International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST523103146

# Artistic Style Transfer using Deep Learning and Style Fusion- a Review

Mohammed Mutahar[1], R Aafrein[2], Maneesh Ragavendra K[3], Mohammed Mustafa Jafer[4], Dr. R. Chinnaiyan[5]
[1,2,3,4]B. Tech CSE – AI ML, School of Computer Science and Engineering, Presidency University , Bangalore, Karnataka, India
[5]Professor & Head- RC, School of Computer Science and Engineering, Presidency University , Bangalore, Karnataka, India

ARTICLEINFO

Article History:

Accepted: 01 June 2023
Published: 11 June 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
785-792

ABSTRACT

In recent years, after the study 'A Neural Algorithm of Artistic Style' published by Gatys et al. in 2016b, research on style transfer boomed drastically. Style transfer is the process of copying an art style from a 'style image' to the contents of the 'content image' and producing a 'draft image' that is on par with respect to quality expectations. This paper explores different techniques of achieving style transformations namely Style Fusion and Convolutional Neural Networks (CNNs). Although CNNs are the state- of-the-art architecture to tackle cognitive visual tasks, and that they clearly perform much better than most conventional algorithms, the imageprocessing-based style fusion method comes close to the CNN in terms of image output quality and supersedes in terms of time and computation and resources complexity. The procedure of both of these methods has beendiscussed in detail in this paper and it was concluded that CNNs have a lot more room for improvement that can be facilitated by the availability of better and larger datasets.

Keywords : Style transfer, style transformations, style fusion, convolutional neural networks, cognitive visual tasks, image processing.

## I. INTRODUCTION

Artistic style fusion using deep learning and style transfer has emerged as a captivating field in computer vision and artificial intelligence in recent years. Style transfer refers to the process of transforming the visual appearance of an image orvideo to emulate the characteristics of a specific style or artistic representation. The style being copied from is referred to as style image and the style being copied to is referred to as content image. There are numerous techniques to achieving style transfer, deep learning techniques involve the use of Convolutional Neural Networks (CNNs) whereas patch pattern matching can also be performed solely by image processing paradigms [1]. Style transfer has found applications in the entertainment industry, enabling the generation of stunning special effects, immersive virtual reality experiences, and the transformation of ordinary photographs into artistic masterpieces. The rise of popularity of image filters on applications such as SnapChat and Instagram has also led to growth ofstudy in the field of style transfer. Researchers have been in pursuit of making the process less resource exhaustive since they now need to run on devices thatdo not offer a lot of computational resources like mobile-phones and smartphones while also making sure that the quality of style transfer remains more than satisfactory. Style transfer is also seen as a way topreserve dying art styles. Style transfer has been inthe spotlight of hot research ever since Gatys et al. [2] first published his paper, "A Neural Algorithm of Artistic Style" in 2015 stating how CNNs can be used to nail the task. There have also been other methods of achieving style transfer that have been later developed that will be discussed further in this paper.

## II. RELATED WORKS

Numerous researchers and experts have conducted investigations into image art style transfer processing, each proposing unique approaches. For instance, Sun et al. [3] introduced a CNN architecture with twopaths that extract object features and textures from images, enabling style recognition and outputting image categories. Lu et al. [4] addressed the limitations of existing deep learning models and proposed a multi-image training model based on an improved neural network, enabling the extraction of multiple image features. Castillo et al. proposed a model for artistic target transfer, focusing on artistic transfer and fusion of individual image objects. For dynamic video style transfer, Ruder et al. extended Gatys' work by extracting artistic style from video images and transferring it to the entire video.

In a study conducted by Elad and Milanfar [5], an alternative way to handling the style-transfer task was discussed that included the derivation of an analogous objective function for minimizing energy. They performed optimization in stages and patches from styling image were mapped onto regions in content image. Although when this study was conducted CNNs were the state-of-the-art architecture for handling style transfer, the method derived in this paper was closer in quality to CNNs but was also much faster and flexible.

