A system for automatically transferring image styles that uses the Gen AI algorithm

Vidya D. Argade[1], Ramalakshmi[2], S Asma Begum[3], Harshitha H[4], Vidhya S G[5], Guru Hugar[6], Rudresha S[7]

[1]Department of Computer Science Engineering, Dhole Patil College of Engineering, Vitthal Nagar, Wagholi, Kharadi, Pune, Maharashtra, Karnataka, India

[2]Department of Computer Science Engineering, Cambridge Institute of Technology, K.R.Puram,Bengaluru, Karnataka, India

[3]Department of Computer Science Engineering, Atria Institute of technology, Bengaluru, Karnataka, India

[4]Department of Computer Science Engineering, Chandigarh Engineering College, Chandigarh Group of Colleges Jhanjeri, Mohali - 140307, Punjab, India

[5]Department of Information Science and Engineering , BGS Institute of Technology, Adichunchanagiri University, BG Nagara, Mandya, Karnataka , India

[6]Department of Computer Science and Engineering, AGM Rural College of Engineering and Technology Varur Hubballi, Karnataka, India

[7]Department of Mechanical Engineering, University of Visvesvaraya College of Engineering, Bengaluru, India

Abstract:

Creative fields have undergone a significant change due to developments in generative AI technologies. Gen AI is employed across various commercial and industrial applications to enhance business results. Image style transfers represent one of the lucrative applications. This job aims to transfer an image from one domain to another, referred to as the target domain, utilizing styling information from the target domain. This research presents a new picture transfer system that allows for strong and adaptable style transitions between two different domains. The proposed design employs a visual transformer to acquire picture information from the original image and the style image. In order to carry out image style transfer while enhancing quality, these feature embeddings are input into cyclic GANs that can operate in both forward and backward modes.„Um die Wirksamkeit der empfohlenen Vorgehensweise zu zeigen, wird auch ein Vergleich mit anderen hochaktuellen Methoden vorgenommen. The results demonstrate that the proposed model transfers visual styles more effectively. Moreover, other transfer learning methods may be explored in future research to reduce the training duration of the model.

Keywords: Image style transfer, visual transformer, Cyclic GAN, FSIM, MMSIM, feature embeddings, style information

1.Introduction `

Generative artificial intelligence (Gen AI) is causing enormous transformations in nearly all sectors by extending limits further than ever. Numerous fields are presently moving toward a new era of creativity and efficiency (Chan et al., 2023). These systems evolve through the application of advanced machine learning (ML), deep learning (DL), and artificial intelligence (AI) methods. Due to their ability to generate unique artistic forms, the use of AI-based algorithms for creating original and innovative content has grown in recent times.Im Jahr 2024, et al. Sharanya Generative AI fundamentally relies on machine learning techniques to analyze large data sets and identify patterns and structures. This allows the AI system to generate new content based on the knowledge it has acquired. The method involves using existing creative works, such as paintings or music, to train the AI model. This model is then used to generate new and original creations.Consequently, the creation of original content, the optimization of workflows, and the uncovering of previously unutilized potential are all now possible. A technology known as Gen AI, which employs generative models (Leong et al., 2024), is transforming how individuals perceive creative expression. To generate synthetic data that mirrors the patterns in the actual data learned by the computers, systems and computers are trained effectively with a diverse set of datasets.This indicates that Gen AI can generate and incorporate creative writing, music, graphics, and even films.Its capacity for use in innovative applications is the most enticing aspect of Gen AI. Gen AI can be utilized by web designers, advertising agencies, content creators, and various other digital platforms. By helping artists and designers create original and imaginative works, they challenge the boundaries of conventional and traditional artistic forms. Gen AI can be applied in various areas, such as creating incredible music and art, enhancing corporate profits, and improving workflow. Gen AI can be utilized in business and industry sectors like marketing and communications to develop more profitable products by examining consumer data for intricate and concealed patterns. Many businesses typically use Gen AI to automate certain tasks and processes, thereby enhancing efficiency and reducing manufacturing and operational costs (Huang et al., 2023).A more contemporary version of Gen AI is employed to create interactive features and generate content that is tailored to the individual by analyzing user preferences and behavior. This Gen AI feature will enhance the application's popularity and usability by optimizing the user experience. Within the design industry, where Midjourney, ChatGPT, and

Gen AI enjoys great popularity owing to these attractive characteristics.While the primary aim of employing Gen AI in design work is to save time, its application goes beyond just that. For simple data analysis, unique visuals and text content are generated using Gen AI technologies. Their designs, which are tailored and customized to individual needs, are widely recognized. Although there are numerous benefits to using Gen AI technology, the absence of emotional depth and originality poses the greatest concern for artists and creators. This technology's extensive application is viewed as a threat to human employment and jobs.

