

Exploring the Evolution and Future Trends of Image Style Transfer Techniques: A Comprehensive Literature Review

¹Zainab Hussain Mahdi, ²Kassem Danach, ³Marwa Hamade

¹Faculty of Engineering, Islamic University of Lebanon, Lebanon

^{2,3}Faculty of Business Administration, Al Maaref University, Lebanon

Abstract - The rapid evolution of image style transfer techniques, a fascinating intersection of art and technology, represents a significant area of research in the domain of computer vision and machine learning. This comprehensive literature review critically examines the development and progression of various methodologies in image style transfer, tracing their evolution from initial neural algorithm-based approaches to more advanced generative adversarial networks (GANs) like CycleGAN, DiscoGAN, and StarGAN. By scrutinizing studies ranging from foundational works to recent innovative approaches, this paper aims to provide a thorough understanding of the techniques, their effectiveness, and the challenges they address. Emerging trends, such as the incorporation of domain-specific information, attention mechanisms, and human perception-inspired loss functions, are highlighted, reflecting the field's shift towards more context-aware and semantically meaningful image translations. The review identifies gaps in systematic comparative studies, particularly concerning the efficacy of CycleGANs against other prevalent methods, indicating areas ripe for future research. This paper serves as a foundational guide for understanding current image style transfer techniques and sets the stage for exploring new horizons in the blending of artistic creativity and technological innovation.

Keywords: Image Style Transfer, Generative Adversarial Networks (GANs), CycleGAN, Neural Algorithms, Computer Vision, Machine Learning.

I. Introduction

The realm of image style transfer represents a fascinating intersection of art and technology, where the essence of artistic creativity is seamlessly blended with the precision of computational algorithms. This field, evolving rapidly due to advancements in machine learning and computer vision, delves into the challenge of transforming the stylistic elements of one image onto the content of another, thereby creating a novel visual experience that retains the essence of the original while adopting an entirely different aesthetic.

The concept of image style transfer is not merely an artistic endeavor but also an exploration into the capabilities of neural networks and their ability to mimic complex human cognitive functions such as creativity and perception. From the pioneering works utilizing deep convolutional neural networks for artistic style transfer to the latest innovations in Generative Adversarial Networks (GANs) and CycleGANs, the field has witnessed a surge in research interest and practical applications. These techniques have not only enabled the recreation of famous artistic styles on everyday images but have also opened avenues for advancements in design, entertainment, and even therapeutic applications.

This paper aims to provide a comprehensive overview of the evolution and current trends in image style transfer techniques. We delve into the foundational methodologies that laid the groundwork for the field, explore the myriad of ways these techniques have been adapted and enhanced over the years, and highlight the emerging trends that are shaping the future of image style transfer. By examining the key contributions and limitations of various approaches, from Neural Style Transfer to CycleGANs, we aim to offer a nuanced understanding of how these technologies work and their potential implications in both art and technology. As we explore this dynamic and rapidly evolving field, we underscore the continuous interplay between artistic creativity and technological innovation that drives the advancement of image style transfer.

II. Problem Statement

The field of image style transfer is at the forefront of blending artistic creativity with technological innovation, offering a unique platform where the styles of various artists can be applied to diverse content images using advanced computational methods. However, the journey of image style transfer from its conceptualization to its current state has encountered numerous challenges and limitations, which need to be addressed to further its advancement.

One of the primary challenges in this domain is the effective and accurate translation of artistic styles onto different content images. This involves not only the replication of visual styles but also the maintenance of the structural integrity and semantic meaning of the original content. Techniques like Neural Style Transfer and Generative Adversarial Networks have made significant strides, yet they often struggle with maintaining a balance between style adaptation and content preservation.

Additionally, the existing algorithms generally require high computational resources and can be time-intensive, making them less accessible for real-time applications or users with limited computational capabilities. This limitation hinders the widespread adoption and practical usage of style transfer technologies in various fields such as real-time design, interactive media, and online content creation.

Moreover, most current methods focus primarily on transferring the styles of famous artworks or predefined styles, limiting the scope of creative exploration. The need for algorithms that can understand and adapt to a wider range of artistic styles, including those that are less defined or more abstract, is yet to be fully realized.

Furthermore, the field faces challenges in terms of data availability and diversity. Many style transfer models are trained on limited datasets, which may not encompass the wide variety of artistic expressions across different cultures and historical periods. This lack of diversity can lead to biases in style translation and restrict the applicability of these models to a broader range of artistic styles.