In another study conducted by Zhang et al. [6], component analysis approach was used to decompose the style image into 3 parts, draft, paint and edge. Then, only the paint and edge components were transferred to the content image. This method of style transfer preserves the content image well.

## III. METHODS

There have been many studies that have proposed different techniques of tackling the task of style transfer. Style transfer can either be performed by CNNs or image processing paradigm to synthesize style from texture synthesis. Both the techniques havebeen discussed in detail in this paper in further sections.

A. Image Processing and Style Fusion

Image processing has been prevalent in many papersin the past years after Kwatra et al. first in 2005 published a paper on texture synthesis which has since then been expanded into style synthesis for styletransformation. The results are similar to neural networks, but the speed and robustness of thealgorithms are high.

The core algorithm has 6 different steps: Style fusion, Patch Matching, Style Synthesis, Content Fusion, Color Transform, Denoising.

1) First step, color transform: Applying color transform from style image to content image. The

transformation function from Elad's work is a good minimization function. The algorithm was found to minimize the cost function to produce an image that balances style and content according to a weight mask.

$$\frac{1}{c}\Sigma_{(i,j\,\epsilon\Omega_{L,n})}\min||R_{ij}^{n}X-Q_{kl}^{n}S||_2^r+||X-C||_W^2+\lambda\{X\}|$$

In this equation:
X represents the estimated image
C represents in content image
S represents the style image
$Rijn$ represents extraction of the i, j-th patch of size nxn
$Qkln$ represents extraction of the k, j-th patch of size nxn
W represents the weight mask
L represents the working scale
r and c represent regularization and normalization factors
λ represents image prior statistics

2) Second step, add Gaussian noise: Noise is the addition of a statistical variation in the image created from random process. Adding noise is done to allow patch matching. Patch matching is required to avoid getting repeated patterns, especially in areas having high uniformity.

Fig. 1 – Added Gaussian Noise

Using noise, we generate a "draft" image that does not contain the content rather a blurred version of it over which the style is applied from the style image. Style fusion depends upon and uses the draft image, but in this stage, we are not concerned about preserving the content. The image generated will be later used in the algorithm.

3) Third step, style fusion: Style fusion uses the "draft" image from the second step and applies a weighted average in each iteration in order to maximize the style transferred while preserving the content. A 75% style fusion would convert almost every aspect of the content image to a style from the style image, which will result in a huge loss of original content from the content image. A 25% style fusion retains much of the content from the original image and applies the style only to the most resembling parts of the image between the style image and the fusion image.

Fig. 2 – A comparison of 75% style fusion vs 25% style fusion

4) Fourth step, patch matching: This step is all about extracting the areas of the style image that matches or closely mimics the areas from the content image. We do this to analyze which area of the style image can be mapped onto the area of the content image so that our style transformation is sound logically. This may include identifying what part of the style image can be blindly copied onto the content image, for instance the art style of sky in the style image can be blindly put into the content image in most cases.

While matching patches we need to make sure that our style does not repetitively iterate over the content image since the areas close to similar patches may also be similar. To make sure of this, we deploy the nearest-neighbors (NN) search in the style image. The patch with the lowest L2 norm is chosen from the style image to with respect to the content patch.

The optimization problem can be formulated as:

$$\{k^*,l^*\}=Argmin_{(k,l)}||R_{ij}^{n}X-Q_{kl}^{n}S||_2$$

Skipping similar patches in close vicinity generates better results than when considering every possible patch from the image.

Patch matching can also be optimized by the principal component analysis approach by projecting every possible style into a smaller space before computing the norm. This saves a lot of memory while minimization. We need to keep in mind that even though a patch may be matching the content image, its energy contribution to obtaining the estimated image might be insubstantial, therefore a threshold for including the top 'x' number of vectors is to be included. This can be decided by allowing only those vectors to contribute whose energy contribution is above 'y'% and all other vectors can be considered to carry non- critical information.

After calculating the L2 norm of each patch, we add noise and generate draft images during run-time. The noise (Gaussian noise) added should be a percentage of the minimum L2 norm metric.

