Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks

Jonayet Miah[1] ✉ Duc M Cao[2], Md Abu Sayed[3], Md Sabbirul Haque and Rhine Hoque[5]
[1]*Department of Computer Science, University of South Dakota, South Dakota, USA*
[2]*Department of Economics, University of Tennessee, Knoxville, TN, USA*
[3]*Department of Professional Security Studies, New Jersey City University, Jersey City, New Jersey, USA*
[4]*IEEE Professional Community, IEEE*
[5]*Researcher, Dhaka, Bangladesh*

Corresponding Author: Jonayet Miah, **E-mail**: Jonayet.miah@coyotes.usd.edu

| ABSTRACT
Artistic style transfer, a captivating application of generative artificial intelligence, involves fusing the content of one image with the artistic style of another to create unique visual compositions. This paper presents a comprehensive overview of a novel technique for style transfer using Convolutional Neural Networks (CNNs). By leveraging deep image representations learned by CNNs, we demonstrate how to separate and manipulate image content and style, enabling the synthesis of high-quality images that combine content and style in a harmonious manner. We describe the methodology, including content and style representations, loss computation, and optimization, and showcase experimental results highlighting the effectiveness and versatility of the approach across different styles and content.

| KEYWORDS

Artistic style transfer, Generative AI, Convolutional Neural Networks (CNNs), Image processing, Texture transfer, Image synthesis, Content representation, Style representation, Loss computation

| ARTICLE INFORMATION

ACCEPTED: 02 November 2023 **PUBLISHED:** 17 November 2023 **DOI:** 10.32996/jcsts.2023.5.4.9

1. Introduction

In the realm of image processing and artistic expression, the quest to combine the style of one image with the content of another has led to a captivating endeavor known as style transfer. This creative technique involves imbuing an image with the visual aesthetic of another, resulting in captivating combinations that merge artistic styles with recognizable subjects. However, achieving this fusion is no trivial task; it necessitates a delicate balance between preserving the underlying content of the target image and infusing it with the stylistic essence of the source image.

This pursuit of harmonizing style and content is inherently linked to the broader challenge of texture transfer, wherein the objective is to synthesize a texture from a source image while retaining the semantic content of a designated target image. Various algorithms have tackled this challenge by exploiting non-parametric methods that resample pixel data from a source texture, yielding photorealistic textures that align with the texture of a source image. Previous methods have leveraged correspondence maps, image analogies, and edge orientation information to guide the texture transfer process.

Yet, these methods share a common constraint—they primarily utilize low-level image features from the target image to steer the texture transfer process. Ideally, a comprehensive style transfer algorithm should decipher the semantic content of the target image, including objects and overall scenery, and then apply a texture transfer technique to render this semantic content in the style of the source image. Such an algorithm hinges upon the ability to disentangle semantic content from style in image representations.

Recent strides in the field, particularly through the advent of Deep Convolutional Neural Networks (CNNs), have ushered in new possibilities. These high-performing CNNs have proven adept at extracting intricate semantic information from images, as evidenced by their prowess in tasks like object recognition and texture classification. Leveraging the feature representations learned by these networks, an innovative algorithm known as "A Neural Algorithm of Artistic Style" emerges, offering a novel approach to image style transfer. In Figure 1, we would like to share the Neural Style Transfer Overview.

Figure 1: Entire Workflow

At its core, this algorithm revolves around the notion of constraining a texture synthesis method using feature representations extracted from state-of-the-art CNNs. By aligning the texture model with deep image representations, the process of style transfer transforms into an optimization challenge within a single neural network. The algorithm effectively generates new images by embarking on a pre-image search, matching feature representations of exemplar images. This approach, which draws from previous endeavors in texture synthesis and deep image representation analysis, combines a parametric texture model rooted in CNNs with an inversion technique for image representations. The foundation of this work rests upon the VGG network, a powerful CNN that excels in object recognition and localization tasks. By leveraging the normalized feature space of this network, the algorithm dissects images into layers of convolutional and pooling stages. These layers encode complex filters that respond to specific image attributes, ultimately capturing the essence of the image at varying degrees of complexity.

In essence, this algorithm navigates the intricate interplay between content and style in images, guided by the rich knowledge encoded within deep CNNs. The subsequent sections will delve into the technical intricacies of content and style representations, illustrating how these components converge to facilitate the magic of image style transfer. Through the synergy of deep image representations and advanced neural network techniques, the algorithm promises to unlock a new realm of artistic expression, merging content and style in unprecedented ways.