This paper aims to address these challenges by exploring advanced methodologies in image style transfer, examining their effectiveness in various scenarios, and proposing enhancements to existing techniques. By doing so, it seeks to contribute to the development of more efficient, versatile, and inclusive image style transfer technologies that can democratize artistic expression and expand the boundaries of creative digital media.

In this section, we will take a step by step approach to reviewing studies on image to image translation. We will start with the work of J. Y. Zhu et al. (2017) which laid the foundation for this field. Subsequent research efforts have tackled different challenges, including disentangling representations, translating across domains, and incorporating mechanisms inspired by the human visual system [1].

An important study by L. A. Gatys, A. S. Ecker and M. Bethge (2015) introduced a neural algorithm for creating images based on Deep Neural Networks. They were inspired by how content and style interact in human artistry [2]. By

leveraging neural representations, their algorithm is capable of separating and recombining content and style elements in a way that mimics the process of humans. While this study was groundbreaking, it focuses on the interplay between content and style and does not fully explore all the complexities involved in artistic creation.

Additionally A. Radford, L. Metz and S Chintala (2015) introduced Deep Convolutional Generative Adversarial Networks (DCGANs) as a type of CNNs used for unsupervised learning [3]. They demonstrated that DCGANs are effective in learning hierarchical representations ranging from object parts to scenes. The adversarial pair that has undergone training produces designs and learned features that can be used for different tasks. This highlights their usefulness, in both supervised learning and generative modeling. The research acknowledges some forms of model instability noting that while models are trained for longer periods, there might be occasional collapse of a subset of filters to a single oscillating mode.

In their 2017 study, J. Y. Zhu et al introduced a technique called CycleGAN for translating images without paired data [1]. They used cycle consistency losses to achieve style transfer and photo enhancement. However, they faced difficulties when it came to handling changes, in the aspects of the images. While CycleGAN excelled in transforming colors and textures it, fell short in terms of handling geometric changes. This limitation can be attributed to the architecture of the generators used and the characteristics of the dataset employed. Despite these challenges, this method contributes significantly to advancing unsupervised image to image translation by demonstrating the value of utilizing unpaired data. The architecture of CycleGAN is depicted in Figure (2.1).

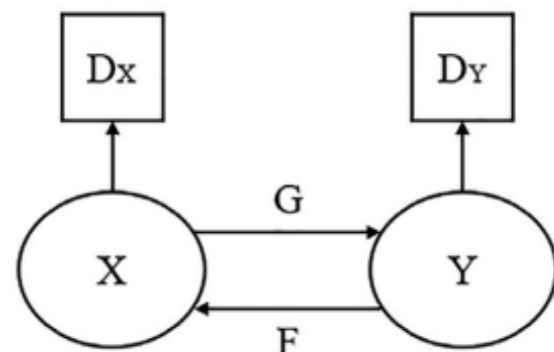


Figure 2.1: Architecture of CycleGAN

In the study conducted by P. Isola, J. Y. Zhu T. Zhou and A. A. Efros in 2017, they delved into the use of conditional adversarial networks, for various image to image translation tasks. The researchers demonstrated the effectiveness of this

approach in synthesizing photos and reconstructing objects and colorizing images [4]. While acknowledging its potential in different graphical outputs, the study also highlighted the need for further exploration to assess its scalability, real time applications, and address any potential biases that may arise in synthesized outputs.

Another notable work by Kim et al. (2017) introduced DiscoGAN as a generative adversarial network that allows for learning cross domain relations without explicit pair labels [5]. This method excels at transferring styles between different domains and across diverse datasets while preserving important attributes. Although DiscoGAN shows promise, it is currently mostly focused on image domains; hence, there is room for exploration to extend its capabilities to handle mixed modalities like text and image.

Y. Chen et al. (2017) work modification of CycleGAN, incorporates adversarial, cycle-consistency, and identity losses to find representation disentanglement similar to InfoGAN [6]. Even though CycleGAN produces realistic images, it does have limitations stemming from the image to image framework or cycle consistency loss or insufficiency of dataset. Despite trying to forecast latent values, disentangling still faces limitations indicating the need for more enhancements. The research acknowledges realism but also acknowledges shortcomings stemming from the small and contaminated dataset, which affects performance in comparison to other GAN models. Attempts were made to improve performance by incorporating latent value prediction. Although they have been showing some success, space for more improvement is still needed.

