

PAPER • OPEN ACCESS

Research on Neural Style Transfer

To cite this article: Xueyuan Liu 2021 *J. Phys.: Conf. Ser.* **2079** 012029

View the [article online](#) for updates and enhancements.

You may also like

- [Fabrication and characterization of patterned micrometre scale interpenetrating Au-TiO₂ network nanocomposites](#)
Abu Samah Zuruzi, Marcus S Ward and Noel C MacDonald
- [Microarrayed Nanostructured Titania Thin Films Functionalized for Hydrogen Detection](#)
A. J. Monkowski, A. Morrill and N. C. MacDonald
- [STATISTICAL DISTRIBUTION OF SIZE AND LIFETIME OF BRIGHT POINTS OBSERVED WITH THE NEW SOLAR TELESCOPE](#)
Valentyna Abramenko, Vasyl Yurchyshyn, Philip Goode et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



Research on Neural Style Transfer

Xueyuan Liu¹

¹Material Science and Engineering Institute, Jilin University

*E-mail: liuxy1618@jlu.edu.cn

Abstract: The process of using CNN (Convolutional Neural Network) to blend the contents of a picture with different styles is called neural style transfer (NST). The purpose of this paper is to introduce current progress of NST, and introduce in detail the classification of the main NST algorithms based on deep learning, and make qualitative and quantitative comparisons of different algorithms, and then analyze the application prospects of image style migration in related fields, and finally summarize the existing problems and future research directions of NST.

1. INTRODUCTION

The use of image style migration technology has become more and more extensive, usually to migrate the artistic style of one picture to another picture, is the image to ensure that the content characteristics do not change significantly without major changes in the art style. With the rise of deep learning, Gatys et al.[1] used convolutional neural network models to achieve image style migration, which attracted wide attention. NST has been an important research direction for two decades. Before the emergence of neural networks, style migration was a field of non-photorealistic rendering, and then, the neural network-based texture synthesis technology provided new ideas for style migration. Although style migration has a good visual effect, there are still some problems to be solved. How to improve the efficiency of the algorithm under the premise of ensuring the quality of stylized images is the direction that needs to be studied at present. In this article, we will provide an overview of recent developments in NST.



Fig.1.Example of NST algorithm to transfer a style of a Chinese painting onto a given photograph.

2. A TAXONOMY OF NEURAL STYLE TRANSFER ALGORITHMS

The current NST algorithm can be divided into two categories: neural methods based on image iteration and neural methods based on model iteration. This section focuses on the main ideas of the two algorithms and discusses their advantages and disadvantages.



2.1 NEURAL METHODS BASED ON IMAGE ITERATION

Neural methods based on image iteration was first proposed by DeepDream[2], which uses deep neural network to extract image features, and then reverses CNN on random noise images to iterate to update noise image pixels so that they have the moral characteristics of the content image and the style characteristics of the style image.

One of the keys to NST is the representation of style, which is divided into two categories depending on the style loss function: an algorithm based on statistical parameters and an algorithm based on non-statistical parameters.

2.1.1 NEURAL METHODS WITH SUMMARY STATISTICS

Gatys et al. Applying the VGG network and Gram matrix to the style migration, he uses the output of the high-level activation function of the VGG network to represent the content characteristics of the image, and then uses the Gram matrix to describe its style characteristics. NST can be achieved by minimizing the difference between the content characteristics and style characteristics of the generated image and the input image.

The total loss function is defined as:

$$L_{total} = \beta L_{style} + \alpha L_{content} \quad (1)$$

β and α are the weight of the style loss function and the content loss function in the total loss function.

But Gatys et al. algorithms don't maintain fine structure and detail consistency well in stylization, because CNN features avoid losing some low-level information for a moment. And Gatys et al. algorithm does not take the stroke changes of the image, semantic information and depth location information into account, which leads to unreasonable stylization.

Berger and Memisevic et al.[3] have built on Gatys to make the generation of images consistent over time by adding a Markov structure to advanced features for global symmetry texture generation and image season transformation.

Rissuer et al.[4] introduced additional statistical histogram loss to characterize the distribution information of image features, which solved the problem of gram matrix instability, but the algorithm is very complicated in calculation.

In view of the loss of low-level information in content image, Li et al.'s introduction of Laplace loss adds constraints to the underlying features and uses the Laplace matrix to describe the low-level information in the content image, and complements the high-level semantic information extracted in the VGG network in detail.

