DATA MINING AND MACHINE LEARNING II Content Loss in Neural Style Transfer using CNN

Karan Sonawane
x22167005
MSc in Data Analytics

Shreyansh Gupta
x21239347
MSc in Data Analytics

Soumiya kanwar
x21218323
MSc in Data Analytics

***Abstract*—In order to represent the content of one image in the creative style of another, neural style transfer uses deep convolutional neural networks (CNNs). With having the ability to produce appealing outputs, most neural style transfer techniques experience significant loss of important features from the original content image. The minute details of the material are twisted and lost as style shifts are made to fit the creative elements of a reference style image. The output images' visual quality is significantly lowered by this ongoing content loss during neural style transfer. Perceptual loss functions on high-level CNN activations to better retain content structure, novel network structures for balanced style and content transfer, and updated training procedures are some of the new techniques used in recent research to try to solve the problem. However, quantitative study shows, current approaches still show a clear loss of content features from the original image. The need to better understand the primary causes of content loss in CNN-based style transfer and investigate new mitigation methods arises. Style parameterization led by content structure, style mixing for balance, matching higher-order feature distributions, and specialized network architecture are some possible paths. For better neural artistic style transfer, improving content preservation is still a challenge.**

Keywords: **Neural style transfer, Content loss, Convolutional neural networks, Image stylization, Artistic style transfer**

I. INTRODUCTION

Neural style transfer is a class of deep learning technique that render image content in different artistic style. With support of deep convolutional neural networks (CNNs), Neural style transfer has allowed the creation of striking images that blend content of one image on to the artistic essence of another. Style transfer involves manipulating image representations extracted from a pre-trained CNN to match the style of one image and the content of another. While capable of producing aesthetically impressive results, most neural style transfer approaches suffer from significant destruction of the original content image. As the CNN transformations are applied to match the reference style, key details in the content get distorted and lost. This content loss during style transfer severely degrades the quality of results.

Several recent works have attempted to address the problem of content preservation. Some methods have proposed using perceptual loss functions defined on high-level CNN features to improve content retention. Others have developed novel network architectures and training procedures for balanced style and content transfer. However, as we quantitatively demonstrate, existing techniques continue to suffer from noticeable content loss.

MOTIVATION:

Since the content loss poses a critical challenge, The motivation arises to preserve this loss and improve the quality of the images. For example, stylizing a portrait photo while maintaining the identity and facial features of the subject remains difficult for CNNs. The final image may adopt the colors and features of the resulting image, but the content gets dismantled or altered. Previous work has tried to address this through modified objectives functions and network training procedures. But balancing style transfer effectiveness & pixel preservation remains an open challenge. This research is motivated by the need to better interpret and reduce content loss in CNN-based neural style transfer.

II. LITERATURE REVIEW

This Paper by Ruta et al. [1] presents new techniques for neural style transfer – HyperNST. HyperNST leverages hyper networks and metric spaces. It focuses on transferring a single style to a content image. However, a different paper by Tuyen et al. [2] presented DeCorMST, which uses a deep correlation loss function. It generates multiple stylized outputs from one content/style pair. The key relationship between the two papers and the loss of content in the CNN-based style transfer approach includes suboptimal tradeoffs between retaining style versus content. HyperNST demonstrates superior preservation of content structure by using semantic maps to guide style parameterization. This helps mitigate the loss of content details. DeCorMST does not directly address content loss, though its style blending approach could potentially balance style and content retention if optimized well.

A paper by Zhao et al. [3] utilizes a local style loss function to retain intricate style details and a global style loss function to capture structural aspects. They proposed combining local and global style loss functions. In relation to this Paper, Zhao's local style loss function is intended to help preserve intricate style details, while the global loss captures content structure. This relates to minimizing loss of style and content.

The Paper [4] demonstrates the benefits of using very deep convolutional neural networks (CNNs) with up to 19 layers for image classification on ImageNet. It shows that increasing network depth leads to higher accuracy, provided other hyper-parameters are properly tuned. The very deep networks outperform shallower ones and set new state-of-the-art results on ImageNet, surpassing previous best-performing architectures. The learned features generalize very well to other datasets, outperforming prior hand-engineered and learned features. In summary, the key contributions are showing the efficacy of very deep CNNs through strong empirical results on ImageNet, and demonstrating their learned features transfer well to other tasks. This highly influential work helped drive the adoption of much deeper architectures for CNNs in computer vision.

