×256 resolution), and use FID (35.71% lower than traditional CycleGAN) and IS (1.852 higher) as indicators. Its merits are better content preservation, steady training, and no need for paired data. Its demerits are insufficient capture of complex textures, limited style diversity, and purely applicable to illustration scenarios.

3 Applications

These technologies have already established clear application scenarios in fields such as digital art, portrait customization, commercial design, and entertainment.

DRB-GAN, with its general style transfer capabilities, can transfer the styles of artists such as Picasso and Ukiyo-e to any content image, providing designers with diverse artistic inspiration. DualStyleGAN supports 1024 × 1024 high-resolution portrait stylization and is widely used in personalized headshot customization, advertising portrait design, and film and television character prototype generation. Its precise structure and color control can meet users' refined needs for portrait style. WeditGAN's few-sample transfer capability is suitable for scenarios such as reusing styles from a small number of museum collections and quickly generating brand-specific styles, significantly reducing the style transfer cost in small dataset scenarios. Multi-scale CycleGAN performs outstandingly in the field of illustration creation, enabling the conversion of real images to illustration styles and improving the creation efficiency of comics, picture books, etc.

4 Discussion

4.1 Limitations

Existing applications still have significant problems. In terms of detail processing, DualStyleGAN is prone to losing non-facial details (such as hats and backgrounds), and Multi-scale CycleGAN is insufficient in capturing complex textures such as oil painting brushstrokes and ink wash. With respect to data dependence, DualStyleGAN is greatly affected by dataset preferences (such as difficulty in handling anime curly hair), and WeditGAN's transfer effect decreases when there is a large difference between the source and target domains (such as landscape → abstract art). As for generalization, DRB-GAN is poorly suited to unnatural styles, and WeditGAN has restricted adaptability to unseen styles. Concerning deployment cost, the high-resolution generation of DRB-GAN and DualStyleGAN is resource-intensive, and although Multi-scale CycleGAN resorts to a lightweight design, it remains to require optimization for large-scale applications.

These four technologies face four principal challenges in the field of image style transfer. The first challenge is insufficient generalization of unseen styles. Models such as DRB-GAN and WeditGAN have weak ability to transfer styles outside the training set, making it difficult to adapt to niche art styles or cross-domain style requirements. The second difficulty is balancing details and content. DualStyleGAN's loss of non-facial details and the inadequacy of Multi-scale CycleGAN in capturing complex textures reflect the problem of models in keep a well balance between style fidelity and detail preservation in style transfer. The third challenge is limitations in data efficiency. DRB-GAN requires millions of data points, and while WeditGAN supports few samples, it depends on domain relevance, and transfer in zero-sample scenarios has not yet been achieved. The fourth obstacle is insufficient real-time capability and lightweight design. The 1024×1024 generation of DualStyleGAN and the complex structure of DRB-GAN make it difficult to deploy on edge devices such as mobile phones and tablets, and satisfy the real-time style transfer requirements.

4.2 Future works

To address these challenges, future prospects can focus on four directions. In terms of cross-modal fusion, combining cross-modal models such as CLIP to achieve "text description → style transfer" (e.g., city photos in the style of Van Gogh's Starry Night), improving the model's generalization and style controllability for unseen styles. As for detail enhancement, adding a non-facial attention refinement module to DualStyleGAN and introducing a super-resolution texture enhancement unit to Multi-scale CycleGAN to optimize the quality of fine-grained detail transfer. Concerning data efficiency optimization, integrating meta-learning and contrastive learning to improve the adaptability of models and reduce reliance on labeled data is feasible. Regarding to lightweight deployment, using Neural Architecture Search (NAS) to design efficient networks, combined with quantization and pruning techniques, to achieve real-time deployment of high-resolution models such as DRB-GAN and DualStyleGAN, while extending to more complex scenarios such as 3D model style transfer and video style transfer (optimizing temporal consistency), driving the technology from static images to multi-modal content.

5 Conclusion

In conclusion, this paper analyses the core theories, frameworks and working logic of these four technologies including DRB-GAN, DualStyleGAN, WeditGAN and Multi-scale CycleGAN. Given its advantages and disadvantages in real experiments and applications, this article demonstrates the potential problems and gives some advice on the future work. Early basic GANs (such as CycleGAN and StyleGAN) achieved unpaired data transfer and high-resolution generation, but suffered from problems such as single task and data dependency. DRB-GAN broke through the limitation of single task and achieved unified style transfer for collective styles. DualStyleGAN focuses on specific domains and optimizes high-resolution control of portraits. WeditGAN solves the data limitation and achieves efficient transfer with few samples. Multi-scale CycleGAN targets scene details and improves the quality of illustration transfer. Overall, it shows an evolutionary trend of "general → segmented", "extensive → refined" and "high data dependency → data efficiency". In summary, there is no "one-size-fits-all" solution among the four technologies, and the choice should be made based on the specific application scenario: DRB-GAN is preferred for general art style transfer, DualStyleGAN for high-resolution portrait customization, WeditGAN for low-sample cross-domain scenarios, and Multi-scale CycleGAN for illustration creation. Future technological breakthroughs need to be promoted in four directions: generalization, detail control, data efficiency and lightweight design. Through cross-modal fusion, detail enhancement, and efficient network design, style transfer technologies can be driven from laboratory researches to large-scale industrial applications, while expanding multi-modal content transfer scenarios to further unleash the application value of the technology in digital art, entertainment, commercial design, and other fields.

References

1. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems, 27. (2014)
2. Aggarwal, A., Mittal, M., & Battineni, G.: Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights, 1(1), 100004. (2021)
3. Wang, R.: Research on image generation and style transfer algorithm based on deep learning. Open Journal of Applied Sciences, 9(08), 661. (2019)
4. Pang, Y., Lin, J., Qin, T., & Chen, Z. Image-to-image translation: Methods and applications. IEEE Transactions on Multimedia, 24, 3859-3881. (2021)
5. LAI, C.-Y.: Image generation models based on style transfer and their applications in the medical field [Master's thesis]. Zhejiang Gongshang University. (2021) https://doi.org/10.27462/d.cnki.ghzhc.2024.000163
6. Xu, W., Long, C., Wang, R., & Wang, G.: Drb-gan: A dynamic resblock generative adversarial network for artistic style transfer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6383-6392). (2021)
7. Yang, S., Jiang, L., Liu, Z., & Loy, C. C.: Pastiche master: Exemplar-based high-resolution portrait style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7693-7702). (2022)
8. Duan, Y., Niu, L., Hong, Y., & Zhang, L.: Weditgan: Few-shot image generation via latent space relocation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 38, No. 2, pp. 1653-1661). (2024)
9. Liang, Y., & Yan, Y.: The style transfer model of illustration images based on multi-scale CycleGAN. International Journal of Information and Communication Technology, 26(7), 1-16. (2025)
10. Ma, X.: A comparison of art style transfer in Cycle-GAN based on different generators. In Journal of Physics: Conference Series (Vol. 2711, No. 1, p. 012006). IOP Publishing. (2024)