Csatillo et al.[5] introduce semantic segmentation of instances in Gatys et al. methods to achieve zone-specific style migration. Luan et al. map the semantic features of content images and style images through artificial control, so that style migration occurs in sub-regions with the same semantics, preventing the overflow of block styles in each region. Penhouest and Sanzenbacher et al. have built on this to simplify workflow by introducing automatic segmentation of image semantics.

Li et al.[6] have developed a new explanation for NST through research, which is seen as an area adaptation problem. Suppose the training and test data come from different distributions. The goal of domain adaption is to adapt a model trained on labelled training data from a source domain to predict labels of unlabeled testing data from a target domain. One way for domain adaption is to match a sample in the source domain to that in the target domain by minimizing their distribution discrepancy, in which Maximum Mean Discrepancy (MMD) is a popular choice to measure the discrepancy between two distributions. Li et al.'s algorithm explains the principle of Gram matrix matching from a mathematical level and proves that matching style images and Gram matrix generating images are essentially MMDs that minimize the distribution between two domains. Therefore, the MMD algorithm using different nuclear functions can be used for NST, so that the principle of NST is clearer.

Although the above methods have improved the algorithms of Gatys et al. to solve the problems of instability, loss of detail and lack of semantic information, they have not solved the problems of stroke change and loss of depth location information, which cannot be ignored for the impact of image quality.

2.1.2 NON-PARAMETRIC NEURAL METHODS WITH MRFs

The method based on non-statistical parameters first divides the two images into many area blocks, and then matches the most similar areas between the two images to achieve style migration, which is good in maintaining local characteristics. Non-parametric image-optimisation-based NST is built on the basis of Non-parametric Texture Modelling with MRTs. Li and Wand[7] found that the early traditional MFRs-based style migration method captured only the relevance of each pixel feature, without constraining its spatial layout, and proposed combining MRTs with deep convolutional neural networks (dCNNs). Using the MRTs model, they split the image feature mapping extracted by dCNNs, obtained many area blocks, and matched the regions of the two images by capturing the feature information of the local pixels of the image. The following is MARs loss:

$$L_s = \sum_{l \in \{l_s\}} \sum_{i=1}^m \| \Psi_i(\mathcal{F}^l(I)) - \Psi_{NN(i)}(\mathcal{F}^l(I_s)) \|^2 \quad (2)$$

m represents the total number of area blocks; $\mathcal{F}^l(I)$ is a set of regional blocks of the layer l

in the neural network model, $\Psi_i(\mathcal{F}^l(I))$ represents area block i in $\mathcal{F}^l(I)$; $\Psi_{NN(i)}(\mathcal{F}^l(I_s))$ is the best matching area block of the area block $\Psi_i(\mathcal{F}^l(I))$, which is calculated by normalization. The advantage of it is that it performs especially well for photorealistic styles, or more specifically, when the content photo and the style are similar in shape and perspective, due to the patch based MRF loss. However, when the difference between content image and style image is too large, stylization results are not ideal, and the algorithm does not preserve the details of the image and global semantic information well.

2.2 MODEL-OPTIMISATION-BASED PFFLINE NEURAL METHODS

Although the results of NST method based on image iteration are considerable, efficiency is the main limitation because of the need to iterate on each pixel in the image. Model iteration-based NST solves the problem of speed and calculation cost by reconstructing the stylized results through the image intermediate technique based on model iteration and transferring the image generation process to a trained feed-forward stylized network.

2.2.1 A METHOD BASED ON A FEED-FORWARD STYLIZED MODEL

The representative of the feed-forward stylized model algorithm is Johnson et al. and Ulyanov et al., who share similar ideas, and stylize the image by pre-training the feed-forward stylized model, which differs from the model structure. Johnson et al. used residual blocks and step-by-step convolution in Radford's proposed model structure, while Ulyanov et al. used a multi-scale architecture as a build network.

Johnson et al.[8] based on Gatys' algorithm to improve, pioneered the introduction of a fast style migration method based on the feed-forward stylized model, and introduced the sense of loss function. Ulyanov et al.[8] use a multi-scale architecture to learn the characteristics of input images in different dimensions, which is to generate images with richer detail. Ulyanov et al. then discovered that using instance normalization (IN) instead of the match normalization (BN) in the original model, i.e., normalizing each image individually, could greatly produce image quality. But their algorithms still follow the models of Gatys et al., so they have the same limitations as they do, and the quality of the resulting images is slightly worse than that of Gatys et al.