While employing generative AI has notable benefits, worries exist regarding its potential impact on industrial innovation. There is concern that entertainment produced by artificial intelligence (AI) may not possess the emotional subtleties and creativity found in the work of human artists. Another concern is that AI could lead to greater automation in the creative sector, putting jobs at risk and causing human workers to leave (Takaffoli et al., 2024). leading to the biggest upheaval in the art business. It allows anyone who loves the art form to create impressive professional-quality work, even if they don't have much experience. This will help people create a new wave of innovation and creativity.

2. Utilization of generative AI in artistic expressions.

Gen AI tools are frequently used to produce creative artistic work by examining the textual descriptions provided to them (Ingerøyen et al., 2024). The artists can investigate new and inventive compositions, ideas, and styles that would not have been possible with traditional techniques by utilizing these tools. This creative democracy brings together a diverse array of original artistic expressions by seamlessly combining human ingenuity with artificial intelligence. Moreover, AI-powered creative tools can enable individuals with less expertise to create artwork of professional quality. This accessibility allows a larger audience to take part in the creative process. Fig. 1 shows the abstract art forms created by Gen AI.

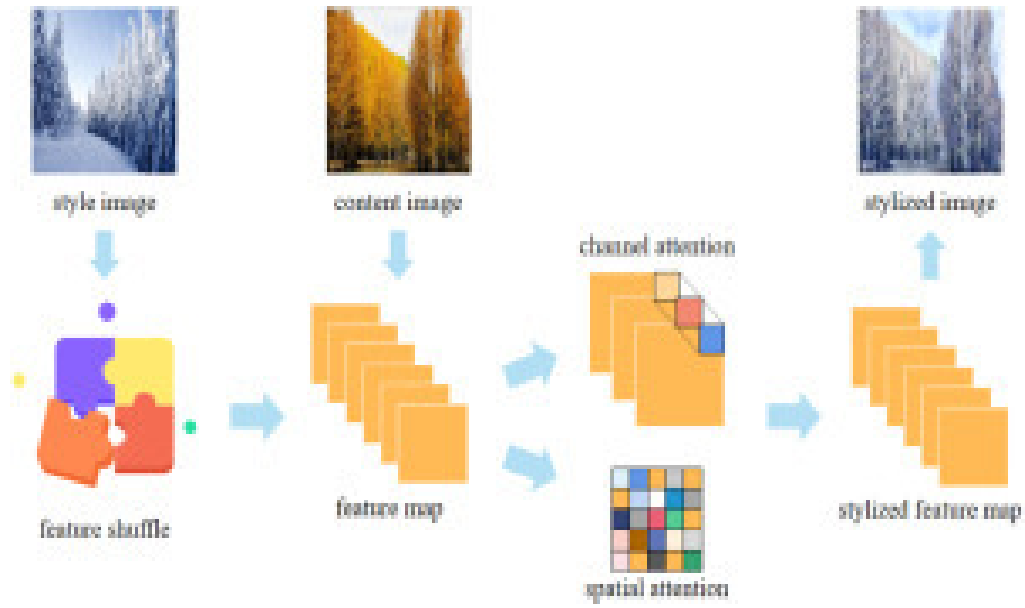

Fig 1: AI-generated abstract landscapes

Through a pre-designed transfer network, image style transfers embed the style of one image into another. The output image preserves the content of the latter while displaying the style of the former. A highly skilled painter is usually required to spend a considerable amount of time recreating the content image for this task. Nevertheless, due to deep learning's swift advancement, the task is now considerably easier to realize. Two primary style transfer technologies exist: GAN-based (generative adversarial network) methods and CNN-based (convolutional neural network) methods.In addition, it is important not to underestimate the role that Gen AI can play in restoring damaged art forms (Gür et al., 2024). The Gen AI models could predict and recreate the damaged and missing components by studying existing patterns and styles. This capacity is highly esteemed in relation to historical monuments, as it can preserve cultural objects and history.

GAN [1], introduced by Goodfellow and colleagues in 2014, is a generative model; it consists of a generator and a discriminator. Unlike the typical application of a random noise vector as input for the generator in the original GAN, in style transfer scenes, the input is a content image, and the output is a stylized image that incorporates both the original content and the target style. The discriminator's function is to ascertain if the produced image embodies the target style. The style transfer method based on GANs does not need an artificial definition of style, as it can learn automatically from the dataset of style images. This technique necessitates a big dataset and can learn only a limited number of specific styles.

The CNN-based style transfer technique is founded on its robust feature extraction abilities. Gatys was the first to implement a CNN for style transfer and introduced an efficient style transfer algorithm. This algorithm employs a deep CNN architecture to extract high-level abstract features from the input image and subsequently executes a style transfer by aligning the distribution of the feature space. This method bears some resemblance to parameter-based transfer learning for cross-domain transfer [8], [9]. It initially posits that the source and target tasks share certain parameters within the model, and then encodes the transferred knowledge into these shared parameters or the prior distribution, thereby facilitating knowledge transfer across tasks. The approach of utilizing CNN-extracted image features for style transfer also presupposes that the features employed for image classification are identical to those necessary for image style transfer, and subsequently transfers the distribution of the target style features through feature fusion. Typically, to prevent negative transfer, the CNN-based style transfer method seeks to acquire highly generalized parameters through extensive training. In contrast to GANs, most CNN-based style transfer techniques invariably necessitate a definition of the criteria for the styles. Style loss can generally be articulated in specific forms, such as the Gram matrix , maximum mean difference , relaxed optimal transport distance [12], and histogram difference . Style transfer images are produced by minimizing both the content loss and the style loss. The content-style fusion module frequently incorporates a predefined style embedding method or structure, such as AdaIN (Adaptive Instance Normalization) and WCT (Whitening and Colouring Transformation) [15], and executes intricate embedding operations to direct the model's fusion process. The fixed style definition and the predefined content-style fusion method allow the model to be utilized for multi-style transfers; however, this method still encounters certain challenges.