2. Literature Review

Gatys et al. (2016) employ image representations acquired from Convolutional Neural Networks that have been explicitly optimized for the purpose of recognizing objects. These networks excel in capturing higher-level details present in images. We introduce an innovative technique named "An Artistic Style Neural Algorithm," which allows for the separation and subsequent recombination of the content and artistic style found in natural images. This algorithm empowers us to create novel images with a robust perceptual quality by merging the content from any given photograph with the visual characteristics present in various well-known artworks. Our results provide novel insights into the inherent attributes of deep image representations that are acquired through Convolutional Neural Networks, displaying their potential for advanced image synthesis and manipulation at an elevated level.

Wang et al. (2022) paper introduces the texture reformulator, a rapid and versatile framework driven by neural networks designed for interactive texture transfer while accommodating user-defined instructions. The primary challenges revolve around three key aspects: 1) handling diverse tasks, 2) utilizing simple guidance maps, and 3) ensuring efficient execution. To address these obstacles, our central concept involves implementing an innovative feed-forward approach that includes multiple perspectives and stages, comprising I) a phase for globally aligning structural elements, II) a phase for locally refining texture details, and III) a phase for comprehensively enhancing the overall effect. This approach facilitates the generation of high-quality results characterized by consistent structures and intricate texture specifics, employing a strategy that progresses from coarse to fine. Furthermore, we introduce a new learning-free operation termed view-specific texture reformulation (VSTR), which integrates a fresh strategy involving semantic map guidance. This strategy enables more precise texture transfer that adheres to the intended semantics and preserves the underlying structure. The effectiveness and superiority of our framework are substantiated through extensive experimentation across a range of application scenarios.

Efros et al. (2023) introduce a straightforward image-based approach to creating a fresh visual appearance, where a new image is generated by combining small patches from existing images. They term this procedure "image quilting." Initially, they employ quilting as a rapid and exceedingly uncomplicated texture synthesis technique, which delivers remarkably satisfactory outcomes for a diverse array of textures. Subsequently, they expand the algorithm's scope to facilitate texture transfer and the rendering of an object using a texture derived from a different object. In a broader context, they demonstrate the capacity to reimagine an image with the artistic style of a distinct image. This method directly manipulates images and does not require the utilization of three-dimensional data.

Güçlü et al. (2015) substantiated the presence of a distinct gradient for feature complexity in the ventral pathway of the human brain through quantitative analysis. This accomplishment involved the meticulous mapping of a multitude of stimulus features with varying degrees of complexity across the cortical sheet using a deep neural network. This methodology not only unveiled a finely detailed functional specialization within the downstream regions of the ventral stream but also facilitated exceptional precision in decoding human brain activity representations. The successful explanation of neural responses by certain stimulus features indicated a clear and explicit tuning of population receptive fields specifically tailored for the purpose of object categorization.

Kingma et al. (2014) reconsider the strategy for semi-supervised learning using generative models and create novel models that enable efficient extrapolation from limited labeled datasets to extensive unlabeled datasets. Up to now, generative methods have been either rigid, ineffective, or lacking scalability. They demonstrate that by utilizing deep generative models and approximate Bayesian inference that harnesses recent progress in variational techniques, substantial enhancements can be achieved. This elevates the competitiveness of generative techniques in the realm of semi-supervised learning.

Kyprianidis et al. (2012) present a comprehensive examination of the non-photorealistic rendering (NPR) field, concentrating on methods that transform 2D inputs like images and videos into artistically stylized representations. To commence, we introduce a categorization of the 2D NPR techniques that have emerged over the past two decades. These are structured based on the unique design characteristics and behaviors inherent in each approach. Following that, we elaborate on the evolution of techniques, progressing from semi-automatic painting systems that were prevalent in the early 1990s to the automated painterly rendering systems that emerged in the late 1990s, driven by the analysis of image gradients. We subsequently explore two interconnected trends identified within the NPR literature, aligning with our established categorization. Primarily, we delve into the fusion of advanced computer vision with NPR, emphasizing the tendency to incorporate scene analysis to guide artistic abstraction and broaden the scope of stylistic variability. Secondly, the shift in local processing methodologies towards edge-aware filtering is explored, particularly concerning the real-time stylization of images and videos.