5) Fifth step, style synthesis: This step includes estimating the output image using the style patches generated using NN in the previous step. Which means that each i, j patch in the output image should roughly match the corresponding NN style patch. We can iterate over re-weighted least squares as used by Kwatra et al. [7] in their paper published in 2005,

$$\chi = Argmin_{\chi} \Sigma_{(i,j \epsilon \Omega)} ||R_{ij}^{n} X - z_{ij}||_{2}^{r},$$

which can be formulated as above.

6) Sixth step, content fusion: Ensuring that the original image is preserved is crucial at this stage of the process. We can do this by either taking the weighted average of the areas between the estimated style image (draft image) and the original content image or each and every individual pixel. Although calculating the weighted average pixel-by-pixel is a computationally demanding process, it provides a much better insight on the quality of the task being performed by the algorithm.

Fig. 3 - A comparison of content fusion before and after applying the Gaussian mask

The foreground of the content image is isolated from the background of the content image by the value of weights assigned to them. The foreground is assigned a larger weight and the background is assigned a smaller weight in comparison. The edges also contain information on the separation of background and foreground therefore they are also assigned a large weight. Content fusion is then the extraction of zero-crossings of the Laplacian of Gaussian operations. To make the transition between high-content areas (areas with high weight) and high-style areas (areas with low weight) smoother, a Gaussian mask is applied to the image.

The equation formulated is used to apply content fusion and for updating the estimate.

$$\mathrm{X} = (W + I)^{-1}(\mathrm{X} + WC)$$

7) Seventh step, color transfer: The color palette of the final image can be chosen to be anything desired, although the colors from the original style image give the best appeal to style transformation. We could also choose to keep the colors from the original content image and just copy the style onto the image. The MATLAB function 'imhismatch' can be used for color palette selection and color transformation since it transforms image I (input) such that the histogram of image J (output) matches the histogram derived from the style image or the content image.

8) Denoising: The image at this stage has been heavily processed and needs to be reverted to close to its original state. This is done by running a filter that smooths weak edges and preserves strong ones over the final output image, preferably a bilateral filter.

Selective Gaussian Blur and Domain Transfer Filter [8] are some examples.

A bilateral filter can be defined by the following formula:

$$I^{filtered}(x) = \frac{1}{W_p}\Sigma_{x_i \epsilon \Omega} I(x_i) f_r(||I(x_i) - I(x)||) g_s(||x_i - x||)$$

Where;

Ifiltered is the filtered image

I is the output image to be filtered

x is the coordinated of the current pixel to be filtered

Ω is the window centered in x

fr is the range kernel for smoothing differences in intensities

gs is the spatial kernel for smoothing differences in coordinates

Before finishing out, we need to keep in mind that for a given patch size, the larger the scale, larger will be the extracted features. To see what works best foryour model you will need to work from towards a finescale from a rough one. We can also ease the smoothing process by choosing different patch sizes around the edges. This increases the variability in distribution of patch choices and refinement of detailsin style transfer. Varying these parameters gives the algorithm an array of options to deploy during patch matching.

This concludes the Image Processing and Style Fusion method; we will now look at how CNNs can be used for style transfer and how they differ from the aforementioned method.

B. Convolutional Neural Networks

CNNs have been a popular choice for image related tasks ever since their inception, hence it is not a matter of surprise when they excel in the field of image transformations as well. CNNs are good at extracting features from images on their own. This property of CNN is leveraged to create quality images from content and style image. The approach to creating a CNN that style transforms images is rather straight forward. Features from content image and style image are extracted by the same network. These features are used to map an estimated output image called the target image. A separate loss function is used for both the content image and the style image called the content loss function and the style loss function respectively. The network adjusts the performance by updating the weights until a satisfactory target image has been achieved.

1) Content loss function: This loss function is used on the content image only. The loss function is based on the input and output content characteristics of each class.

$$L_{content} = \frac{1}{2}\Sigma_{i,j}(F_{ij}^l - P_{ij}^l)^2$$

$F_{ij}^l$ is the trigger function of the i[th] filter at position j in class 1.
$P^l$ , $F^l$ represent their respective features in the class.

