Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/47/20241242

Neural Style Transfer with automatic style weight searching

Xiang Zhou

College of Computer and Cyber Security, Fujian Normal University, Fuzhou, Fujian, 350117, China

121152021072@student.fjnu.edu.cn

Abstract. Neural Style Transfer (NST) has garnered significant attention in the field of computer vision. Previous research in this area has made important breakthroughs, creating various style transfer models and innovative architectural designs, and has achieved success in commercial applications. This paper presents a novel design for automatic style weight determination based on constraining the initial total loss within a healthy range and finding the optimal solution through grid search. This design aims to harness the potential of existing NST models by automating the optimization of neural style transfer hyperparameters. The paper first discusses the impact of content weights and loss weights on the generated images and validates the influence of weight ratios on image quality through experimental adjustments of weight proportions. It also confirms that the initial loss essentially defines the optimization space of the optimizer. The paper explores the significance of the initial loss and proposes a method to improve image generation quality by constraining the initial loss range.

Keywords: Neural Style Transfer, Deep Learning, Automatic Weight Searching.

1. Introduction

In recent years, with the continuous development of deep learning technologies, various subfields, such as computer vision and natural language processing (NLP), have achieved remarkable accomplishments. Computer vision has become a focal point not only in machine learning but also in the broader field of artificial intelligence. Computer vision is a powerful image representation technique that decomposes images into independent factors perceivable by human vision and makes them computationally interpretable [1]. It has successfully bridged the gap between computers and human vision and has found applications in autonomous driving, image generation, super-resolution, style transfer, and many other exciting domains. These innovations have made neural network scientists proud and continue to bring surprises and conveniences to people's lives.

Among these advancements, neural style transfer emerged several years ago and continues to hold significant research value today. Past work has focused on designing new models or loss functions to improve the quality and effectiveness of image generation. For example, previous research concentrated on enhancing the architectures of generators and discriminators and introducing new loss functions to better capture content and style information in images [2]. While these efforts have been successful to some extent, they often required extensive manual tuning of hyperparameters to achieve the best results.

Though there are fewer new models today that significantly boost performance, it does not mean that existing models no longer have their original potential and room for improvement. Most general style

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/47/20241242

transfer models that offer flexible tuning spaces provide users with numerous hyperparameters to adjust for optimal image generation. Hyperparameters play a crucial role in style transfer, but tuning them is not straightforward. Even with extensive adjustments of hyperparameters, there remains a significant risk of generating low-quality images.

Therefore, this paper first explores the significant impact of hyperparameter tuning on image generation in neural style transfer. Subsequently, it provides an automated solution for tuning hyperparameters for a specific model. The concept of automatically matching hyperparameters holds the promise of alleviating the burden of manual hyperparameter tuning, saving time for users and researchers, and improving image generation quality.

In this paper, the author employs the VGG19 recognition network to produce results. By comparing this network with state-of-the-art alternatives, it is evident that AlexNet and GoogLeNet yield similar results, while VGG19, which focuses on depth, is similar to VGG16. Due to the use of large kernels and a stride of 1 in all convolutional layers, VGG networks effectively retain information [3]. Therefore, based on the outstanding performance of the VGG19 network in style transfer, this project selects it as the model to explore automatic weight matching for enhancing image generation quality.

2. Method

2.1. Assumptions

The core task of neural style transfer is to combine the content of an input image with the features of a style image to generate a new image. The image preserving the objects and structures of the original content and presenting the artistic style of another image. This can be viewed as a domain adaptation problem, with the goal of aligning the feature distributions of the generated image with those of the style image [4].

For style transfer, it is challenging to define a perfect standard for measuring the aesthetics of images. However, it is certain that the model should not generate very poor-quality images. The model does not need to produce perfect images for every pair of inputs, but it should not generate severely subpar images. However, the goals of most common models are often perfection-oriented, disregarding the potential lower bounds, which leads to the frequent generation of low-quality images.