In their paper, Y. Choi et al. (2018) introduced a model called StarGAN which's a generative adversarial network designed for image to image translation across multiple domains [7]. This model stands out for its ability to produce high quality images and its flexibility, in handling datasets with varying domain labels, due to its scalability and generalization capabilities. However, it is worth noting that StarGAN may have limitations when dealing with complex or specialized domains that were not specifically addressed in their empirical evaluation. Additionally, there could be practical implementation challenges or computation requirements that need to be considered for real world implementation. Figure (2.2) provides an overview of the two components of StarGAN; the discriminator D and the generator G .

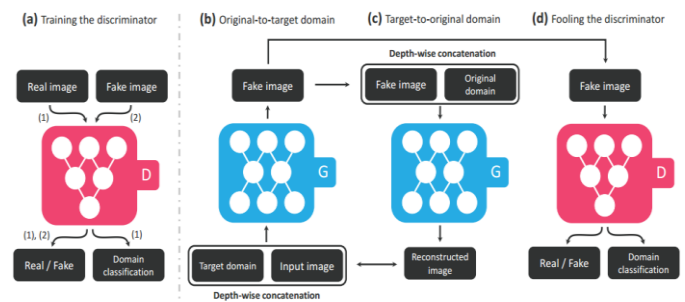


Figure 2.2: An overview of StarGAN network

In another study, Marra et al. (2019) investigated the presence of fingerprints in images generated by adversarial networks (GANs) [9]. They found that each GAN leaves behind fingerprints to how real world cameras mark images with identifiable characteristics. The researchers conducted source identification experiments which demonstrated the potential of these fingerprints providing a means to differentiate between images generated by different networks. The study offers experimental evidence supporting GAN fingerprints. However, the study also highlighted the need for investigation into factors such as network architecture and parameters. It acknowledged that more research is required to validate their results and determine if these findings hold true across datasets. Additionally, they emphasized the importance of developing methodologies for future directions, in this area.

In their 2019 paper, Y. Wu, R. Zhang and K. Yanai introduced an approach called Progressive Growing CycleGAN (PG-Att-CycleGAN) that combines the benefits of Progressive Growing GAN and CycleGAN to achieve image to image translation [10]. The key innovation of this model lies in its ability to progressively scale input sizes from 256x256 to 1024x1024 resulting in improved image quality and stable training. By incorporating attention blocks, the model further enhances domain transfer capabilities and ensures processing of high scale images. Notably, PG-Att-CycleGAN outperforms approaches such as CycleGAN, PG CycleGAN and DiscoGAN by reducing artifacts and generating images with natural backgrounds. Quantitative evaluations using VGG-16 demonstrate its superior performance across most translation tasks. However, it is worth noting that the study does not extensively discuss limitations or computational considerations, leaving room for more exploration in these areas.

In their 2020 paper, W. Zheng, L. Yan, C. Gou and F. Y. Wang introduced JND-GAN as a method for translating images to images based on the human visual system [11]. This model utilizes a loss function called Just Noticeable Difference (JND) to generate realistic images for various tasks even without paired training data. It demonstrates superior performance in fidelity and emphasizing human vision

inspired consistency. While it excels in perceptual metrics, there is room for exploration and improvement regarding its semantic accuracy. Figure (2.3) illustrates the network architecture of the proposed network (i.e. JND-GAN).