3. Utilizing Generative AI for Digital Marketing.

From a marketing perspective, consumer data can be analyzed by generative AI models in order to produce customized marketing content. This includes the development of tailored and individualized emails, social media content, and ads (Bartelt et al., 2023). Due to these advancements, the prerequisites for companies and individuals have been changed, as it has become easier to create marketing campaigns that are more effective and engaging. The technology allows for the development of multiple adaptable versions of the same category of advertisements. Moreover, it has become a lot easier to optimize performance and conduct real-time tests of digital ads and social media content. This greatly diminishes the time and expenses linked to manual approaches, all while improving the effectiveness of the company’s marketing tactics.According to Olivier et al. (2024), virtual assistants and AI-driven chatbots provide more personalized customer support by addressing inquiries and assisting customers throughout their purchasing journeys. These AI systems can simulate human-like interactions, which enhances customer satisfaction and engagement.

3.1. Generative AI-based Content Creation.

A multitude of Generative AI tools produce high-quality written content while also aiding in proofreading and editing. According to Barrett et al. (2023), the output of these textual Generative AI tools has led to a substantial reduction in the manual work needed by writers and editors. They also aid in brainstorming and collaborative writing for new ideas. When humans and Generative AI work together, the result can be content that is more diverse and attractive, thus enriching conventional storytelling.

3.2. Model Development Process in Generative AI .

Creating Generative AI models tailored to particular tasks involves several phases: preparing data, preprocessing it, choosing the right model, training the model, assessing its performance, and deploying it in real time. These actions are usually performed in an iterative manner until satisfactory results are obtained.

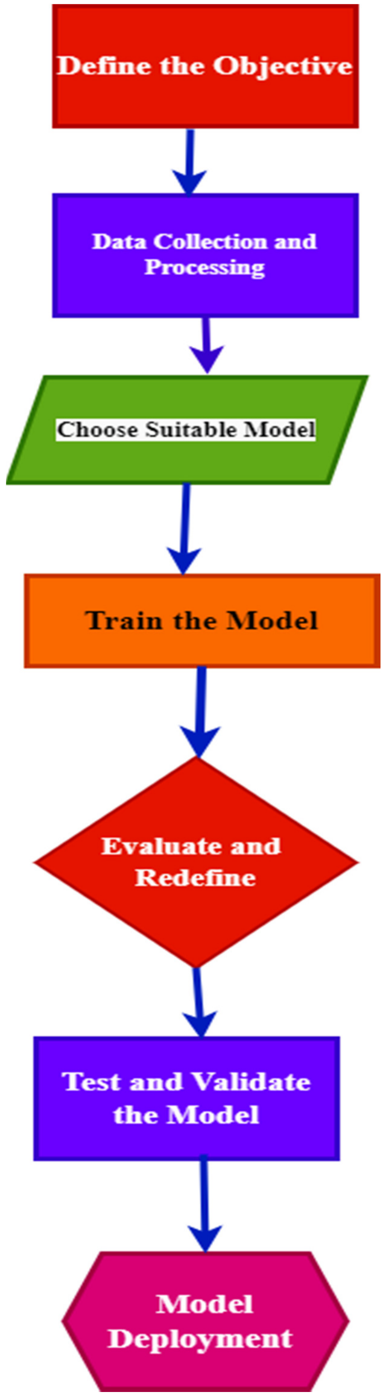

Fig 2: Model development process of Gen AI

3.3. Image Styling Generative AI

has become extensively used in the field of image styling, which includes a variety of tasks such as image transformation, enhancement, and the generation of new images through sophisticated AI algorithms that evaluate a wide range of artistic styles and image datasets. Image style transfer is defined as the method of applying the visual attributes of one image to another. Notable neural network architectures, such as convolutional neural networks (CNN), VGG19, and Generative Adversarial Networks (GANs), are predominantly utilized in this procedure. In the process of styling an image, designers are required to create a model that achieves super-resolution (Li et al., 2019). It is essential to transform low-quality images into high-resolution versions without any data loss. Moreover, addressing damaged or absent sections is also a crucial aspect of image styling. This is generally achieved through the inpainting technique, which employs the information from surrounding pixels to fill in the voids. AI tools like DALL-E and GANs are particularly proficient in effectively restoring missing details. Throughout the image style transfer process, significant effort is necessary to improve various quality dimensions of the image, such as color balance, contrast, and ambient lighting. This study intends to propose a framework for image style transfer using Generative AI, ensuring a high level of accuracy. The framework utilizes cyclic GANs to extract features from existing images and integrate them with the desired artistic components. Additionally, this research explores the potential limitations and advantages of the proposed framework within the context of image styling. Future research avenues and possible improvements to the proposed framework are also addressed in this study.