The generation of low-quality images may be due to the fact that the optimized loss remains substantial. Therefore, before the optimization process, reducing the total loss caused by the feature extractor may alleviate the optimizer's burden, indirectly reducing the post-optimization loss. Building on this, there is a more reasonable assumption that different optimizers each have a healthy and effective initial optimization space, and if an appropriate initial loss is provided for each optimizer, their abilities can be fully unleashed. Initial loss is closely related to hyperparameters, making it the key to automatically setting hyperparameters.

Therefore, two hypotheses are proposed. The first hypothesis is that reducing the initial total loss can alleviate the optimizer's burden. The second hypothesis is that using weights obtained by constraining the initial total loss range can improve the quality of images generated by the model.

2.2. Style Weight Searching

For this project, an efficient and versatile style transfer model is chosen to improve experimental efficiency and demonstrate significant changes in image quality after weight optimization. This work is based on VGG19, which has already been successfully used for image processing tasks, especially style transfer. The depth of deep neural networks influences their feature extraction and classification capabilities, and recognition performance significantly improves when the network depth reaches 16-19 layers [5]. VGG19 possesses excellent feature extraction capabilities, generality, and pretrained weights, making it an ideal choice for this project.

In the image loading function, the maximum size of images is constrained to be within 512 pixels. This decision aims to ensure that the processed images remain computationally manageable while preserving the quality of generated images. The Gram matrix method is employed to transform pixel

DOI: 10.54254/2755-2721/47/20241242

information from the original image into feature information present in the image. This step helps capture the style and texture features of the image.

Regarding optimizer selection, this project decided to use the Adam optimizer with a learning rate of 0.02. Adam optimizer has shown excellent performance in deep learning tasks, as it combines the features of momentum and adaptive learning rates, effectively converging to optimal solutions [6]. This choice helps ensure that the model can optimize the generation of images quickly and stably.

In terms of the number of optimization iterations for images, this project divided the training process into 10 stages, with each stage comprising 100 optimization steps, resulting in a total of 1000 training iterations. This division allows for better monitoring of training progress and gradual improvement in the quality of generated images. Additionally, to prevent the generation of images with excessive high-frequency errors, regularization techniques are applied to process the high-frequency components of images, typically referred to as the total variation loss. Finally, after all training is completed, an additional optimization step is performed to obtain the final generated results.

For data selection, this project used close-up images of various animals as content images to ensure clear experimental targets and a degree of generality. For style images, diverse color images are chosen that contained minimal inherent structure, with their main characteristic being color diversity. This selection allows us to intuitively observe the impact of parameter adjustments on images of different objects and colors.

2.3. Experimental Design

Based on the calculation methods of content loss and style loss functions, it is known that the size of style weights directly affects the losses they produce [7]. Therefore, setting weights essentially means defining the initial losses. To validate the first hypothesis, it is necessary to confirm that style weights have a significant impact on image generation. This can be achieved by controlling the size of content weights while adjusting the size of style weights to observe their influence on image quality.

This work first attempted three cases: two extreme cases and one normal case. The extreme cases involved setting style weights to be extremely small and extremely large. According to the nature of the function, when style weights are adjusted to be infinitesimally small, the initial total loss approaches zero. In this scenario, whether the optimizer becomes ineffective could be determined. In the first experiment, the authors selected two sets of images with three different initial loss ranges: 1e0 to 1e1, 1e7 to 1e8, and 1e9 to 1e10. If low initial loss results in low-styled images, it could be concluded that the size of style weights implies the intensity of style generation.

To validate the second hypothesis, more comprehensive experiments are conducted, including different datasets and loss ranges: 1e5 to 1e6, 1e6 to 1e7, 1e7 to 1e8, 1e8 to 1e9, and 1e9 to 1e10. By using methods such as grid search, it could be found suitable initial loss ranges for the model and applied them in the experiments [8].

If not providing an initial loss or providing a very large loss to the optimizer results in irresponsible outcomes, it could be proved that the necessity of constraining the initial total loss range. Without an initial total loss, each gradient descent step and space will shrink, making progress inconspicuous. By constraining the initial loss within an appropriate range, this work can not only avoids excessively large initial losses but also fully utilizes the optimizer's capabilities, enabling both the model and the images to enter a healthy generation space.