Although the above build model method is two orders of magnitude faster than the previous style migration method based on image iteration, only specific style images can be generated, which is flexible and time-consuming. So, a single-model multi-style generation network (multiple-style-per-model neural methods) began to appear. It integrates multiple styles into one model to improve the efficiency of the feed-forward network. Domoulin et al. put forward the idea of creating a CIN on the basis of Ulyanov et al., changing only the parameters of the IN layer to generate different styles of image conversion networks in 32. Chen et al. proposed the concept of the StyleBank layer, binding style features to a set of parameters in the StyleBank layer, sharing content features, and training a new StyleBank layer separately to implement a new style migration. With the increase of the number of styles learned and the model parameters more and more, Li et al. put forward a style selection model

that includes various styles and uses image pixels as signal input to control stylized image generation. Zhang and Dana, among others, proposed the concept of the CoMatch layer, first by allowing the model to learn a variety of styles, and then by using the target style image features as signal inputs for NST.

Chen and Schmid and others then proposed any style transfer model for the first time, using a pretrained VGG network to extract multiple activation blocks from each content feature and style feature, then each content activation block, and then matching each content activation block to the most similar style activation block to produce an image, but the model stylized slowly. Huang and Belongies, among others, propose the AdaIN, which enables real-time, arbitrary style transfer of input images thanks to Chen and Schmid methods. But the AdaIN algorithm is data-driven and has limitations when generalizing data that has not yet appeared. In addition, it only changes the mean and variance of feature maps, making it difficult to get a plot with rich detail and complex structure. Li et al. proposed an encoder-based model that used witting and colouring -change (WIT) for the first time to enable any style transfer that did not require training for a particular style. Shen et al. propose a new metanet for any style migration, which has a smaller model size that is convenient to run all the time and on the mobile side. Park et al. better balance global and local style patterns. The style-attentional network (SANet) is proposed, which can be used to match style features to content features flexibly according to the semantic spatial distribution of content images.

2.2.2 A METHOD BASED ON GAN NETWORK

The GAN network consists of generators and identifiers. The former is dedicated to generating false data, the latter is committed to identifying false data, the two in the confrontation of learning and progress together.

Li and Wand et al.[10] get realistic images by training MRFs-based feed-forward networks through adversarial training. The results show that their algorithm is superior to Johnson et al.'s feed-forward generation model algorithm, but the effect is poor in texture look-up because semantic correlation is not taken into account.

Mirza et al. proposed CGAN for image generation, and the model added additional information to both the generator and the identifier to guide the model generation direction on the basis of the original GAN model, but the supervised learning algorithm needed to be trained on pre-processed piles of data sets. Zhu et al. proposed CycleGAN without training on paired data sets, a network of two generators and two evaluators for converting between two domains and two evaluators for distinguishing between pictures in two domains. The model not only requires that the image can be converted from the source domain to the target domain, but also that the target image can be converted back to the source domain. Then, Choi et al. proposed SartGAN, a model that trains on multiple cross-domain data sets for multi-domain transformation. The use of GAN provides a new way of thinking for the field of NST, not only to improve the speed of NST, but also to ensure the quality of generated images. At the same time, GAN, which meets different needs, promotes the application of style migration technology in the real business field.

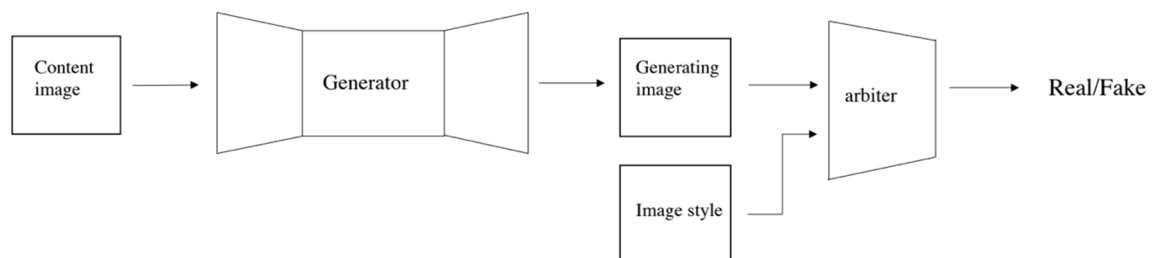


Fig.2. Basic framework of image style transfer based on GAN

3. EVALUATION METHODOLOGY

There are two main methods of NST algorithm, namely qualitative evaluation and quantitative evaluation. Qualitative evaluation depends on the observer's aesthetic judgment, and the evaluation results are related to many factors, such as the age, occupation, educational background, observation conditions and so on. Quantitative evaluation, on the other being, focuses on precise indicators, such as model generation speed, loss changes, and so on. The combination of qualitative and quantitative assessments can make evaluation more comprehensive and objective.