3. Method and Materials

Figure 2 illustrates the intricate image representations within a Convolutional Neural Network (CNN). When an input image undergoes processing stages in the CNN, it becomes a collection of filtered images. As the processing hierarchy advances, the number of filters increases, yet the size of filtered images diminishes due to down-sampling mechanisms like max pooling. This leads to a reduction in the total units per layer. By reconstructing the input image from various layers ('conv1 2' to 'conv5 2' of the VGG-Network), it's evident that lower layers achieve nearly flawless reconstruction (a–c), while higher layers retain high-level content while losing detailed pixel information (d,e). The style reconstructions, built upon CNN activations and capturing texture information, involve computing correlations between features in different layers. Style reconstruction occurs by creating style representations using subsets of CNN layers ('conv1 1' to 'conv5 1'), yielding images that progressively match the style of a given image while omitting global scene arrangement details.

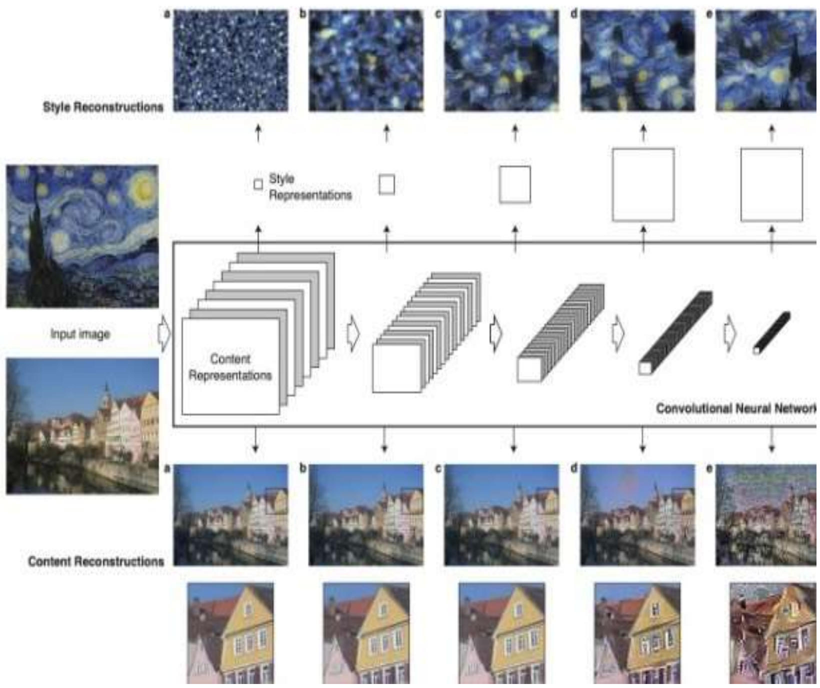

Figure 2: Figure Construction

In the past, factorized representations like these were successfully obtained for controlled portions of real-world images, like faces captured under different lighting conditions or characters written in various font styles. Nevertheless, the separation of content and style within natural images still presents a considerable obstacle. A recent significant advancement comes in the form of Deep Convolutional Neural Networks. These networks have made it possible to extract meaningful semantic information from real images [13,14,15]. These networks are trained for specific objectives such as recognizing objects, and they possess the ability to generalize high-level content across various datasets and even different visual tasks, encompassing aspects like recognizing textures and categorizing artistic styles.

Expanding upon this progress, we introduce "An Artistic Style Neural Algorithm," a novel strategy for transferring artistic style within images [16,17]. This approach leverages the capabilities of Convolutional Neural Networks to autonomously manipulate both content and style aspects in real images. Operating as a technique for transferring textures, it makes use of feature representations. from state-of-the-art CNNs to constrain the texture synthesis process. Since both the texture model and deep image representations contribute, the style transfer process simplifies into a single-network optimization challenge. By matching feature representations of example images through pre-image searches, new images emerge. This methodology, employed in previous texture synthesis work to enhance the comprehension of deep image representations, synergizes a parametric texture model from CNNs with an image representation inversion technique.

3.1 Deep image representations

The outcomes presented in the following sections were produced using the VGG network, which underwent training for tasks like object recognition and localization. This network's comprehensive details can be found in the original publication. In our work, we utilized the feature space that originates from a standardized version of the 19-layer VGG network, which consists of 16 convolutional layers and 5 pooling layers [19,20,22]. The normalization of the network involved adjusting the weights to ensure that the mean activation of each convolutional filter, spanning various images and positions, equated to one. This normalization procedure, which is tailored to the VGG network [24,25,26], does not influence its output since it exclusively employs rectifying linear activation functions and lacks normalization or pooling across feature maps. It's noteworthy that the fully connected layers were not utilized in our methodology. The architecture itself is openly accessible and can be explored within the Caffe framework. In the realm of image synthesis, we observed that substituting the traditional maximum pooling operation with average pooling yielded slightly more visually pleasing outcomes. This preference for average pooling is the reason the showcased images were generated using this method.