2) Style loss function: Even though the formulation of the style loss function happens the same way as the formulation of the content loss function and it follows the same principle, calculating the style loss function is relatively more complicated. Gram matrices can be used to compare the two feature maps obtained after the target image and the style image have been passed to the CNN. A sum of the style of total loss function amounts to:

$$L_{style} = \Sigma_{l=0}^{L} w_l E_l$$

where El is given as:

$$E_l = \frac{1}{4N_l^2 M_l^2}\Sigma_{i,j}(G_{ij}^l - A_{ij}^l)^2$$

Since both these loss functions work on separate images, an aggregation of them is required to synthesize the final image. For synthesizing in a balanced manner, we need to assign weights to the loss functions to avoid one dominating the other to have a new image that is a good representation of its content from the content image through a style from its style image. A simple weighted addition of these loss functions is enough to produce a synthesized loss function.

$$L_{total} = \alpha L_{content} + \beta L_{style}$$

The idea behind the standard conversion model is that it initializes the target image as white noise then proceeds to match the mappings of the content and style image onto the target image. It corrects the loss function until the noisy image correctly translates to the target image. This is generally a slow and time-consuming process and many studies involving deep learning and CNN architecture have since found better and more efficient ways of tackling the task.

## IV. CONCLUSION

The task of style transferring has gained popularity among researchers in the past few years. Many different techniques of achieving good quality images have been formulated. This paper explored two such techniques. CNNs have been excelling in the field of computer vision but we also saw how a more image processing-based approach can be used for tackling the same task with much the same efficiency. CNNs with their robustness and their virtue of extracting features on their own can be improved to a state much better than its current form with the availability of better and larger datasets. CNNs have also been known for reducing the execution time by huge margins that set up new benchmarks that serves as the next level to beat for future image style transferresearch. Current CNN models can be improved to parallelly convert more artwork styles in a single transformation session or the flexibility to adjust the degree of transformation.

## V. REFERENCES

[1]. Abdulhamit Subasia, Bayader Kadasaa, Emir Kremic, “Classification of the Cardiotocogram Data for Anticipation of Fetal Risks usingBagging Ensemble Classifier”, Procedia Computer Science 168 (2020) 34–39

[2]. Alessio Petrozziello, Ivan Jordanov, Aris T.Papageorghiou, Christopher W.G. Redman, and Antoniya Georgieva,” Deep Learning for ContinuousElectronic Fetal Monitoring in Labor”, Preprint, Researchgate

[3]. Artificial Intelligence - Scope and Limitations. IntechOpen, Apr. 24, 2019. doi: 10.5772/intechopen.81872.

[4]. Attallah O, Sharkas MA, Gadelkarim H. Fetal Brain Abnormality Classification from MRI Images of Different Gestational Age. Brain Sciences. 2019; 9(9):231.

[5]. Balachandar S., Chinnaiyan R. (2019) Centralized Reliability and Security Management of Data in Internet of Things (IoT)with Rule Builder. In: Smys S., Bestak R., Chen JZ., Kotuliak I. (eds) International Conferenceon Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore

[6]. Balachandar S., Chinnaiyan R. (2019) Reliable Digital Twin for Connected Footballer. In: SmysS., Bestak R., Chen JZ., Kotuliak I. (eds) International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore

[7]. Centers for Disease Control and Prevention(CDC). https://www.cdc.gov/pneumonia/index.html [8]. Comert Z., Kocamaz A. F., Subha V. (2018). Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. Comput. Biol. Med. 9985–97.

[9]. Daniel LaFreniere, Farhana Zulkernine, David Barber, Ken Martin. “Using Machine Learning to Predict Hypertension

[10]. E.S.L. Gastal and M.M. Oliveira, Domain Transform for Edge-Aware Image and Video Processing, TOG, Vol. 30, No. 4, pp. 169, 2011.