The Paper [5] proposes a neural algorithm for artistic style transfer using convolutional neural networks (CNNs). It shows CNNs can learn separable representations for image content and style. The content representation captures high-level image information, while the style representation matches texture statistics. New images are synthesized by optimizing to match the content representation from one image and style representation from another. This enables combining the content of a photo with the artistic style of a painting. Adjusting the weighting allows control over the content vs style balance. In summary, the key contribution is demonstrating CNNs can separate image content from style and use this to enable artistic style transfer by blending arbitrary content and style images.

The Paper [6] improves on prior work on learning feed-forward networks for artistic style transfer. It identifies issues with artefacts in generated images when training style transfer networks. The key idea is to replace batch normalization with instance normalization in the generator network. Instance normalization normalizes feature maps per instance rather than across a batch. This removes instance-specific contrast information from the content image. Removing this contrast dependence simplifies the learning problem. Replacing batch norm with instance norm substantially improves results for style transfer. Two different generator architectures both benefit from instance normalization. The resulting method produces high-quality stylized images in real time. In summary, the main contribution is showing instance normalization stabilizes training of feed-forward style transfer networks, enabling high-quality real-time artistic style transfer. w6

The Paper [7] proposes StyleBank, an explicit learnable representation of artistic styles in neural style transfer. StyleBank consists of multiple convolutional filter banks, each capturing the style of one artwork. An autoencoder separates content from style. StyleBank then renders content images in different styles. This decouples style from content and enables real-time transfer, unlike optimization-based approaches. The explicit style representation enables style fusion at global and regional levels. The network can efficiently learn new styles through training just the new style filter bank. Results match or improve the quality of previous optimization-based and feed-forward approaches. StyleBank links neural style transfer back to traditional texton-mapping methods for texture synthesis. In summary, the main contributions are the proposed StyleBank explicit style representation, which enables real-time high-quality style transfer, flexible style fusion, and efficient learning of new styles.

The Paper [8] improves neural style transfer by better matching feature distributions between content and style images. It categorizes prior methods into ones using MMD, moment matching, and optimal transport. These only match first and second-order moments (mean and variance). The Paper proposes matching higher-order moments using Central Moment Discrepancy (CMD). CMD matches all moments up to a desired order, not just low-order ones. This aligns content and style distributions better theoretically and visually. Results show CMD captures texture and stroke patterns better. A user study finds CMD preferred over other methods. In simple terms, the main idea is to match higher-order moments, not just mean and variance, between content and style. This matches the distributions better and produces improved style transfer results.

The Paper [9] describes the training of deep convolutional neural networks (CNNs) on the ImageNet dataset for image classification. The architecture has five convolutional layers and three fully-connected layers, with over 60 million parameters. Several innovations are introduced, like ReLU non-linearities, overlapping pooling, and dropout. These improve training speed and reduce overfitting. The network achieves significantly better performance on ImageNet compared to previous state-of-the-art methods. Combining multiple models by averaging their predictions further improves performance. Visualizations show the network learns meaningful features in convolution layers. Results generalize well to other datasets not used for training. In summary, the Paper shows deep CNNs can achieve record-breaking image classification performance on large and complex datasets like ImageNet. The network design ideas have been highly influential.

The Paper [10] proposes a method to invert image representations by reconstructing the input image from the representation. This is formulated as an optimization problem using gradients and natural image priors. The method is applied to visualizing and analyzing SIFT, HOG, and deep convolutional neural networks (CNNs). For SIFT and HOG, it recovers significantly better visualizations than a prior inversion technique. When applied to CNNs, the inversions shed light on the information captured at each layer. Lower layers retain photographic details; higher layers show increasing invariance and abstraction. CNN codes still contain significant visual information about objects, even in deep layers. Reconstructions from subsets of neurons reveal localized receptive fields. The inversion method provides a tool for analyzing and visualizing image representations. In summary, the key contributions are the proposed inversion technique and insights obtained from applying it to interpret internal representations learned by deep neural networks for image analysis.

This Paper [11] proposes neural style transfer methods that don't depend on the style image, so they can generate stylized images for any content-style pair. Uses a unidirectional GAN with cyclic consistency loss to learn mappings between domains. Achieves style transfer with a single feed-forward pass, unlike optimization-based approaches. Can generate stylized images with multiple styles using the same model. The model is lightweight and efficient to train compared to previous style transfer networks. Ensures semantic accuracy of stylized images through adversarial and cyclic consistency losses. Evaluated on COCO and WikiArt datasets and shows improved quality and accuracy over baselines. Limitations include model complexity, large dataset requirements, and lack of widespread adoption so far. Significant contribution in advancing neural style transfer through efficient training of style-independent models with semantic accuracy.