2.4. Evaluation Metrics

The evaluation of neural style transfer algorithms and generated images remains an open and crucial issue in this field. Generally, two main types of evaluation methods are available for the neural style transfer domain: qualitative evaluation and quantitative evaluation. Qualitative evaluation is more subjective and relies on the aesthetic judgment of observers, influenced by various factors. Quantitative evaluation focuses on finding precise evaluation metrics, including time complexity, loss changes, and others. In this project, the authors used qualitative evaluation to assess the quality of generated images and quantitative evaluation to assess the performance of the optimizer.

DOI: 10.54254/2755-2721/47/20241242

Regarding qualitative evaluation, this project heavily relies on the subjective aesthetic judgment of observers to assess the visual appeal of generated images. This involves manual observation of generated images and evaluation based on observer feedback. Although this evaluation method is subjective, it provides intuitive impressions of image quality.

In quantitative evaluation of the optimization process, this project focuses on monitoring loss function reductions and examining intermediate results of image generation. Specifically, the project monitors changes in losses during optimization to ensure they decrease within reasonable ranges. The project also observes intermediate results of images during the optimization process to assess the progressive improvement of generated images. These evaluation metrics help assess the performance of the optimizer and the model in generating images [9].

In summary, this project uses qualitative evaluation to assess the aesthetic appeal of generated images and quantitative evaluation to assess the performance of the optimizer.

3. Results

3.1. Experiment 1: Impact of Reducing Initial Total Loss on Images

In Experiment 1, two sets of images exhibited similar phenomena within three different initial loss ranges. Results are demonstrated in Figure 1. When the initial loss was small (1e0 ~ 1e1), there was almost no generation of style features, only a reduction in the resolution of the content image. In the normal initial loss range (1e7 ~ 1e8), there was a good image generation result with sufficient style intensity. In the high initial loss range (1e9, 1e10), image generation quality decreased, and the style intensity did not significantly increase. This validates that the generation of low-quality images is indeed due to the optimizer still having a large loss after optimization. Therefore, it can be considered that reducing the initial total loss within a certain range effectively improves image quality.



Figure 1. Results at various loss ranges (Figure Credits: Original).

It was also observed that, under the same style weights, different images had different initial losses. If fixed weights were used, random images had the probability of generating extremely high initial losses, resulting in poor-looking generated images. Random images also had the probability of generating lower initial losses, leading to insufficient style intensity in the generated images. All of these findings demonstrate that the model's lower bound is not guaranteed. In Experiment 2, this issue will be addressed.

DOI: 10.54254/2755-2721/47/20241242

3.2. Experiment 2: Impact of Constraining Initial Total Loss Range on Images

In Experiment 2, four sets of images were selected with commonly used ranges (1e5 ~ 1e6, 1e6 ~ 1e7, 1e7 ~ 1e8, 1e8 ~ 1e9, 1e9 ~ 1e10). Results are demonstrated in Figure 2, Figure 3, Figure 4, and Figure 5 respectively. The initial loss value of the latter image was approximately 9 to 11 times that of the previous one, around 10 times. After constraining the range, the specific initial loss values (e.g., different leading digits) were randomly determined.



Figure 2. Example 1 of NST result with various weights (Figure Credits: Original).



Figure 3. Example 2 of NST result with various weights (Figure Credits: Original).

DOI: 10.54254/2755-2721/47/20241242



Figure 4. Example 3 of NST result with various weights (Figure Credits: Original).