4. Literature Survey

Neural network-based image style transfer has been thoroughly examined and explored by a multitude of researchers. Numerous advancements employing advanced and cutting-edge computing techniques have been introduced, concentrating on the extraction of vital information from the input image, comprehending intricate styling patterns, and generating a new image with minimal loss of information or content. A cyclic GAN-based transfer learning framework, which circumvents the extraction of redundant details from images by utilizing a single discriminator and generator pair instead of two, has been proposed by Chan Hu (Chan Hu et al., 2020). The learning process within this framework occurs by analyzing the semantics between the input and generated images through feature maps produced by the VGG model. The model achieves convergence through pre-training and transfer learning. Furthermore, a simple CNN model can also be employed for image style transfer. A novel neural algorithm that acquires artistic styles by integrating and differentiating image styles and features to create an image with a new style has been introduced by Leon A. Gatys (Leon A. Gatys et al., 2016). This algorithm is notably effective as it can generate high-quality images by blending the attributes of any photographs with famous artworks. The capabilities of CNNs in the image style transfer process using the neural style transfer algorithm have significantly increased in various studies. This has resulted in numerous potential applications from both research and industrial perspectives. A thorough review of style transfer algorithms reliant on CNNs, along with a detailed classification taxonomy, has been presented by Jing et al. (Jing et al., 2019). This study evaluates different assessment methods both quantitatively and qualitatively. However, image styling is frequently not robust and tends to diminish the quality of the images. he field of computer vision has experienced an increasing interest in image style transfers as a topic of research. The existing CNN-based style transfer methods have laid the groundwork for a considerable amount of research in the domain of content–style fusion. Deterministic calculations are employed to implement style fusions that are artificially controlled. Due to these methods, the model's ability for automatic learning is limited, resulting in inconsistent style transfer effects. To tackle this problem, we propose a non-definitive style auto-transfer module. This module utilizes an attention submodule that guides the model regarding the channels and areas for content–style fusions. [1]. It has been demonstrated that contrastive learning is effective for cross-domain feature learning and is frequently utilized in image translation tasks. Nevertheless, these methods often overlook the differences between positive and negative samples concerning the model's optimization capability, treating them as equivalent. This oversight reduces the generative models' ability to represent features. This paper introduces a novel image translation model based on asymmetric slack contrastive learning. A new contrastive loss is asymmetrically designed by integrating a slack adjustment factor. Theoretical analysis suggests that it can adaptively optimize and adjust according to various positive and negative samples, resulting in a significant improvement in optimization efficiency. [2].Methods of image translation utilizing contrastive learning have been proposed, which assess various spatial locations to enhance spatial correspondence. However, these methods often neglect the different semantic relationships that exist within the images. To address this concern, we introduce a novel semantic relation consistency (SRC) regularization alongside decoupled contrastive learning. This strategy capitalizes on diverse semantics by focusing on the heterogeneous semantics among the image patches of a single image. We also present hard negative mining that utilizes the semantic relationships to further improve performance. Our approach was validated across three tasks: single-modal and multi-modal image translations, as well as the GAN compression task for image translation. The experimental findings confirmed that our method achieved state-of-the-art performance in all three tasks.[3]. The phrase "ambiguous sketchy image" denotes images that consist of abstract lines. In these images, unclear and weak semantics frequently arise. Stylizing such ambiguous sketchy images presents a challenge but is important for certain individuals, including children or those with difficulties in drawing. Nevertheless, current state-of-the-art (SOTA) techniques predominantly depend on semantic information, and when an ambiguous sketchy image is used as content, these methods face the issue of collapse.[4]. We propose a non-definitive style auto-transfer module to tackle this issue. This module utilizes an attention submodule that guides the model regarding the channels and the areas for content–style fusions. Instead of enforcing artificial definitions on the content–style fusion techniques, it permits the model to learn autonomously. Furthermore, we recommend a feature shuffle operation that reduces the influence of the style image on the content of the output. In addition, to improve the retention of both high-level and low-level information in the image, our loss function integrates a multi-scale content–style loss alongside an edge detection loss.[5]. This module utilizes an attention submodule that guides the model regarding the channels and the areas for content–style fusions. Instead of enforcing artificial definitions on the content–style fusion techniques, it permits the model to learn autonomously. Furthermore, we recommend a feature shuffle operation that reduces the influence of the style image on the content of the output. In addition, to improve the retention of both high-level and low-level information in the image, our loss function integrates a multi-scale content–style loss alongside an edge detection loss. The WikiArt and Microsoft COCO datasets are employed for all our experiments. The outcomes of the experiment indicate that our method can yield visual effects that are more stable and superior to those produced by existing techniques.[6]Recently, there has been an increasing interest in low-rank multi-view subspace clustering within the domain of multi-view learning research. Despite significant progress, most existing methods still encounter two primary challenges. Firstly, while they focus on leveraging the low-rank consistency across different views, they often neglect the low-rank structure present within each individual view. Secondly, they commonly experience substantial time overhead, typically due to the costs associated with matrix inversion and singular value decomposition (SVD) in each iteration. In light of these issues, we propose a method known as Facilitated Low-rank Multi-view Subspace Clustering (FLMSC), which is both efficient and effective.[7]. Due to its efficacy in utilizing multiple data views, multi-view clustering has attracted considerable attention. However, most current multi-view clustering approaches are primarily aimed at either examining the consistency or enhancing the diversity of various views. This paper presents a novel multi-view subspace clustering technique (CSMSC) that integrates both consistency and specificity for the purpose of learning subspace representations. We articulate the multi-view self-representation property by employing a shared consistent representation in conjunction with a collection of specific representations, which is more appropriate for real-world datasets.[8] One of the most prevalent techniques in multi-view clustering is consensus representation learning. However, most existing methods fail to address the importance of cluster separation and compactness within clusters, which hinders their ability to learn discriminative representations that are conducive to clustering. We propose an innovative deep multi-view clustering network that incorporates a dual contrastive mechanism to generate representations that facilitate clustering as a solution to this issue. Our method employs two contrasting losses: a dynamic cluster diffusion loss that aims to enhance the distance between distinct clusters, and a reliable neighbor-guided positive alignment loss that is intended to foster compactness within each cluster. [9]. The time-domain module is responsible for capturing the time-domain features of multivariate time series. By utilizing CNNs and GCNs, it enables the capture of both temporal and spatial dependencies in the time-domain. The frequency-domain module is designed to extract the frequency-domain characteristics of multivariate time series data. This module treats the frequency-domain features as images, thus transforming the multivariate time series classification challenge into an image classification task in a novel manner. [10]