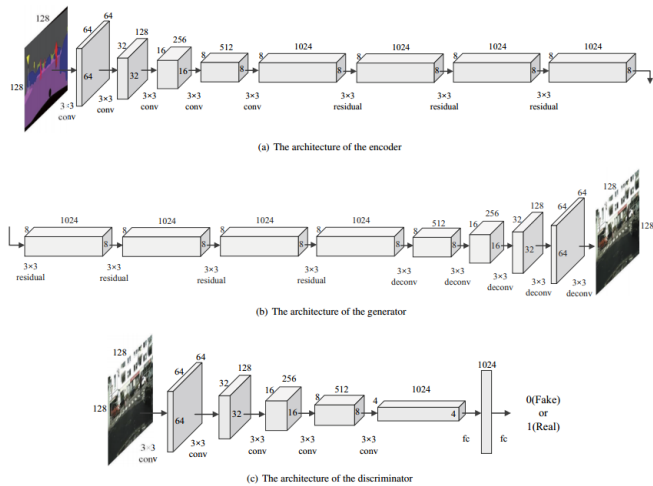


Figure 2.3: Network architecture of JND-GAN

E. Batziou et al. (2021) came up with a new way to combine CycleGAN and FABEMD for image style transfer. They modified the cycle consistency loss by taking into account the differences, between Intrinsic Mode Functions (BIMFs) [12]. The experimental results demonstrated the effectiveness of this method producing outcomes that are comparable to state of the art techniques. This innovative approach not only improved the quality and accuracy when compared to the original CycleGAN and a referenced method but also introduced frequency components in the training process that could be beneficial for style transfer. However, it is important to explore and consider frequency components in the training process beyond just style transfer. Figure (2.4) illustrates the framework for style transfer using Cycle-Consistent Adversarial.

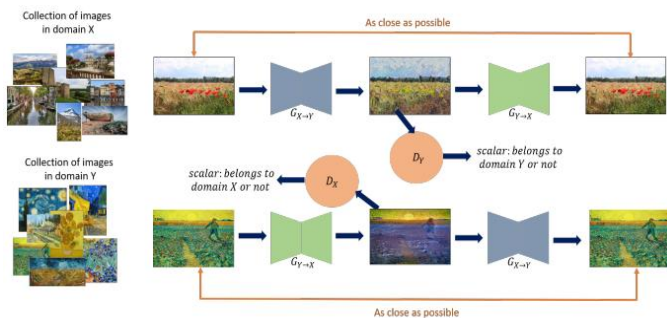


Figure 2.4: Style transfer framework with Cycle-Consistent Adversarial

In a study by Z. Chen et al. (2021), they presented a method called PREGAN that focuses on style translation in weakly paired images with pose errors [13]. The approach

involves using adversarial training along with intentional random pose transformations and with a differentiable non-trainable pose estimator. The effectiveness of PREGAN has been validated through experiments, including tasks like classification and object detection. It particularly excels in translating weakly paired images and thus enhancing interpretability with random poses. The differentiable pose estimator helps to separate image style and pose while adversarial training enhances network learning. However, it is recommended to conduct robustness testing for pose estimation under varying conditions.

Another study by T. Fontanini, F. Botti, M. Bertozzi and A. Prati (2022) aimed to address a limitation in image to image translation related to the cycle consistency loss promoting shortcuts [14]. They introduced an attention consistency loss during training which ensures focus is maintained throughout the translation cycle. This method consistently improved both qualitative and quantitative aspects of the translations across different datasets, including style transfer tasks. While it successfully enhanced accurate translations, there were instances where attention transfer could worsen failures in cases like horse to zebra translations using CycleGANs. Despite these occasional effects, the attention consistency loss consistently improved overall results and proved its efficacy in maintaining focus during the translation cycle. The paper proposes that future experiments should involve testing with other architectures and extending attention maps to other tasks, beyond just image to image translation.

In their work, X. Rang et al. (2022) introduced Cycle-DPN-GAN, an innovative approach for unsupervised multistyle image transfer [15]. By incorporating the PONO-DMS module to enhance information and utilizing MS-SSIM loss to improve perceptions, the model demonstrates outstanding performance across various datasets. It effectively tackles challenges like edge blurring and object loss, which results in high visual quality compared to existing approaches. While demonstrating strengths in several evaluations, there is still a space for improvement in refining the network structure and exploring applications in other domains such as Chinese ink paintings and cartoon style images.

In a different study by Wang, Wang and Chen (2022), a new approach called ESA-CycleGAN was introduced [16]. This approach combines edge features and self-attention in a cycle adversarial network for style transfer. The architecture consists of a generator and a discriminator and an edge feature extraction network that utilizes self-attention to capture global features. By enhancing detail processing, ESA-CycleGAN achieves superior Inception Score (IS) and Fréchet Inception Distance (FID) values compared to existing models, across

four datasets. It excels in preserving image details and improving quality through its effective global feature capture. However, the study also acknowledges limitations when dealing with complex geometric changes, which highlights the importance of further exploring the models robustness across diverse datasets. Please refer to Figure (2.5) for the generator structure of ESA CycleGAN.