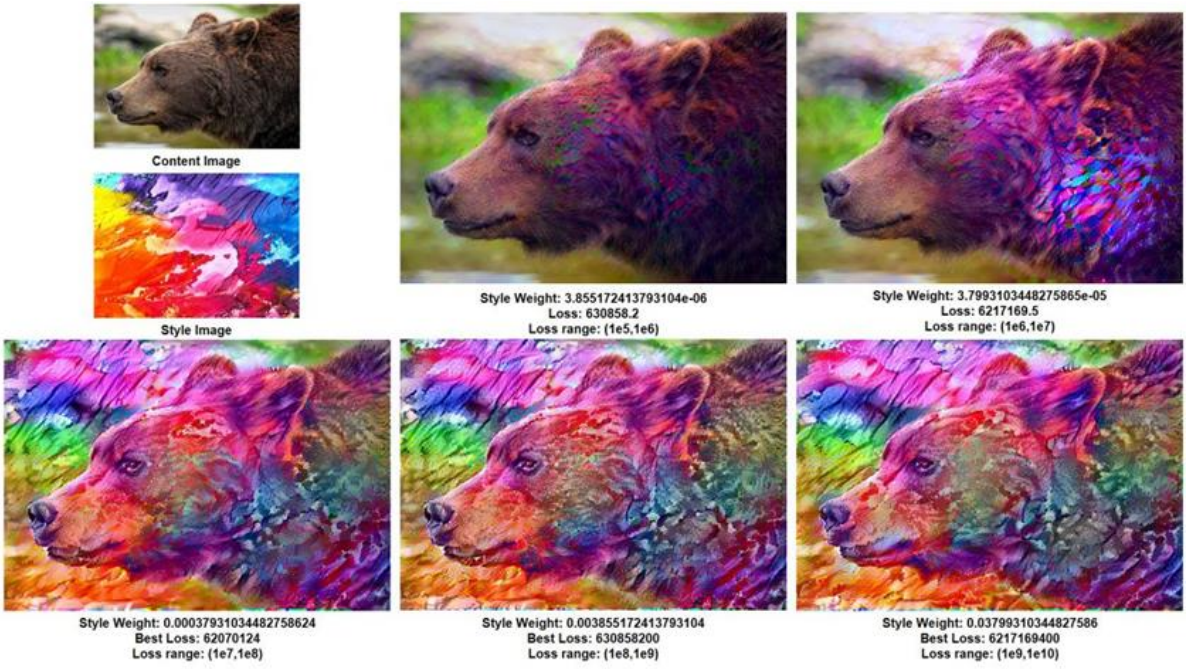


Figure 5. Example 4 of NST result with various weights (Figure Credits: Original).

The initial performance of image generation was characterized by a continuous increase in style intensity. When the intensity reached a certain point, it no longer increased, but the loss of the results increased, leading to a decrease in image quality. This implies the existence of an optimal initial loss value. Below this optimal value, the initial loss is positively correlated with style intensity, and above it, the initial loss is negatively correlated with image quality.

If the weights are determined based on this optimal value, then at this point, the style intensity of the images will be optimal, and image quality will also be optimal. Since different users use different images,

DOI: 10.54254/2755-2721/47/20241242

fixed weights are not feasible. Instead, they can be determined based on this optimal value and corresponding random images. By introducing this optimizer into style transfer without changing the model's original structure, the quality of style transfer images can be significantly improved.

Comparing the image quality within these five ranges, it can be observed that images generated when the initial loss is in the range of (1e7, 1e8) or (1e8, 1e9) exhibit usable results. Therefore, it could be concluded that the optimal initial loss value lies within this range. To gain a more intuitive understanding of gradient descent optimization, this work will monitor the changes in loss during the optimization process and examine the intermediate results of images during optimization.

3.3. Optimizer and Loss Function

To visually observe the gradient descent optimization when the initial loss is in the range of (1e7, 1e8), the project experimented with 30 intermediate optimization images and selected the most representative 12 samples. Each sample had an interval of 50 training steps, forming an approximate power function curve by plotting their mathematical relationship. The horizontal axis was measured in 50 training steps, and the vertical axis was measured in 200,000 losses.

At the beginning of optimization (0, 200), the loss decreased rapidly, and the quality and style of the image improved significantly. When the number of training steps was in the range of (200, 400), the rate of optimization decreased rapidly, but the image's quality continued to improve, and the style became richer. When the number of training steps was in the range of (400, 800), the loss decreased by only a few thousand in each optimization step, and the image did not change significantly. Finally, when the number of training steps approached 1000, the image reached around 1,500,000 losses. When the initial loss was in the range of (1e7, 1e8), the loss decreased by about 95% to 98%.