5. Framework for Image Style Transfer Using Vision Transformer Encodings.

The proposed framework employs a vision transformer to capture stylistic features. Following this, cyclic GANs utilize these embeddings to perform image transfer on the original image. This section delineates the comprehensive workflow. The vision transformer is notably effective in demonstrating significant alterations in image processing (Ranftl et al., 2021). Unlike conventional CNN architectures, they are extensively used in computer vision applications. The input image is segmented into fixed-length patches of 16 x 16 pixels, a method known as image patching. This technique allows the model to treat images as discrete tokens. After the patching process, the next layer executes both flattening and embedding. Each patch is transformed into a vector, which is subsequently linearly mapped into a lower-dimensional space via a patch embedding layer. This conversion changes the patched spatial image data into a format that is amenable to further processing by the transformer. Figure 3 depicts the generation of embeddings. Transformers do not possess the capability to understand and interpret the input tokens. As a result, positional embeddings are incorporated to form patched embeddings. These embeddings aid the model in preserving spatial awareness concerning the position of each patch within the framework of the original image. Another vital element of the visual transformer is the transformer encoder. A series of encoder blocks are stacked to create a complete model. The proposed encoder transformer module comprises the following components: • Multi-Head Self-Attention (MHSA): This feature allows the proposed model to assess the significance of different patches in relation to each other. As a result, this process enables the model to maintain both global and local dependencies. The outputs from multiple heads in this layer are aggregated for collective processing.