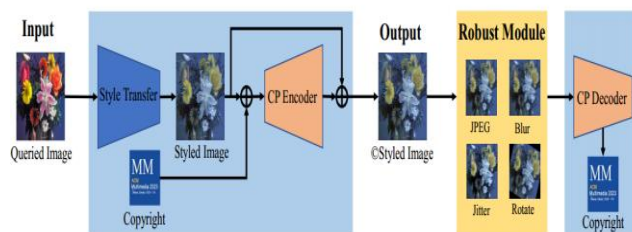


Figure 2.6: An overview of the proposed plug-and-play method

In their study, Abbasnejad et al. (2023) introduced SCONE-GAN, a method that utilizes graph convolutional networks, for end to end image translation. This approach effectively captures the relationships between objects and maintains the structure of the images [18]. Through incorporating a style reference image, SCONE-GAN enhances diversity by maximizing mutual information. The results obtained by testing it on four datasets confirm its ability to generate varied realistic and scenery images, thus surpassing state of the art techniques. However, it is worth noting that the reliance on contrastive learning and segmentation frameworks might affect the quality of outputs, particularly for images not specifically trained for the given task. This observation was made during experiments conducted with Cityscape and Monet2photo datasets. Nevertheless, SCONE-GAN makes contributions, towards improving both the quality and diversity of generated images.

Table (2.1) provides a summary of research in the domain of image style transfer, highlighting the year, researcher name, methodology, main contribution, and limitations of each method.

Table (2.1): Summary of research in the domain of image style transfer

Table 2.1: Summary of research in the domain of image style transfer

Year	Researchers	Methodology	Key Contributions	Limitations
2015	Gatys, Ecker, Bethge	Neural Algorithm of Artistic Style	Neural algorithm for artistic image creation and the limited exploration of diverse elements.	Limited exploration of diverse elements.
2016	Radford, Metz, Chintala	DCGANs	Introduction of DCGANs for unsupervised learning and the model instability with prolonged training.	Model instability with prolonged training.
2017	Zhu et al.	CycleGAN	CycleGAN for unpaired image translation and the challenges in geometric transformations.	Challenges in handling diverse transformations.

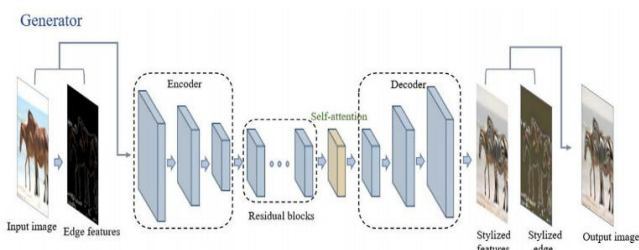


Figure 2.5: Structure of the generator of ESA-CycleGAN

In their study published in 2023, D. Wang et al. Introduced a method for protecting copyrights in applications that utilize networks (DNNs) with a focus on style transfer [17]. Their approach involves encoding copyright information into defensive perturbations, which ensures that the decoding process remains robust when faced with distortions like JPEG compression. Both objective and subjective experiments were conducted to validate the effectiveness of this method and real world tests on social media platforms confirmed its ability to accurately extract information from encoded images. However, it is important to note that while this method proves robust for protecting copyrights in DNN based applications, there are still challenges to address such as distortion when decoding images and susceptibility to text watermarks. Further refinements are necessary to make it more applicable in other scenarios and improve its resilience, against watermarking techniques. For a representation of the proposed plug and play method, please refer to Figure (2.6).

2017	Isola, Zhu, Zhou, Efros	Conditional Adversarial Networks	Effective for various tasks and the limited discussion on scalability and biases.	Limited discussion on scalability and biases.
2017	Kim et al.	DiscoGAN	Effective in transferring styles between domains and focus on image domains.	Limited exploration beyond image domains.
2017	Chen et al.	Modified CycleGAN	Realistic image generation but limitations with a small dataset.	Suboptimal performance compared to other GAN models.
2018	Choi et al.	StarGAN	Superior image quality and the potential limitations in complex domains.	Practical implementation challenges and the computational requirements.
2018	Benaïm, Wolf	One-shot unsupervised cross-domain translation	Effective in diverse environments and the conservative for multi-one-shot adaptations.	Conservative for multi-one-shot adaptations.
2019	Marra et al.	GAN fingerprints	Potential value for forensic analyses and the preliminary results.	Preliminary results and the open questions regarding generalizability.
2019	Wu, Zhang, Yanai	PG-Att-CycleGAN	Reduction of artifacts and the lack of explicit discussion on limitations.	Lack of explicit discussion on limitations.
2020	Zheng, Yan, Gou, Wang	JND-GAN	Competitive in domain adaptation and the potential weaknesses in semantic accuracy tasks.	Potential weaknesses in semantic accuracy tasks.
2021	Batzou et al.	CycleGAN + FABEMD	Improved outcomes, introduction of frequency components.	Qualitative considerations on introducing frequency components.
2021	Chen et al.	PREGAN	Effective in downstream tasks and the potential sensitivity in pose estimation.	Potential sensitivity in pose estimation.
2022	Fontanini et al.	Attention consistency loss	Significant improvement and the occasional effects.	Occasional effects and the need for further exploration.
2022	Rang et al.	Cycle-DPN-GAN	Enhanced visual quality and need for a more sophisticated structure.	Need for a more sophisticated structure.
2022	Wang, Wang, Chen	ESA-CycleGAN	Superior scores and potential limitations in handling complex changes.	Potential limitations in handling complex changes.
2023	Wang et al.	DNN-based copyright protection	Robust against distortions and potential distortion in clean images.	Potential distortion in clean images and susceptibility to text watermarks.