The Paper [12] proposes a neural algorithm for artistic style transfer using convolutional neural networks (CNNs). Uses CNNs to separate image content and style from natural images. Content representation captures high-level image information from deeper CNN layers. Style representation matches texture statistics based on correlations across layers. New images are synthesized by minimizing the distance between content/style CNN features. Results show high-quality stylization combining arbitrary photos and artwork styles. Demonstrates CNNs can effectively learn representations for image style transfer. Provides a new formulation of style transfer as matching CNN feature distributions. The approach is flexible to different content and style images. Limitations include computational expense and artefact generation. Highly influential work, with over 10,000 citations and impact on applications. Inspired many follow-ups works to improve neural style transfer. In summary, the key contribution is a neural algorithm for artistic style transfer using CNNs to separate and recombine content and style. This enabled new state-of-the-art results and sparked significant research interest. Image style transfer involves combining the content of one image with the style of another using deep learning.

Earlier methods in this Paper [13] relied on traditional image processing techniques and were computationally expensive. Convolutional neural networks (CNNs) revolutionized this field by learning to extract semantic content and style features from images. Transfer learning using pre-trained CNNs like VGGNet simplified style transfer by leveraging learned feature representations. A key contribution is the insight content and style can be independently manipulated in CNN activations. This enables the creation of novel images blending content and style from diverse images. Style transfer reveals how deep image representations are learned in CNNs. Challenges remain in improving quality and efficiency and expanding applications. Future directions involve enhancing algorithms, reducing resource needs, and wider adoption. Potential for integrating style transfer into broader domains like mobile apps. Overall, CNNs have enabled great progress in neural style transfer through learned feature representations. But opportunities exist to further improve techniques and applications going forward. In summary, I highlighted how CNNs and deep learning have advanced style transfer, key insights like independent content/style manipulation, remaining challenges and opportunities, and the potential for expanding these methods to new domains in the future.

This Research [14] proposes a new approach to neural style transfer using the Wasserstein distance to match feature distributions. It argues matching feature distributions is better than traditional feature matching methods. Wasserstein distance measures the distance between style and content feature distributions. Results show the approach transfers style more accurately while preserving details. Interprets style as simply the distribution of features. Views neural style transfer as a generative adversarial network problem. The new approach is a valuable contribution, outperforming traditional methods. Limitations include lack of evaluation so far, technical complexity, and not addressing all issues. It Does not consider factors like colour or composition that influence style. It does not address style transfer for videos or control over the amount of transfer. Overall, an important advancement despite limitations. A new interpretation of the style is valuable. In summary, the key ideas are using Wasserstein distance to match distributions for style transfer and interpreting style as a distribution. This is a promising new approach but has some limitations to be addressed in future work.

The Paper [15] proposes using feature perceptual loss based on deep CNNs to improve image quality in variational autoencoders (VAEs). Leveraging pre-trained CNNs provides greater perceptual understanding compared to the pixel-wise loss. This captures visual attributes better and enhances image quality. The learned latent space also shows the ability to represent semantic information, enabling applications like facial attribute manipulation. While results are competitive, further improvements may come from combining feature loss and GANs. Overall, the Paper demonstrates the promise of VAEs for high-quality image generation and learning semantic representations when augmented with deep CNN feature loss. In summary, the key ideas are using deep CNN feature loss to improve VAE image quality and showing the latent space captures semantic information - advancing VAE generative abilities.

The research study [16] presents a revolutionary deep-learning approach for generating artistic images and transferring styles, specifically for turning sketches into stylized graphics. The study accurately translates images on tiny datasets using a Conditional Confrontation Network (CCN) model that integrates Generative Adversarial Networks (GANs). It contrasts their style transfer process with the Gatys et al. baseline method, noting considerable speed gains. The CCN and style transfer models are also combined in the article to demonstrate improved picture production and creative stylization. In conclusion, the authors highlight the effectiveness of the method and suggest areas for further Research, such as capturing detailed style features and accelerating training convergence.

In this paper [17], the authors have proposed a novel method to create specific network parameters through one feed-forward propagation for neural style transfer. To overcome the issue of enormous iterations of stochastic gradient descent, they have built a meta-network which creates an image transformation network directly. These meta networks can handle an arbitrary new style within 19 milliseconds on one modern GPU card.