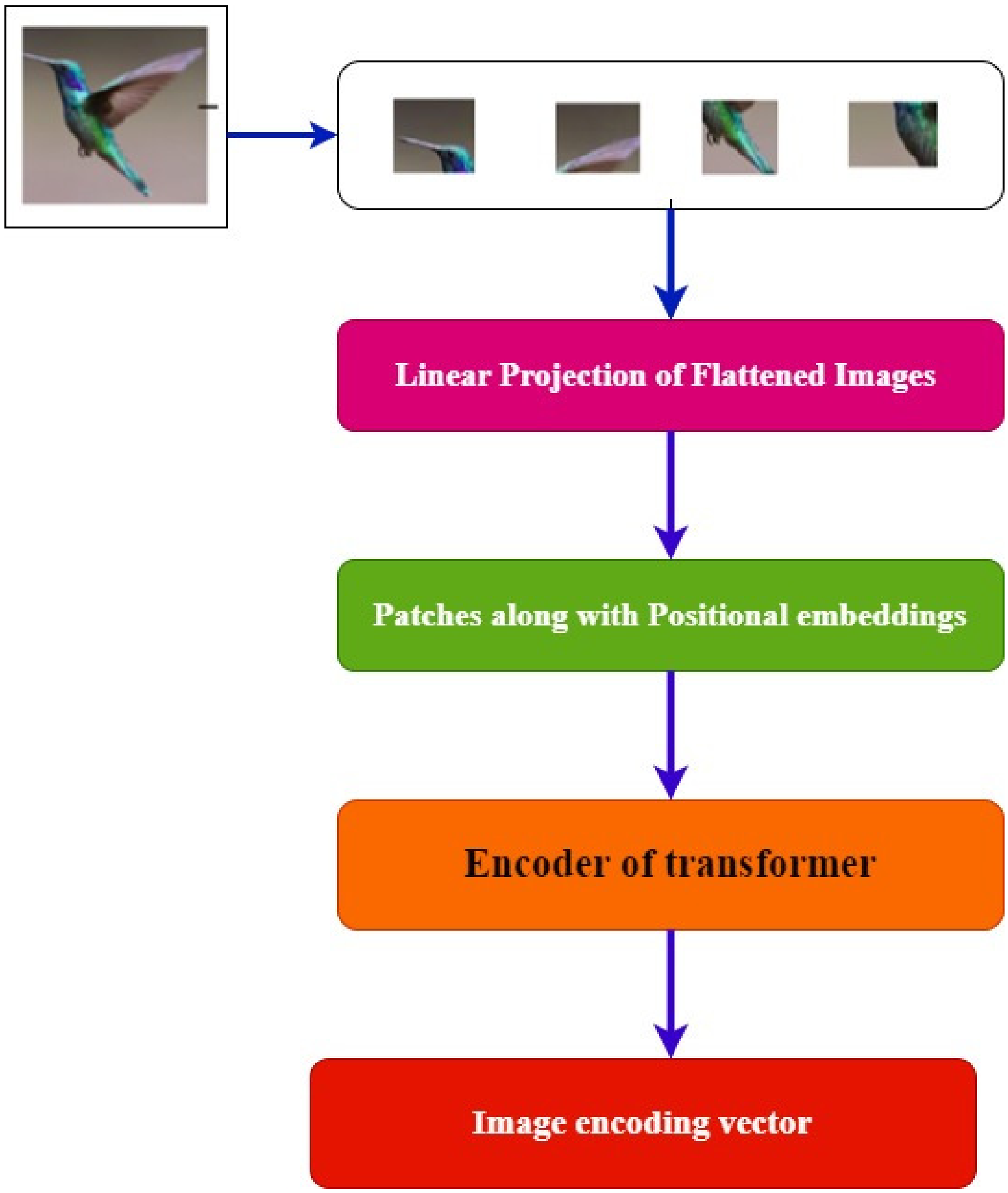

Fig 3: Generation of image embeddings

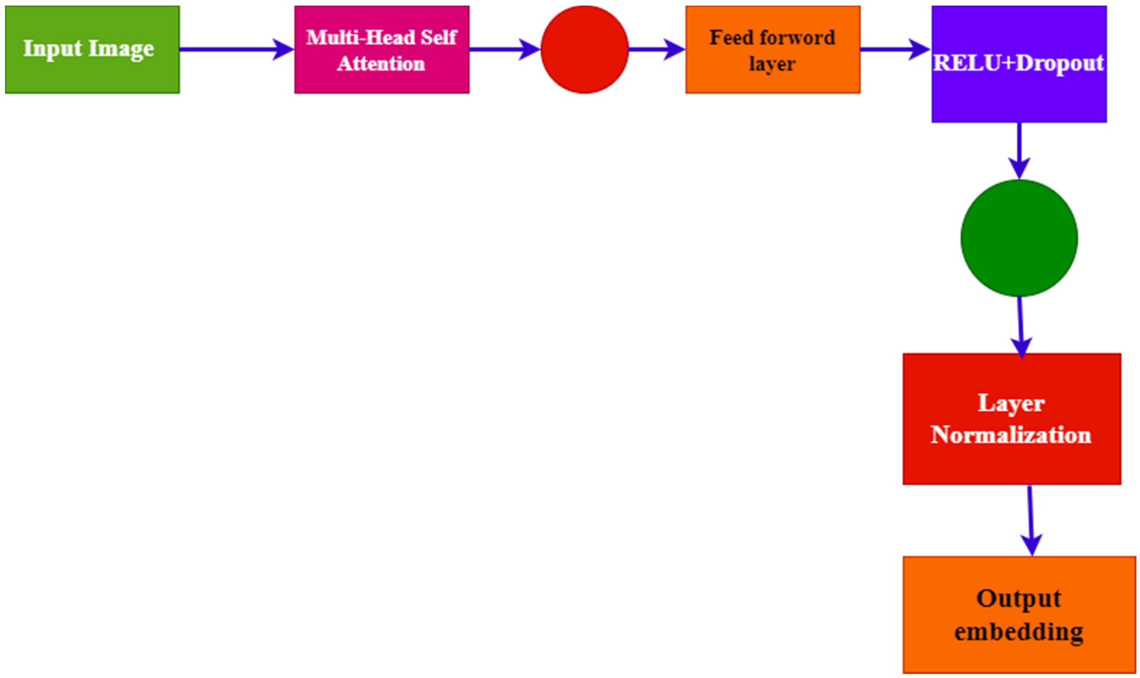

Fig 4: Architecture of Encoder transformer.