2023	Abbasnejad et al.	SCONE-GAN	Effective in generating diverse scenery images and the output quality deterioration in specific cases.	Output quality deterioration in specific cases and the reliance on contrastive learning may affect robustness.
------	-------------------	-----------	--	--

III. Discussion of Methodologies Employed by Different Researchers

Different researchers in the field of image to image translation employ different methodologies that showcase both similarities and dissimilarities, to address the challenges in this field. For instance, J. Y. Zhu et al. (2017) and P. Isola, J. Y. Zhu T. Zhou and A. A. Efros (2017) utilize adversarial networks to achieve unpaired image translation. They rely on generators and discriminators within a training framework to learn the mapping between domains without paired data.

In contrast, S. Benaim and L. Wolf (2018) introduce variations in their approach by employing a one shot unsupervised cross domain translation method that involves adapting layers based on input samples and utilizing autoencoders. Although effective, this method differs significantly from traditional adversarial training, demonstrating the diverse techniques explored in this research community.

Another aspect addressed by Y. Chen et al. (2017) and other researchers is representation disentanglement, where they modify CycleGAN to explore disentangled representations. However, the specific techniques employed to achieve vary among these researchers highlighting the nuanced differences in implementation.

In a study conducted in 2018, Choi et al. introduced a method called StarGAN for multi domain translation. They highlighted its scalability and ability to handle translation

tasks, distinguishing it from one shot methods proposed by Benaim and Wolf. This shows the breadth of strategies employed in tackling translation challenges.

Furthermore, other researchers have focused on aspects within this field. For instance, Marra et al. (2019) explored GAN fingerprints, while Fontanini et al. (2022) introduced attention consistency loss to address shortcuts in unpaired translation. These specific focuses not only contribute to the variety of methodologies used but also highlight the nature of particular techniques.

Despite these differences, there is an acknowledgment among researchers regarding challenges and limitations in their methodologies. For example Y. Wu, R. Zhang and K. Yanai (2019) recognized potential model instability during training durations while D. Wang et al. (2023) acknowledged difficulties related to decoding clean images and susceptibility to text watermarks in their copyright protection approach.

In short, researchers working on image to image translation employ common elements like adversarial networks; however they also demonstrate distinct approaches such as unique training methods and a focus on specific challenges. These variations contribute to the variety of the research landscape in this evolving field. Table (3.1) highlights key strengths, weaknesses, and associated researchers for each approach in image to image translation.

Table 3.1: Weaknesses and strengths of each approach in image-to-image translation

Approach	Strengths	Weaknesses	Researchers
Adversarial Networks	Captures complex mappings without paired data	Challenges in handling geometric changes	J.-Y. Zhu et al. (2017), P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros (2017)
One-shot Unsupervised Cross-domain	Effective with limited training samples	Conservative for multi one shot adaptations and limited in complex scenarios	S. Benaim and L. Wolf (2018)
Representation Disentanglement	Aims for better control through latent representation	Challenges in achieving full disentanglement and affected by dataset limitations	Y. Chen et al. (2017)