[11]. G Sabarmathi, R Chinnaiyan (2019),Envisagation and Analysis of Mosquito Borne

Fevers: A Health Monitoring System by Envisagative Computing Using Big Data Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 630-636. Springer, Cham

[12]. G. Sabarmathi, R. Chinnaiyan (2016) , Big Data Analytics Research Opportunities and Challenges - A Review, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.6 , Issue.10, 227-231

[13]. G. Sabarmathi, R. Chinnaiyan, Investigations on big data features research challenges and applications, IEEE Xplore Digital LibraryInternational Conference on Intelligent Computing and Control Systems (ICICCS), 782 – 786.

[14]. Gatys, L., Ecker, A., & Bethge, M. A Neural Algorithm of Artistic Style (2015) URL: https://arxiv.org/pdf/1508.06576.pdf

[15]. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., & Yang, M., Universal Style Transfer via Feature Transforms in arXiv:1705.08086 URL: https://doi.org/10.48550/arXiv.1705.08086

[16]. M Swarnamugi, R Chinnaiyan (2019), IoT Hybrid Computing Model for Intelligent Transportation System (ITS), Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018), 802-806.

[17]. M. Elad, P. Milanfar, Style-Transfer via Texture-Synthesis, Google Research, September 21, 2016

[18]. M. Swarnamugi ; R. Chinnaiyan, “IoT Hybrid Computing Model for Intelligent Transportation System (ITS)”, IEEE Second InternationalConference on Computing Methodologies and Communication (ICCMC), 15-16 Feb. 2018.

[19]. M. Swarnamugi; R. Chinnaiyan, “Cloud and Fog Computing Models for Internet of Things”, International Journal for Research in Applied Science & Engineering Technology, December2017.

20]. R.Vani, “Weighted Deep Neural Network BasedClinical Decision Support System for the Determination of Fetal Health”, International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-8 Issue-4, November 2019,8564-8569.

[21]. Ragab DA, Sharkas M, Attallah O. Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers. Diagnostics. 2019; 9(4):165.

[22]. S. Balachandar, R. Chinnaiyan (2019), Internet of Things Based Reliable Real-Time Disease Monitoring of Poultry Farming Imagery Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 615- 620. Springer, Cham

[23]. S.Balachandar , R.Chinnaiyan (2018), A Reliable Troubleshooting Model for IoT Devices with Sensors and Voice Based Chatbot Application, International Journal for Research in Applied Science & Engineering Technology,Vol.6,Iss.2, 1406-1409.

[24]. S.Balachandar , R.Chinnaiyan (2018), Centralized Reliability and Security Management of Data in Internet of Things (IoT) with Rule Builder, Lecture Notes on Data Engineering and Communications Technologies 15, 193-201.

[25]. S.Balachandar , R.Chinnaiyan (2018), Reliable Digital Twin for Connected Footballer, Lecture Notes on Data Engineering and Communications Technologies 15, 185-191.

[26]. T. Sun, Y. Wang, J. Yang, and X. Hu, “Convolution neural networks with two pathways for image style recognition,” IEEE Transactions on Image Processing, vol. 26, no. 9,pp. 4102–4113, 2017.

[27]. V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, Texture Optimization for Example-Based

Synthesis,ACM ToG, Vol. 24, No. 3, pp. 795-802, 2005

[28]. Wang, E., & Tan, N. Artistic Style Transfer in EE 368: DIGITAL IMAGE PROCESSING, STANFORD UNIVERSITY

[29]. Zhang, Wei, Chen Cao, Shifeng Chen, Jianzhuang Liu, and Xiaoou Tang. Style transfer via image component analysis IEEE Transactions on Multimedia 15, no. 7 (2013): 1594-1601

Cite this article as :

Mohammed Mutahar, , R Aafrein[2], Maneesh Ragavendra K[3], Mohammed Mustafa Jafer[4] and Dr. R. Chinnaiyan, "Artistic Style Transfer using Deep Learning and Style Fusion- a Review", International Journal of Scientific Research in Science and Technology(IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 785-792, May-June 2023. Available at doi : https://doi.org/10.32628/IJSRST523103146

Journal URL : https://ijsrst.com/IJSRST523103146