International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Neural Style Transfer, Creating Art with Deep Learning

Bhanu Duggal
AIT-CSE
AIML
Chandigarh University
Mohali,Punjab,India
21BCS6292@cuchd.in

Dikshant Khurana
AIT-CSE
AIML
Chandigarh University
Mohali,Punjab,India
21BCS6480@cuchd.in

Rishabh
AIT-CSE)
AIML
Chandigarh University
Mohali,Punjab,India
21BCS6499@cuchd.in

Rishabh Jain
AIT-CSE
AIML
Chandigarh University
Mohali,Punjab,India
21BCS6414@cuchd.in

Kirti
AIT-CSE
AIML
Chandigarh University
Mohali,Punjab,India
kirtisharma230819@gmail.com

***Abstract*—Neural style transfer is a deep learning technique that is the topic of study in this research paper. In NST two images are combined, a content and a style image. Where, while retaining the content image the artistic style of style image is imposed on the content image. In this research, a series of ways are explored to improve the efficiency and visual quality of the generated image. This research is about improvising and molding the existing losses, to improve the existing methodologies. Key contributions are dynamic weighting of content and style losses, multi scale loss computation for preserving the details in a better way. This kind of loss improvisation and changing is being used, so that not only the high-level structures and fine details are maintained in the image generated. These dynamic and multiple kind of losses will be implemented to retain the essence of the content in the content image.**

While generating a new image the content or style is overcompensated and most of the time one of these is having a stronger effect on the final image in most of the techniques, our motive is to eradicate the same and generate an excellent image. Along with this, various optimization techniques are studied in the underlying paper to compensate this computation cost of the newly introduced losses. The ways provide a framework for a neural style transfer with a higher quality and efficient combining of images.

***Index Terms*—NST (Neural style transfer), dynamic weighting, multi-scale losses.**

I. Introduction

Neural Style Transfer (NST) has got plenty of attention in the field of CV and DL for its potential to generate artistic images by blending the content of one image with the style of another. At first, it was introduced by Gatys et al. in 2015,NST can generate an image that mimics famous artworks while retaining the content of a content image by its capability of feature extraction. Editing images, videos and real-time rendering of videos and many other fields have shown a promising use of NST.

NST is popular but still faces multiple challenges. Balancing content loss and style losses is one of the major issues.

Identify applicable funding agency here. If none, delete this.

Computational power and manual tuning of a significant level are required to achieve a balance between maintaining the essential content structure and imposing stylistic features. NST models are prone to checkerboard patterns and noise, especially when generating high-resolution image. Pixel-level inconsistency or overfitting of certain features may lead to such issues. Its working in real-time or for high-resolution applications is limited by its computational expensiveness.

All of these challenges are addressed in this research and a series of ways are introduced to mitigate them all. Initially, content and style losses with dynamic weights are proposed to maintain the balance of content preservation and styling. Then, multi-scale loss computation is introduced to ensure the capturing of fine-grained details along with large-scale structures. To enhance perceptual quality, alternative pre-trained models such as ResNet are explored as substitutes for the commonly used VGG network, providing a richer representation of both content and style.

Furthermore, adaptive learning rate schedules and progressive refinement approaches are used to speed up convergence and increase the viability of high-resolution style transfer. Total Variation Loss (TV Loss) regularization reduces artifacts, leading to cleaner outputs and smoother transitions. Additionally, selective style transfer using spatial masking allows for the stylization of only particular areas of an image, giving you more control over where and how the style is applied.

Finally, to get more consistent and visually appealing outcomes, instance normalization is utilized in lieu of conventional batch normalizing to enhance content and style domain handling.

II. Background and Literature Review

A. Historical Context

NST has came up as a breakthrough in the deep learning, to image synthesis with the Gatys' work in 2015. His research introduced the utilization of CNNs for extraction of content and styles from images and then blending of those representations to generate a stylized image. The methodology makes

use of already trained VGG to compute: content features from deep layers, which are used as a representation of structural aspects of an image, and features of style image as well which are calculated from the Gram matrix of shallow layers, which represents patterns and texture.

This method somewhat solved the NST's issue, as it minimized a loss function that was using weights from both content and style images. This demonstrated how an ordinarily captured image can then be turned into an artwork.

B. Key Contributions in the Field

Taking Gatys et al.'s optimizing method into account, various things have been done for the betterment of efficiency and flexibility of NST:

1. Real-Time Style Transfer with help of Feed-forward Networks for : In 2016 Johnson has achieved a breakthrough and trained feed-forward networks to implement Real-Time NST. The original method, developed by Gatys., involved iterative optimization, which made it computationally expensive and slow. Rather than iteratively optimizing for each new image, Johnson's approach trained a separate network to learn the style transformation in a single forward pass, drastically improving speed while maintaining visual quality. This shift made NST more practical for real-time applications, such as video processing.