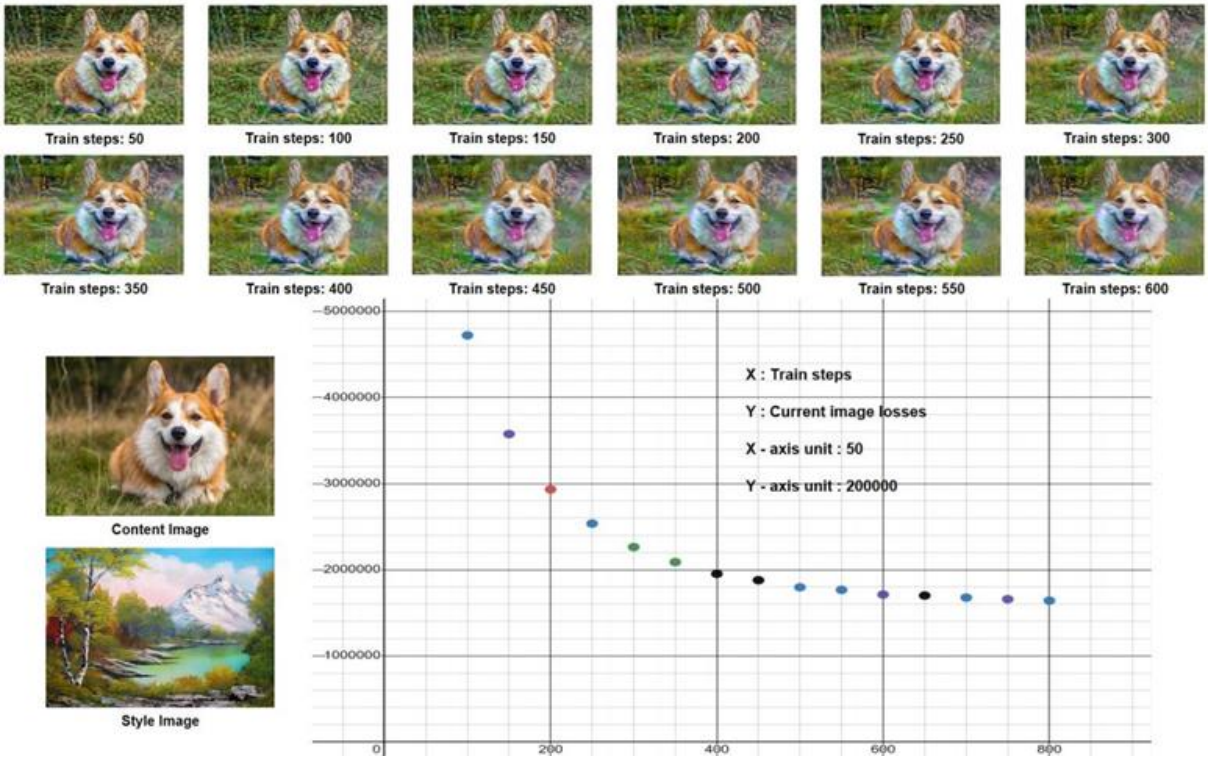


Figure 6. Intermediate result of NST (Figure Credits: Original).

4. Discussion

From the above, it can be visually observed that when the style weight is increased relative to the content weight, the style features of the image increase. Conversely, when the style weight is decreased relative

DOI: 10.54254/2755-2721/47/20241242

to the content weight, the content features of the image increase. Therefore, by controlling the weight ratios, the intensity of the generated image's style can be determined.

Different content and style images resulted in different initial total losses. This is because the feature extractor calculates different initial losses for different images. When excessively large initial total losses are generated due to image selection, it is beneficial to reduce the initial loss appropriately to increase image quality and reduce the risk of generating poor-quality images.

However, at the same time, it was found that excessively small initial losses result in meaningless results. This implies that if the initial loss is continuously reduced blindly, the optimizer and gradient descent will have difficulty making progress, leading to the generation of shallow or even style-less images. Thus, the second hypothesis could be successfully demonstrated that the initial total loss should neither be too high nor too low. Therefore, the goal and direction of setting weights were obtained, which is to select an initial loss that is most suitable for the optimizer's operation. By using this initial loss to automatically set the weights, the optimizer's effectiveness can be maximized to a higher degree.