This Paper [18] introduced a neural algorithm of artistic style that can separate and recombine the image content and style of natural images. This algorithm allowed to produce high-quality images that combine arbitrary photographs with numerous well-known artworks.

III. METHODOLOGY

In the evolving world of data-driven research, the Knowledge Discovery in Databases (KDD) process stands as a key methodology. It is a systematic approach that converts raw data into meaningful patterns and knowledge, going through several stages. From understanding the domain, selecting and preprocessing data to the core data mining phase, followed by evaluation and knowledge application. KDD's principles are immensely adaptable, making it suitable for deep and diverse domains. In the context of this paper, the framework of KDD is used to guide the exploration. By sticking to its stages, a comprehensive, rigorous, and systematic analysis of the process is ensured. By the means of this methodology, the aim is to not only enhance the understanding but to also test the versatility and efficiency of the process.

A. *UNDERSTANDING THE DOMAIN -*

Neural style transfer has bridged the gap between art and technology, allowing artists and even novices to explore and create novel artworks by merging diverse styles. However, one of the core challenges of NST is ensuring that the content of the original image remains recognizable after the style has been applied. While some styles work seamlessly with any content, others might produce artifacts or undesired results. Understanding and mitigating these inconsistencies is crucial.

B. *DATA SELECTION*

In Neural Style Transfer, the choice of data is twofold: the content image that provides the structure and the style image that dictates the artistic style. The success of the stylization heavily relies on the quality and compatibility of these images. In this research, two types of datasets are sourced – "Content Images" and "Style Images". For optimal results, both content and style images should ideally have similar resolutions and aspect ratios. This ensures that the style patterns are transferred appropriately without distortions. To adhere to this fact, 'food and drinks' class was selected among the four different classes present in the dataset as the images in this class were perfectly fitting with the style images.

On the other hand, style image dataset contains a number of artists and their specific art works. Using diverse content images, such as portraits, landscapes, buildings, and more, allows for a more robust evaluation of how well the style is applied across different content types. Thus, ten random art works were selected from each one of those artists.

When sourcing images, especially art pieces, it's essential to consider copyright restrictions and give appropriate credits when necessary. Keeping this in mind Both the datasets are sourced from Kaggle and are publicly available.

C. *DATA PREPROCESSING*

In machine learning and deep learning applications, the preparation and transformation of raw data into an appropriate format is a very crucial step. This step ensures that the algorithm intakes information in a step-wise and optimized manner. Within this neural style transfer framework, there underly two primary pre-processing operations: Image loading and Normalization.

*1) **Image Loading :*** The first and foremost step of pre processing involves the extraction and decoding of raw image data to render it suitable for subsequent operations. A designated function, '**load_image**', undertakes several tasks:

- **Raw data reading -**
  The image is fetched from the source destination, utilizing TensorFlow's **'tf.io.read_file'** function. This results in a raw byte representation.
- **Decoding –**
  This raw byte sequence is then decoded into a tensor structure using **'tf.image.decode_image'**, ensuring an RGB format with all the three channels.
- **Batch dimension –**
  Even when the framework processes singular images, deep learning models usually anticipate a batched input.

*2) **Normalization:*** Normalization is done to ensure convergence stability and speed for neural networks.

A critical step is performed within the **'load_image'** function. With the use of **'tf.image.convert_image_dtype'** the image tensor's datatype is converted to 'float32'. Perhaps, this function scales the original image pixel intensities, which conventionally lies between 0 and 255, to a normalized range of [0,1].

The choice of this range is supported by the observation that neural networks, particularly when activating functions like sigmoid or tanh. This provides enhanced performance due to the prevention of gradient saturation phenomenon. These steps carefully transform raw image data into structured and normalized format, laying the foundation for effective neural style transfer operations.

D. *MODEL BUILDING*

With the completion of data selection and preprocessing steps comes another pivotal phase in the KDD process where algorithms and techniques are employed to recognize patterns or relation within the data. Within the reach of this neural style transfer study, this step involves the applications of deep learning models to combine the distinct artistic styles and content.

*1) **Model selection** :* The basic aspect of the data mining endeavor is the choice of an appropriate model that captures the quality of the problem at hand.