The Cyclic GANs utilize the output embeddings to convey styles. These generative models, referred to as GANs, are extensively employed for transforming images from the X domain to the Y domain (XY). This transformation is accomplished through a paired dataset where instance X of the image is closely associated with instance Y. However, cyclic GANs are particularly advantageous for style transfer applications as they do not necessitate these (X, Y) image mappings. To prevent the discriminator from distinguishing between the original image and the style-transferred image (y=G (x)), the generator (G) in the proposed architecture modifies the distribution derived from the embeddings of style X to that of style Y. The main challenge with this method is that the generator tends to modify each image uniformly, which hinders stylistic diversity. This contradicts the objective of establishing a robust framework for style transmission. To address this issue, the cyclic GAN employs a different generator (F) to enhance a constraint. This generator is responsible for reversing the transformation of G, ensuring that FGx=x is achieved with the assistance of this network. This guarantees that no single reduced example exists in any transformation of x. During the training process, a novel type of loss function is introduced to monitor the cycle consistency loss. Which modification should validate the two essential characteristics outlined below • FGx=x • G (Fy)=y Furthermore, a discriminator function will be incorporated by the discriminator network for G. The two elements of the soil's cyclic consistency are as follows: 1) The loss between x and their reconstruction 2) The difference between y and the corresponding reconstruction. The discriminators and generators of the proposed architecture will strive to optimize the same loss function, which consists of the two adversarial loss functions as illustrated in Fig. 5, similar to any generative adversarial network. Equation 1 outlines the overall loss function of the proposed architecture for image style transfer. In this equation, L represents the total adversarial loss function, while LGAN and Lcyc signify the loss functions generated by the cyclic and GAN, respectively.

$$L(G, F, D_{x'}, D_y) = L_{GAN}(G, D_{y'}, X, Y) + L_{GAN}(F, D_{x'}, Y, X) + L_{cyc}(G, F) \quad \textbf{(1)}$$

6.Results and Discussions

The proposed framework undergoes testing and training with a range of styling configurations and randomly selected input images. Several styling parameters are omitted, allowing the pre-trained Image Net dataset to be utilized for training purposes. The Imagenet dataset consists of images depicting various animal species set against diverse backgrounds. Subsequently, more generic objects were incorporated into the dataset. The results of applying style transfer through the proposed framework are illustrated in Fig. 7. The styling image is presented in Fig. 7a), while the image designated for style transfer is depicted in Fig. 7b). The latter is identified as being in domain y, whereas the former is categorized as being in domain x. The original image, represented in Fig. 7c, is merged with the styling selections to produce the final style transferred output.

Fig .5 Examples of invoking styles in image generation

When generative AIs are understood as 'style engines', our relationship with them shifts towards creative writing, visual work, conversation, and more expansive interactions, rather than merely retrieving information or searching for answers to factual inquiries. At its core, this involves applying a style to pre-existing content by uploading a text or image to a generative AI system for manipulation. This encompasses, for instance, shortening, summarizing, or cleaning texts (using conciseness as a stylistic approach), converting text into bullet points, Genuinely generative uses involve the invocation of one or more styles in the production of new content. In the context of generating images, prompting can be seen as a useful means of combining multiple styles. As an illustration, merging the aesthetics of Scandinavian design, armchair-ness, and minimalism produces a representation akin to that found in IKEA catalog imagery (Fig. 3a). Once more, considering that the fundamental qualities of objects, their materials, and any other patterns are recorded as patterns in the latent space, AI models facilitate the visual exploration of objects that might otherwise be challenging or even impossible to examine.

a) Style image **b) Original Image** **c) Style transferred image**

Fig 6: Style transferred image with reference style and input image

The suggested framework can be employed to produce a target image that boasts improved color quality and stylistic coherence, alongside enhancing the overall image quality. This method maintains the creative essence of the original image while ensuring that its details remain unaltered. The outcomes of the proposed method indicate that the model retains the structure of the actual image while adding a distinctive artistic flair. While the aforementioned output appears to demonstrate a very smooth style transfer of the picture, it is crucial to evaluate the production quality through quantitative assessments as well. To determine the effectiveness of this work, it is vital to compare the performance of the proposed architectural framework against other advanced techniques using quantitative metrics.

6.1. Feature similarity measure (FSIM)

Three important factors—brightness, contrast, and structure—are used by the structural similarity to evaluate the quality of the image. FSIM extends the concept of structural similarity, with the key difference being that it does not assign equal weight to all pixels (Sara et al., 2019).

$$FSIM = \frac{\sum_{x\in\rho} S_L(x) PC_m(x)}{\sum\sum_{x\epsilon\rho} PC_m\,(x)} \tag{2}$$

The expression for evaluating the FSIM between the original and output images is presented in Equation 2. PCm denotes phase consistency, whereas SL is the function that depicts the gradient in the process of image style transfer.