There is reason to believe that different optimizers may have an optimal initial loss value when dealing with different losses. The initial loss, as a space for optimizer optimization, may be more effective for some optimizers when the loss is large, while others may only be effective when the loss is small. Here, the initial loss was only restricted from having too large of a deviation, resulting in significant risk reduction in generating poor-quality images. Unfortunately, although the weights obtained through grid search were much better than manual testing one by one, grid search is time-consuming and has low search precision [10].

Future research can explore more efficient ways of weight search, such as random search and Bayesian optimization, to find higher-precision and better-fitting hyperparameters. There is a wide range of research related to hyperparameter optimization, and the viewpoints presented in this paper are highly applicable—in the context of ensuring style intensity, calculate and search for the corresponding weights based on the most suitable initial loss for the optimizer and set them as hyperparameters of the model.

All neural network scientists pursue the optimal answers. However, the author believe that there is a philosophical perspective that, many times, the path to finding the optimal solution always leads to the most suitable one.

5. Conclusion

this paper presented two hypotheses: 1. Reducing the initial total loss can alleviate the optimization pressure on the optimizer. 2. Confining the initial total loss range to obtain weights can improve the quality of generated model images. To validate the first hypothesis, the project first attempted three categories of style weights and observed that lowering the style weight under certain conditions significantly improved image quality. When the initial loss was in the single digits, generated images almost no longer contained style, confirming the clear correlation between style intensity and style weight and confirming the critical role of hyperparameters in neural style transfer.

To validate the second hypothesis, this project selected a wide range of initial losses and set weights through grid search. The experiment demonstrated that this method could effectively improve the quality of generated images while ensuring style intensity and reducing the risk of generating low-quality images. This provides a new approach to optimizing hyperparameters for neural style transfer models.

Then, this work compared images generated with different initial losses and observed that for a neural style transfer model based on VGG19, when the initial loss was in the range of 1e7 ~ 1e8, the style weights found through the search produced higher-quality images. Furthermore, this work observed the changes in the loss function and intermediate results of image generation when the initial loss was in the range of 1e7 ~ 1e8.

Through the experiments mentioned above, it could be concluded that the initial total loss is essentially the optimization space for the optimizer, and different optimizers have their optimal optimization space. When this optimization space is within a healthy range, the optimizer can achieve its best performance.

DOI: 10.54254/2755-2721/47/20241242

This paper presents an automated logic and method for adjusting the weights of neural style transfer models, which works effectively with simple search methods. This paper also opens up possibilities for automating hyperparameter tuning in other neural network models.

The author firmly believe that existing models still have untapped potential and room for improvement. Therefore, optimizing hyperparameters and developing new models are equally important. While the birth of a new model is exciting, hyperparameter tuning and model maintenance often require researchers even more time. Thus, the automated optimization of model hyperparameters is a crucial issue in the field of deep learning.

References

[1] Gatys, L. A., Ecker, A. S., Bethge, M., Hertzmann, A., & Shechtman, E. (2017). Controlling perceptual factors in neural style transfer. In Proceedings of the IEEE conference on computer vision and pattern recognition, 3985-3993.

[2] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. IEEE transactions on visualization and computer graphics, 26(11), 3365-3385.

[3] Nikulin, Y., & Novak, R. (2016). Exploring the neural algorithm of artistic style. arXiv preprint arXiv:1602.07188.

[4] Li, Y., Wang, N., Liu, J., & Hou, X. (2017). Demystifying neural style transfer. arXiv preprint arXiv:1701.01036.

[5] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[6] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[7] Nikulin, Y., & Novak, R. (2016). Exploring the neural algorithm of artistic style. arXiv preprint arXiv:1602.07188.

[8] LaValle, S. M., Branicky, M. S., & Lindemann, S. R. (2004). On the relationship between classical grid search and probabilistic roadmaps. The International Journal of Robotics Research, 23(7-8), 673-692.

[9] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In 14th European Conference Computer Vision, 2(14), 694-711.

[10] Liashchynskyi, P., & Liashchynskyi, P. (2019). Grid search, random search, genetic algorithm: a big comparison for NAS. arXiv preprint arXiv:1912.06059.