Multi-domain Translation (StarGAN)	Scalable and generalizable	Potential limitations in complex domains and practical challenges	Y. Choi et al. (2018)
GAN Fingerprints	Unique fingerprints for source identification	Preliminary results and questions about generalizability and the lack of methodologies	Marra et al. (2019)
Progressive Growing GAN (PG-Att-CycleGAN)	Improved image quality, stability in training	Lack of explicit discussion on limitations and further exploration needed	Y. Wu, R. Zhang, and K. Yanai (2019)
Just-Noticeable-Difference (JND-GAN)	Emphasis on human vision inspired consistency	Potential limitations in semantic accuracy and the need for further exploration	W. Zheng, L. Yan, C. Gou, and F.-Y. Wang (2020)
CycleGAN with FABEMD	Efficient style transfer and improved results	Considerations regarding frequency components and the exploration beyond style transfer	E. Batziou et al. (2021)
PREGAN for Style Translation	Effective in weakly paired images with pose errors	Need for robustness testing and effective in downstream tasks	Z. Chen et al. (2021)
Attention Consistency Loss (ESA-CycleGAN)	Improved focus maintenance in translation cycle	Potential exacerbation of failures, consistent improvement despite occasional effects	T. Fontanini, F. Botti, M. Bertozzi, and A. Prati (2022)
Cycle-DPN-GAN	Enhanced structural information, improved visual perceptions	Areas for improvement in network structure, potential applications in specific tasks	X. Rang et al. (2022)
SCONE-GAN	Graph convolutional networks for end to end translation	Reliance on contrastive learning, segmentation and the potential impact on output quality	I. Abbasnejad et al. (2023)
Plug-and-Play Copyright Protection (DNN-based)	Robust encoding of copyright information and effective against distortions	Challenges in decoding clean images and the susceptibility to text watermarks and refinements needed	D. Wang et al. (2023)

IV. Gaps in the Literature

Image style transfer is a combination of art and technology. It involves transforming images to embody particular artistic styles while retaining their original content, thus creating a connection between creativity and technological innovation.

However, there are still gaps in the research on this subject. The effectiveness of CycleGAN for image style transfer has not been thoroughly evaluated, in comparison to methods like DCGANs. Additionally there has been limited

exploration of variations within the CycleGAN framework indicating a need for more investigation.

Existing studies have not systematically assessed how well CycleGAN can capture and transfer styles or compared it comprehensively with conventional methods. Furthermore, the impact of different architectural choices on the efficiency of style transfer has not been extensively explored.

This paper aims to address these gaps and contribute to the advancement of image style transfer in both art and technology domains. The research strives to uncover the strengths, weaknesses and unique capabilities of CycleGAN

compared to methods like DCGANs and Neural Style Transfer.

Essentially, our primary objective is to investigate how different architectural choices influence the effectiveness and efficiency of image style transfer. We will delve into understanding CycleGANs capabilities by making comparisons with existing methods and examining how architectural variations, within the CycleGAN framework, affect its performance. These endeavors are driven by the changing demands of art, and design, and entertainment, and diverse industries, which play a role in pushing forward the frontiers of computer vision and machine learning.

V. Emerging Trends

In the field of unpaired image to image translation, there are noticeable emerging trends and shifts in focus. Recent studies suggest a growing emphasis on incorporating domain-related information for better translation. For example, Abbasnejad et al. (2023) introduced semantic contrastive learning based generative adversarial networks (SCONE GAN) which highlights the use of graph convolutional networks to capture object dependencies, and maintain image structure in translation. This trend reflects an understanding of the importance of context aware approaches in achieving realistic and semantically meaningful results.

Furthermore, attention mechanisms are increasingly being integrated into these approaches. Wu, Zhang and Yanai (2019) demonstrated this with their work on Progressive Growing CycleGAN (PG-Att-CycleGAN) where they combined Progressive Growing GAN with CycleGAN and incorporated attention blocks to enhance domain transfer and improve training stability. Attention mechanisms proved effective in addressing problems like checkerboard artifacts thus enhancing the quality of high scale image translation.

Another noteworthy trend involves exploring loss functions that take into account human inspired perceptual considerations. The introduction of the JND loss, in JND-GAN (Zheng et al., 2020), exemplifies this shift by aiming to prioritize perceptual fidelity inspired by human visual systems. This trend highlights a shift away from traditional adversarial and cycle consistency losses, indicating the desire to make the generated images more closely align with the perceptual expectations of humans.

These emerging trends show that the field is maturing, as researchers increasingly recognize the importance of considering context and attention mechanisms and human perception when improving unpaired image to image translation models.