3.2 Style representation and transfer

To obtain an interpretation of the artistic style present in an input image, we utilize a deliberately designed feature space to encompass texture intricacies [27,29]. This feature space can be formulated based on the filter responses present in any layer of the network. It comprises correlations computed from the various filter responses, with these correlations being calculated across the spatial dimensions of the feature maps. These correlations based on features are encapsulated within the Gram matrix Gl ∈ RNl×Nl, where Gl ij represents the inner product that exists.

To impart the style of an artistic creation ~a onto a photograph ~p, we generate a fresh image that concurrently aligns with the content depiction of ~p and the style depiction of ~a (as depicted in Fig 2). In essence, we simultaneously decrease the disparity between the feature representations of a white noise image and the content portrayal of the photograph in a particular layer while also reducing the discrepancy between the feature representations of the artwork's style, defined across several layers of the Convolutional Neural Network between the vectorized feature maps of i and j within layer l:

3.3 Loss Functions

To achieve the desired image, it is necessary to establish a loss function that guides the optimization process towards the intended outcome. In this context, we will employ the concept of per-pixel losses. The Per Pixel Loss stands as a metric used to evaluate disparities between images at the pixel level. It involves comparing the pixel values in the generated output with those of the input. (Another approach, perpetual loss functions, will be briefly discussed in subsequent sections of this article.) However, per-pixel loss occasionally falls short of capturing all the essential characteristics. This is where perpetual losses come into play, addressing such limitations. The loss components that will be emphasized are:

1. Content Loss
2. Style Loss

Content Loss The objective of content loss is to effectively encapsulate the desired content within the generated image. Empirical observations reveal that Convolutional Neural Networks (CNNs) tend to encode content-related information predominantly in the upper layers of the network, while the lower layers are primarily concerned with individual pixel-level attributes.

Crafting the loss function for style entails a more intricate process than content, mainly due to the engagement of multiple layers in computation. The essence of style information is gauged through the extent of correlation existing among feature maps within each layer. To quantify style loss, the Gram Matrix is employed [29,30,31]. The Gram matrix acts as a yardstick for capturing the distribution pattern of features across a collection of feature maps within a specific layer. Consequently, when working on calculating or minimizing style loss, the underlying endeavor is to align the distribution characteristics of features in both the original style and the generated images.

4. Result and Discussion

The primary discovery of this research paper lies in the discernible separability of content and style representations within the Convolutional Neural Network [32,34]. This signifies the ability to independently manipulate these representations to create novel, perceptually significant images. To illustrate this breakthrough, we generate images that amalgamate the content representation of a photograph portraying the Neckar riverfront in Tubingen, Germany, with the style representations extracted from various renowned artworks spanning diverse artistic eras (as depicted in Figure 3). The images showcased in Figure 3 are synthesized by aligning the content representation from the 'conv4 2' layer and the style representation from layers 'conv1 1', 'conv2 1', 'conv3 1', 'conv4 1', and 'conv5 1' (with a weight ratio of 1/5 in these layers, and 0 weight in all other layers). The α/β ratio is set to either 1 × 10^-2 (Figure 3B), 7× 10^-3 (Figure 3C) or 6× 10^-3 (Figure 3D).

Figure: Result Images

4.1 The trade-off between content and style matching

Achieving a complete disentanglement between image content and style is not always feasible. When attempting to create an image that merges the content of one image with the style of another, it's often challenging to find an image that precisely satisfies both requirements simultaneously. Nevertheless, due to the nature of the loss function employed in image synthesis – which is a linear combination of content and style loss functions – we can seamlessly control the degree of emphasis on either faithfully reconstructing the content or faithfully replicating the style.