4. APPLICATIONS

In recent years, NST technology has been widely used in commercial applications as it has been improving in image quality and speed.

1、 Digital simulation. Style migration can be used for the reconstruction of ink painting, oil painting, traditional Chinese painting and other artistic paintings. With the help of technology, people can easily draw the image style they want without the need for professional skills.

2、 Film and television production. In the film and television industry, if you can use NST instead of manual hand painting, you can greatly improve efficiency and reduce production costs.

3、 Image processing software. The NST algorithm based on deep neural network makes image hard software no longer limited to simply adding a filter to the picture, but uses neural network to learn the style, texture and other details of its works of art.

5. FUTURE CHALLENGE

Progress in NST is evident, and some algorithms have been used in industry. Although the current results have yielded good performance, there are still some challenges and shortcomings.

1、 The standard for generating image quality assessments is inadequate. At this stage, the evaluation standard in the field of NST is greatly influenced by subjectivity, and a standard evaluation system is not formed, which is not scientific and standardized enough. You can specify an algorithm as a specification for comparison, and then, when selecting a quality evaluator, cover the population as generally as possible and develop a fixed and comprehensive evaluation form.

2、 The trade-off between speed, flexibility and quality. Image-optimized NST delivers superior performance in quality, but at a higher computational cost. Although PSPM can be stylized in real time, it requires a separate network of aesthetic Chinese types. MSPM increases flexibility by combining multiple styles into one model, but still needs to pre-train the network for a set of target styles. Although ASPM algorithms successfully transmit any style, they are not satisfactory in terms of perceived quality and speed.

3、 The limit of shape change. At this stage, most style migration is only for the texture of the image, color changes, but ignore the impact of its shape. But in a particular scene, people want to generate an image shape as close as the target image shape. For example, when transforming a real face into a comic face, not only does it require a change in style, but it also needs to be exaggerated like an anime character in terms of shape contours. Therefore, the combination of image geometry transformation and style migration is an important way for NST to develop further.

6. CONCLUSIONS

Over the past few years, NST has grown rapidly in the arts, academic research, and commercial applications. This paper first summarizes the style migration methods based on image iteration and build model iteration, then introduces the evaluation method and application scenario of the algorithm, and finally summarizes the existing problems and challenges in NST field. In short, image style migration based on deep learning not only promotes the development of computer field, but also receives wide attention in other fields, so the development of NST has important research significance and broad application scenarios.

REFERENCES

- [1]. Leon Gatys, et al. "A Neural Algorithm of Artistic Style." *Journal of Vision* 16.12(2016): doi:10.1167/16.12.326.
- [2]. Mordvintsev, Alexander, Christopher Olah, and Mike Tyka. "Inceptionism: Going deeper into neural networks." (2015).
- [3]. Berger, Guillaume, and Roland Memisevic. "Incorporating long-range consistency in cnn-based texture generation." *arXiv preprint arXiv:1606.01286* (2016).
- [4]. Risser, Eric, Pierre Wilmot, and Connelly Barnes. "Stable and controllable neural texture synthesis and style transfer using histogram losses." *arXiv preprint arXiv:1701.08893* (2017).
- [5]. Castillo, Carlos, et al. "Son of zorn's lemma: Targeted style transfer using instance-aware semantic segmentation." *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.
- [6]. Li, Yanghao, et al. "Demystifying neural style transfer." *arXiv preprint arXiv:1701.01036* (2017).
- [7]. Li, Chuan, and Michael Wand. "Combining markov random fields and convolutional neural networks for image synthesis." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [8]. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European conference on computer vision*. Springer, Cham, 2016.
- [9]. Ulyanov, Dmitry, et al. "Texture networks: Feed-forward synthesis of textures and stylized images." *ICML*. Vol. 1. No. 2. 2016.
- [10]. Li, Chuan, and Michael Wand. "Precomputed real-time texture synthesis with markovian generative adversarial networks." *European conference on computer vision*. Springer, Cham, 2016.