VI. Conclusion

The exploration of image style transfer techniques in this comprehensive literature review reveals a dynamic and evolving field, characterized by a diversity of approaches and methodologies. We have observed significant advancements from the foundational work of Gatys et al. and Zhu et al., leading to more sophisticated methods like CycleGAN, DiscoGAN, and StarGAN, each addressing unique aspects of style transfer challenges. The integration of domain-related information, attention mechanisms, and human perception-inspired loss functions marks a notable shift towards more context-aware, realistic, and semantically meaningful image translations.

The research community's continuous efforts to refine these techniques underscore the importance of image style transfer not only in artistic and creative endeavors but also in practical applications across various industries. However, despite these advancements, our review identifies gaps in systematic comparisons of different methodologies, particularly in the effectiveness of CycleGANs in comparison to other methods. This gap points to the need for further empirical studies and experimentation to fully understand the strengths, limitations, and potential of each approach.

Emerging trends in the field, such as semantic contrastive learning and graph convolutional networks, indicate a promising direction towards more nuanced and context-sensitive image translation methods. These developments are crucial in meeting the ever-increasing demands for quality and versatility in applications ranging from art generation to entertainment and beyond.

In conclusion, while the field of image style transfer has made significant strides, it remains a fertile ground for innovation and exploration. The trends and gaps identified in this review not only provide a comprehensive understanding of the current state of the field but also pave the way for future research that can further push the boundaries of creativity and technology in image style transfer.

REFERENCES

- [1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223-2232.
- [2] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.
- [3] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional

- generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125-1134.
- [5] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *International conference on machine learning*, 2017: PMLR, pp. 1857-1865.
- [6] Y. Chen, M. Tang, C. W. Y. AG, and Y. V. Wang, "CycleGAN and Representation Disentanglement."
- [7] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789-8797.
- [8] S. Benaim and L. Wolf, "One-shot unsupervised cross domain translation," *advances in neural information processing systems*, vol. 31, 2018.
- [9] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "Do gans leave artificial fingerprints?," in *2019 IEEE conference on multimedia information processing and retrieval (MIPR)*, 2019: IEEE, pp. 506-511.
- [10] Y. Wu, R. Zhang, and K. Yanai, "Attention Guided Unsupervised Image-to-Image Translation with Progressively Growing Strategy," in *Asian Conference on Pattern Recognition*, 2019: Springer, pp. 85-99.
- [11] W. Zheng, L. Yan, C. Gou, and F.-Y. Wang, "JND-GAN: human-vision-systems inspired generative adversarial networks for image-to-image translation," in *ECAI 2020*: IOS Press, 2020, pp. 2816-2823.
- [12] E. Batziou, P. Alvanitopoulos, K. Ioannidis, I. Patras, S. Vrochidis, and I. Kompatsiaris, "Cycle-consistent adversarial networks and fast adaptive bi-dimensional empirical mode decomposition for style transfer," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021: IEEE, pp. 2360-2367.
- [13] Z. Chen *et al.*, "Pregan: Pose randomization and estimation for weakly paired image style translation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2209-2216, 2021.
- [14] T. Fontanini, F. Botti, M. Bertozzi, and A. Prati, "Avoiding Shortcuts in Unpaired Image-to-Image Translation," in *International Conference on Image Analysis and Processing*, 2022: Springer, pp. 463-475.
- [15] X. Rang, C. Ma, Q. Zheng, and G. Zhao, "A Generative Adversarial Network for Multistyle Unsupervised Image Transfer Based on Dynamic Position Normalization," *IEEE Access*, vol. 10, pp. 96284-96295, 2022.
- [16] L. Wang, L. Wang, and S. Chen, "ESA-CycleGAN: Edge feature and self-attention based cycle-consistent generative adversarial network for style transfer," *IET Image Processing*, vol. 16, no. 1, pp. 176-190, 2022.
- [17] D. Wang, W. Yao, T. Jiang, W. Zhou, L. Lin, and X. Chen, "A Plug-and-Play Defensive Perturbation for Copyright Protection of DNN-based Applications," *arXiv preprint arXiv:2304.10679*, 2023.
- [18] I. Abbasnejad *et al.*, "SCONE-GAN: Semantic Contrastive Learning-Based Generative Adversarial Network for an End-to-End Image Translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1111-1120.

Citation of this Article:

Zainab Hussain Mahdi, Kassem Danach, & Marwa Hamade. (2024). Exploring the Evolution and Future Trends of Image Style Transfer Techniques: A Comprehensive Literature Review. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(6), 71-80. Article DOI <https://doi.org/10.47001/IRJIET/2024.806009>