For this study, the selection took turn towards a pre-trained model from TensorFlow Hub due to various reasons:

- **Transfer Learning** : pre-trained models are trained on a vast number of datasets, capturing generalized features that can be tuned or adapted to specific tasks, even with limited data. This helps in achieving better performance without the need for extensive training.
- **Efficiency** : Using an already trained model typically reduces computational demands and accelerates the style transfer process.
- **Model credibility** : The chosen model from TensorFlow Hub, 'https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2', is a well-known and tested model for image stylization, ensuring reliability and quality in results.

*2) **Model application**:* With the model in place, the next step in the data mining process is the application phase, where pre-trained model is employed on the dataset.

- **Input Presentation** : The preprocessed content and style images are presented as inputs to the model. The content image provides the foundational structure or subject, while the style image imparts its unique artistic flair to the final result.
- **Stylization Process** : The model, using its internal architectures, typically convolutional layers and other mechanisms, overlaps and combines the two images. This involves retaining the content of the content image while overlaying it with the style of the style image.
- **Output Retrieval** : The end product of this procedure is a 'stylized image'—an artistic version of the original content image.

In conclusion, the data mining steps in this neural style transfer study demonstrates the selection and application of a reliable pre-trained model to execute the complex task of fusing content and style. The result is an image that combines the main features of one picture with the artistic style of another.

IV. EVALUATION & RESULTS

A. *Visual Evaluation*

Examining the stylized image to determine how well the style from the reference image has been transferred while retaining the content of the original image. As shown in Fig. 1, Fig .2 and Fig.3

B. *Quantitative Evaluation*

- **Mean Squared Error (MSE)** This metric calculates the pixel-wise difference between the content and style images after resizing the style image. A lower MSE indicates a closer match between the two images in terms of pixel intensity.

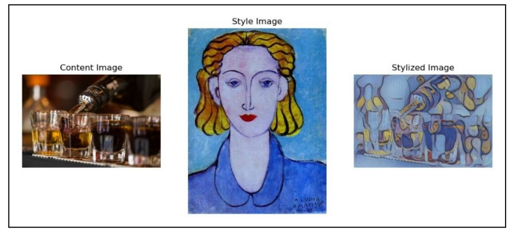

Fig. 1

Fig. 2

Fig. 3

Fig. 4

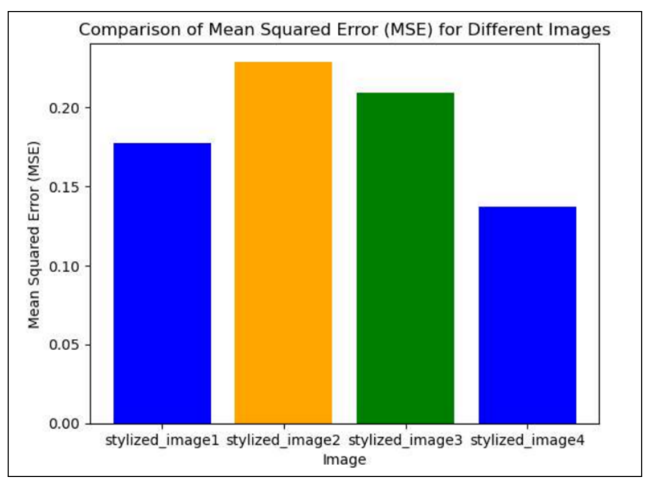

Fig. 5: MSE

- **Perceptual Loss** Using the VGG16 model, the perceptual loss between the content and stylized images is calculated. This metric measures the difference in content representations of the two images.

Sample	Perceptual Loss
1	73.822
2	44.781
3	44.088
4	171.49

Fig. 6: Table 1

- **Gram Matrix Loss** This metric measures the difference in style representations between the stylized image and the style image using their respective gram matrices

Sample	Gram Matrix Loss
1	0.00085
2	3.13237
3	0.00132
4	0.00051

Fig. 7: Table 2

- **Total Variation Loss** This measures the amount of noise in the stylized image. A lower total variation loss indicates a smoother image.

Sample	Total Variation Loss
1	67402.312
2	406243.75
3	852280.37
4	277654.90

Fig. 8: Table 3

V. CONCLUSION & FUTURE WORK

The structured framework of the Knowledge Discovery Databases (KDD) process served as the platform for our research into neural style transfer, which resulted in insightful information about the relationship between technology and artistic expression. With the help of a pre-trained model from TensorFlow Hub, we were able to successfully combine several artistic styles into content images. The effectiveness of this model in maintaining the integrity of the content while smoothly incorporating the complexity of the selected style was highlighted by visual evaluations. Quantitative measurements added to our knowledge by focusing on the model's strengths and places for improvement. Furthermore, through the calculation of various losses, such as perceptual, MSE, gram matrix, and total variation loss, we quantitatively measured the quality of the style transfer, offering a more technical perspective on the results.