4.2 Impact of Various Layers in the Convolutional Neural Network.

The selection of layers within the Convolutional Neural Network (CNN)[35,36,37] has a significant impact on the process of synthesizing images. As previously discussed, the style representation encompasses various layers of the neural network and spans multiple scales. The specific layers chosen and their arrangement influence the extent of style matching at different local scales, resulting in distinct visual outcomes (refer to Figure 1, style reconstructions). It is observed that by aligning the style representations with higher layers in the network, finer image details on larger scales are retained, contributing to a smoother and more cohesive visual perception. Consequently, the images that are most visually appealing are typically generated by harmonizing the style features within the upper layers of the network. This explains why, in all the showcased images, we match the style characteristics in layers 'conv1 1', 'conv2 1', 'conv3 1', 'conv4 1', and 'conv5 1' of the network.

5. Conclusion and Future Work

In conclusion, the paper "Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks" presents an innovative approach to artistic style transfer by leveraging the power of Convolutional Neural Networks (CNNs) to manipulate and combine image content and style. The technique offers a way to synthesize images that harmoniously blend the content of one image with the artistic style of another. The authors highlight the capability of deep CNNs to extract high-level semantic information from images, enabling the separation of content and style representations for manipulation.

The key contributions of the paper include the introduction of a loss function that incorporates both content and style information, guiding the optimization process to create images that align with the intended outcomes. The use of feature representations extracted from the VGG network's [40,41,42] layers allows for the disentanglement of content and style, providing a foundation for the style transfer process.

However, the paper also acknowledges some limitations of the proposed method. The computational complexity of the optimization process can be demanding, particularly for higher-resolution images. This restricts the real-time and interactive use of the algorithm. Additionally, some synthesized images might exhibit minor noise, which could affect photorealism, especially when both content and style images are photographs.

Arguably, the most limiting factor pertains to the resolution of the generated images. The complexity of the optimization problem and the number of units within the Convolutional Neural Network both increase linearly with the count of pixels. Consequently, the speed of the synthesis process is heavily reliant on the resolution of the image. The images presented in this paper were produced at an approximate resolution of 512 × 512 pixels, and the synthesis procedure could take up to an hour utilizing an Nvidia K40 GPU [43] (subject to specific image dimensions and gradient descent stopping criteria). While the current performance levels impede the real-time and interactive implementation of our style transfer algorithm, it is conceivable that forthcoming advancements in deep learning will enhance the method's efficiency. An additional concern is that the synthesized images may occasionally exhibit minor low-level noise. While this holds less significance in artistic style transfer, it becomes more noticeable when both the content and style images are photographs, which could potentially impact the photorealism of the generated image.

However, this noise appears to possess distinct characteristics resembling the network's unit filters. Consequently, it might be feasible to formulate effective denoising approaches for post-processing images after the optimization procedure. Future advancements in deep learning and optimization techniques are likely to enhance further the efficiency, quality, and flexibility of this technique.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers

References:

[1] Bellala, G., Sricharan, K., & Srinivasa, J. A. (n.d). Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients.

[2] Benaim, S., Wolf, L. (2017).One-sided unsupervised domain mapping. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December. 752–762.]

[3] Cabral, B., & Leedom, L. C. (1993, September). Imaging vector fields using line integral convolution. In *Proceedings of the 20th annual conference on Computer Graphics and interactive techniques* (pp. 263-270).

[4] Collomosse, J. P., & Hall, P. M. (2002, June). Painterly rendering using image salience. In *Proceedings 20th Euro graphics UK Conference* (pp. 122-128). IEEE.

[5] Chen, Y., Lai, Y.K., Liu, Y.J. (2018). Cartoongan: Generative adversarial networks for photo carbonization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. 9465–9474.

[6] Efros, A. A., & Freeman, W. T. (2023). Image quilting for texture synthesis and transfer. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* 571-576

[7] Efros, A.A., Freeman, W.T. (2001). Image quilting for texture synthesis and transfer. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 341–346.

[8] Efros, A.A., Leung, T.K. (n.d). Texture synthesis by non-parametric sampling. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Corfu, Greece, 20–27 September. 1033–1038.

[9] Gatys, L.A., Ecker, A.S., Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June. 2414–2423.

[10] Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005-10014.

[11] Haque, M. S., Amin, M. S., Miah, J., Cao, D. M., & Ahmed, A. H. (2023). Boosting Stock Price Prediction with Anticipated Macro Policy Changes. *Journal of Mathematics and Statistics Studies*, *4*(3), 29-34.

[12] He, B., Gao, F., Ma, D., Shi, B., Duan, L.Y. (2018). Chipgan: A generative adversarial network for chinese ink wash painting style transfer. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October. 1172–1180.