6.2. MSSIM (Mean SSIM index)

The MMSIM is another crucial measure for evaluating performance; it reflects the average Structural Similarity Index between the original and generated images. Equation 3 (Want et al., 2004) provides the expression for estimating the MSSIM.

$$MSSIM\,(X) = \frac{\sum_{i=1}^{N} SSIM(x_i, y_i)}{N} \tag{3}$$

X and Y denote the input image and output image, respectively. The terms xi and yi denote the image contents with i representing the window size. The number of local windows is indicated by N. The proposed framework's performance is evaluated against other state-of-the-art techniques, such as StyleGAN and CycleGAN (without embeddings), using the metrics mentioned above. Figure 8 presents the results of the comparison. It can be noted that the proposed framework demonstrates enhanced outcomes regarding both evaluation metrics. This indicates that the visual embeddings produced by the visual transformer have successfully captured the image features in both the style image and the original image.

Fig 7: Comparative analysis of the proposed framework with cyclic GANs and Style GANs

The two discriminator and generator networks are identical, sharing the same layers and dimensions. Generally, the generator neural network is composed of residual blocks along with multiple convolutional layers. To incorporate nonlinearity into the learning process, the RELU activation function is utilized. For each training iteration of the proposed architecture, two groups of images, x and y, are selected. The generator network produces generic images of the target domain. Furthermore, adversarial loss is computed, a process we refer to as forward consistency. The same procedure was repeated with the y image embeddings, which we designate as backward consistency.

Consequently, the visual transformer is employed in the proposed design to extract information from the images. The entire style transfer process is then executed by the proposed cyclic GAN architecture, utilizing these embeddings through two generator neural networks equipped with skip connections.

4. Conclusions AND FUTURE WORKS

The suggested framework can be employed to produce a target image that boasts improved color quality and stylistic coherence, alongside enhancing the overall image quality. This method maintains the creative essence of the original image while ensuring that its details remain unaltered. The outcomes of the proposed method indicate that the model retains the structure of the actual image while adding a distinctive artistic flair. While the aforementioned output appears to demonstrate a very smooth style transfer of the picture, it is crucial to evaluate the production quality through quantitative assessments as well. To determine the effectiveness of this work, it is vital to compare the performance of the proposed architectural framework against other advanced techniques using quantitative metrics. Comparative analysis of the proposed framework with cyclic GANs and Style GANs

References

1. Wang, J., Xu, L., Yu, X., Lu, H., Huang, Z., Wang, H., & Pang, C. (2023). A non-definitive auto-transfer mechanism for arbitrary style transfers. *Knowledge-Based Systems*, *260*, 110171.https://doi.org/10.1016/j.knosys.2022.110171.
2. Zhang, Y., Li, M., Gou, Y., & He, Y. (2024). Asymmetric slack contrastive learning for full use of feature information in image translation. *Knowledge-Based Systems*, *299*, 112136.https://doi.org/10.1016/j.knosys.2024.112136.
3. Jung, C., Kwon, G., & Ye, J. C. (2022). Exploring patch-wise semantic relation for contrastive learning in image-to-image translation tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18260-18269).
4. Chen, Y., Zheng, Y., Li, P., Luo, W., Kang, S., & He, W. (2025). Content-activating for artistic style transfer with ambiguous sketchy content image. *Neurocomputing*, *620*, 129217. https://doi.org/10.1016/j.neucom.2024.129217.
5. Wang, J., Xu, L., Yu, X., Lu, H., Huang, Z., Wang, H., & Pang, C. (2023). A non-definitive auto-transfer mechanism for arbitrary style transfers. *Knowledge-Based Systems*, *260*, 11017 https://doi.org/10.1016/j.knosys.2022.110171.
6. Wang, J., Xu, L., Yu, X., Lu, H., Huang, Z., Wang, H., & Pang, C. (2023). A non-definitive auto-transfer mechanism for arbitrary style transfers. *Knowledge-Based Systems*, *260*, 110171. https://doi.org/10.1016/j.knosys.2022.110171.
7. Zhang, G. Y., Huang, D., & Wang, C. D. (2023). Facilitated low-rank multi-view subspace clustering. *Knowledge-Based Systems*, *260*, 110141. https://doi.org/10.1016/j.knosys.2022.110141.
8. Luo, S., Zhang, C., Zhang, W., & Cao, X. (2018, April). Consistent and specific multi-view subspace clustering. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1). **DOI:** https://doi.org/10.1609/aaai.v32i1.11617.

9. J. Cui, Y. Li, H. Huang and J. Wen, "Dual Contrast-Driven Deep Multi-View Clustering," in *IEEE Transactions on Image Processing*, vol. 33, pp. 4753-4764, 2024, doi: 10.1109/TIP.2024.3444269. **DOI:** 10.1109/TIP.2024.3444269.

10. Lei, T., Li, J., & Yang, K. (2024). Time and frequency-domain feature fusion network for multivariate time series classification. *Expert Systems with Applications*, *252*, 124155. https://doi.org/10.1016/j.eswa.2024.124155.

****