In context of our results, several interesting directions for further study are revealed. Although the majority of our research focused on a particular pre-trained model from TensorFlow Hub, there is room to expand the scope of this study. Future research could explore other neural architectures or even models developed using various datasets, with the goal of identifying minute variations in performance. The world of hyperparameters is another amazing aspect. The balance between content preservation and style immersion could be adjusted to produce even more captivating results by going further into characteristics like style weight, content weight, and overall variation weight. Beyond the simple transfer of one style to another, the combination of various artistic styles onto a single content image opens up an interesting new world and opens the way for artworks that are a mix of various artistic inspirations.

In addition, neural style transfer's accessibility also is interesting. It can be possible for people, even those without significant technological knowledge, to take part in this art by encouraging the development of interactive platforms or applications and selecting their own content and style combinations. Last but not least, the changing world of video poses a brand-new difficulty. Imagine that a video's every frame was styled in real time while yet keeping a seamless flow and the video's overall integrity. Such a development of our research might significantly change the way we see video content by enhancing motion with the arts.

A. *ACKNOWLEGMENT*

We would like to thank Professor Noel Cosgrave for providing guidance and instruction for the data mining and machine learning module project.

REFERENCES

[1] D. Ruta, A. Gilbert, S. Motiian, B. Faieta, Z. Lin, and J. Collomosse, "Hypernst: Hyper-networks for neural style transfer," in *Computer Vision – ECCV 2022 Workshops*, ser. Lecture Notes in Computer Science, L. Karlinsky, T. Michaeli, and K. Nishino, Eds., vol. 13801. Springer, Cham, 2023.

[2] N. Q. Tuyen, S. T. Nguyen, T. J. Choi, and V. Q. Dinh, "Deep correlation multimodal neural style transfer," *IEEE Access*, vol. 9, pp. 141 329–141 338, 2021.

[3] H. H. Zhao, P. L. Rosin, Y. K. Lai, M. G. Lin, and Q. Y. Liu, "Image neural style transfer with global and local optimization fusion," *IEEE Access*, vol. 7, pp. 85 573–85 580, 2019.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Vision and Pattern Recognition (cs.CV)*, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[5] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015. [Online]. Available: http://arxiv.org/abs/1508.06576

[6] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016. [Online]. Available: http://arxiv.org/abs/1607.08022

[7] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stylebank: An explicit representation for neural image style transfer," *CoRR*, vol. abs/1703.09210, 2017. [Online]. Available: http://arxiv.org/abs/1703.09210

[8] N. Kalischek, J. D. Wegner, and K. Schindler, "In the light of feature distributions: moment matching for neural style transfer," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2021, pp. 9377–9386.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet

[10] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” *arXiv preprint arXiv:1412.0035*, 2014, arXiv: 1412.0035 [cs]. [Online]. Available: http://arxiv.org/abs/1412.0035

[11] P. N. Deelaka. (2023) Neural artistic style transfer with conditional adversaria. arXiv.org. Accessed: Aug. 13, 2023. [Online]. Available: https://arxiv.org/abs/2302.03875

[12] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016, pp. 2414–2423.

[13] A. Khan, A. Kumar, N. Kulkarni, A. Kamat, and D. Kumar, “Image style transfer using cnn,” *http://www.ijert.org*, 2023.

[14] E. Huang and S. Gupta. (2023) Style is a distribution of features. Accessed: Aug. 14, 2023. [Online]. Available: https://arxiv.org/abs/2007.13010

[15] X. Hou, L. Shen, K. Sun, and G. Qiu, “Deep feature consistent variational autoencoder,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, USA, 2017, pp. 1133–1141.

[16] R. Wang, “Research on image generation and style transfer algorithm based on deep learning,” *Open Journal of Applied Sciences*, 2023, available: https://www.scirp.org/journal/paperinformation.aspx?paperid=94650 (accessed Aug. 14, 2023).

[17] F. Shen, S. Yan, and G. Zeng, “Neural style transfer via meta networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Shen_Neural_Style_Transfer_CVPR_2018_paper.html