[13] Islam, N., Shah A A., Islam, M. and Faisal, F. (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*. Springer, Singapore

[14] Jonayet M, Razib H K, Sabbir A, and Isatyaq M. (2023). A comparative study of detecting covid 19 by using chest X-ray images – A deep learning approach," In 2023 IEEE World AIIoT Congress (AIIoT) Regular Research Paper, 2023.

[15] Khan, R.H., Miah, J. Hasan M.M and Mamun, M. (2023). A study of forecasting stocks price by using deep Reinforcement Learning, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, pp. 0250-0255, doi: 10.1109/AIIoT58121.2023.10174358.

[16] Khan, R.H., Miah, J., Abed N and Islam. (n.d). A Comparative Study of Machine Learning classifiers to Analyze the Precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.

[17] Kayyum S. (2020). Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020. 1-7, doi: 10.1109/ICCCI48352.2020.9104104.

[18] Kingma, D. P., Mohamed, S., Jimenez Rezende, D., & Welling, M. (2014). Semi-supervised learning with deep generative models. *Advances in neural information processing systems*, *27*.

[19] Kyprianidis, J. E., Collomosse, J., Wang, T., & Isenberg, T. (2012). State of the" art": A taxonomy of artistic stylization techniques for images and video. *IEEE transactions on visualization and computer graphics*, *19*(5), 866-885.

[20] Karras, T., Laine, S., Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June. 4401–4410.

[21] Krizhevsky, A., Sutskever, I.., Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* , *60*, 84–90.

[22] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature* 2015, *521*, 436–444.

[23] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H. (2017). Universal style transfer via feature transforms. *Adv. Neural Inf. Process. Syst.*, *30*, 386–396.

[24] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*

[25] Larson, E C., and Damon M. C (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*: 011006-011006.

[26] Li, Q and Zhou W. (2009). Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE journal of selected topics in signal processing* 3: 202-211.

[27] Miah, J., Mamun, M., Rahman, M. I. (2022). MHfit: Mobile Health Data for Predicting Athletics Fitness using Machine Learning Models 2022 2^{nd} International seminar on machine learning, Optimization, and Data Science (ISMODE), (Preprint)

[28] Mokhayeri, F., Granger, E. A. (2020). Paired sparse representation model for robust face recognition from a single sample. *Pattern Recognit.* 2020, *100*, 107129.

[29] Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., Jiang, J. (2020). Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion* 2020, *62*, 110–120.

[30] Ninassi, A., Le Meur, O., Le Callet, P., & Barba, D. (2008, October). Which semi-local visual masking model forwavelet based image quality metric. In *2008 15th IEEE International Conference on Image Processing* (pp. 1180-1183). IEEE.

[31] Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, *40*, 49-70.

[32] Sun, W., & Wang, R. (2018). Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 474-478.

[33] Sheikh, H. R., Muhammad F. S and Alan C. B. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*: 3440-3451.

[34] Tenenbaum, J. B., & Freeman, W. T. (2000). Separating style and content with bilinear models. *Neural computation*, *12*(6), 1247-1283.

[35] Upchurch, P., Gardner, J., Pleiss, G., Pless, R., Snavely, N., Bala, K. and Weinberger, K. (2017). Deep feature interpolation for image content changes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 7064–7073.

[36] Wang, Z, Alan C. B, Hamid R. S, and Eero P. S (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13: 600-612.

[37] Wang, Z, Alan C. B and Ligang L (2002). Why is image quality assessment so difficult. In *2002 IEEE International conference on acoustics, speech, and signal processing*. IV-3313. IEEE, 2002.

[38] Wang, Z and Alan C B. (2006). Modern image quality assessment." PhD diss., Morgan & Claypool Publishers.

[39] Wang, Z., Zhao, L., Chen, H., Li, A., Zuo, Z., Xing, W., & Lu, D. (2022, June). Texture reformer: towards fast and universal interactive texture transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence*

[40] Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, *111*(23), 8619-8624.

[41] Yu, Y., Gong, Z., Zhong, P., Shan, J. (2017). Unsupervised representation learning with deep convolutional neural network for remote sensing images. In Proceedings of the International Conference on Image and Graphics, Shanghai, China, 13–15 September. 97–108.

[42] Ye, S., Su, K., & Xiao, C. (2008, May). Video quality assessment based on edge structural similarity. In *2008 Congress on Image and Signal Processing* (Vol. 3, pp. 445-448). IEEE.

[43] Zhu, J.Y., Park, T., Isola, P., Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2223–2232.