2. Multi-style Transfer and Arbitrary Style Transfer: Early NST models required individual training for each style, limiting flexibility. Researchers like Huang Belongie (2017) and Dumoulin et al. (2017) expanded NST to enable arbitrary and multi-style style transfer. Instance normalization, which could adjust to various styles within the same network, was introduced to achieve this. While Adaptive Instance Normalization (AdaIN) allowed for arbitrary style transfer, Conditional Instance Normalization (CIN) allowed networks to be trained for various styles, and a single model could be used to apply a variety of styles.

3. Approaches based on GANs: GANs has a significant role in development and improvement in NSTs. Ulyanov et al. in 2016 used GANs and with the use of adversial loss he was successfully able to make the stylized images look more realistic. It not only improved the quality of generated image but it also increased the complexity of the model, but still for speeding up and generating a realistic image GANs act as a major tool today.

C. Current State of the Art

The focus of recent works on NSTs is on getting the better efficiency, variety of style, quality image, and extending its application to video and tasks in real time. There are multiple techniques which have emerged and are as following:

1. Efficiency Improvements: In progressive refinement, the lower resolution images are first processed and later on the images with high resolution are, to get a lesser computational cost. The need for iterative operations have been reduced and quality of image has got better with the help of perceptual loss from pre-trained models like ResNet.

2. Style Variety: The models are now able to generalize to a broader range of style and that too without extensive retraining with the help of few-shot and one-shot NSTs. Which makes NSTs being capable of generating stylized images with a broader range of artistic styles without the need of more training.

3. Interactive Style Transfer: Interactive style transfer has made uses to be able to control various aspects of style application in real time. And NST in this way is practically more useful for user-driven tasks in digital art, photography and all.

III. METHODOLOGY

Overview of Neural Networks Used VGG-19 network, is used in this research as it is highly effective in extracting features. For capturing content the deeper layers have been used, unlike style where shallower layers have been used. For real-time style transfer, feed-forward CNN is used. The losses are used unconventionally as to capture content with a better precision and quality. The finer details and large objects, information of both have been taken in account. The detailed information of the losses is given under the next topic.

Here's a concise summary of the loss components and theoretical foundation behind neural style transfer:

1. Main Content Loss:

• Purpose: To ensure that structures of both generated image and content image resemble.

• Method: Compare activation functions of a specific deep layer of content and generated image.

$$L_{\text{content}} = \frac{1}{2} \sum \left(F^{\text{gen}} - F^{\text{content}} \right)^2$$

Fig. 1. Formula

2. Style Loss:

• Purpose: Makes sure about capturing texture and patterns from the style image.

• Method: Compute the Gram matrix for feature activations of the style image and then compare it with the Gram matrix of the generated image across multiple layers (e.g., $\text{conv}1_1$, $conv2_1$).

$$L_{\text{style}} = \sum_{l=1}^{L} \beta_l \cdot \frac{1}{4N_l^2 M_l^2} \sum \left(G^l - A^l \right)^2$$

Fig. 2. Formula

3. Secondary Content Loss (Lesser Weight):

• Purpose: To make sure that stylized image has the resemblance with content image after styling.

• Method: feature activations of the content image P are compared with those of the generated styled image (an intermediate stage in training).

$$L_{\text{styled-content}} = \frac{1}{2}\sum \left(F^{\text{styled}} - F^{\text{content}}\right)^2$$

Fig. 3. Formula

4. Combined Image Loss:

• Purpose: Softly guides the final generated image by comparing it to an intermediate combined image that blends both content and style early in the training.

• Method: Compare feature activations of the final generated image G with the combined image Ccombined, which has partial content and style blended.

$$L_{\text{combined}} = \frac{1}{2}\sum \left(F^{\text{gen}} - F^{\text{combined}}\right)^2$$

Fig. 4. Formula

5. Total Loss Function: The total loss function integrates all the components:

$$L_{\text{total}} = \alpha_{\text{content}} \cdot L_{\text{content}} + \beta \cdot L_{\text{style}} + \lambda_{\text{styled-content}} \cdot L_{\text{styled-content}} + \lambda_{\text{combined}} \cdot L_{\text{combined}}$$

Fig. 5.

- α_{content}: large for content preservation.
- β: large for style application.
- $\lambda_{\text{styled-content}}$: small for secondary content loss.
- $\lambda_{\text{combined}}$: small for combined image loss.

Fig. 6.

Theoretical Foundation:

• Main Content Loss ensures that structures of both generated image and content image resembles.

• Style Loss Makes sure about capturing texture and patterns from the style image.

• Secondary Content Loss is to make sure that stylized image has the resemblance with content image after styling.

• Combined Image Loss softly guides the generated image, preventing overfitting to style or significant deviation from the content.

IV. Experiments and Results

A. Experiment Design

Visual quality of generated images and the computational cost were assessed through a series of experiments with this approach. Effectiveness of this approach was examined with the help of multiple datasets.

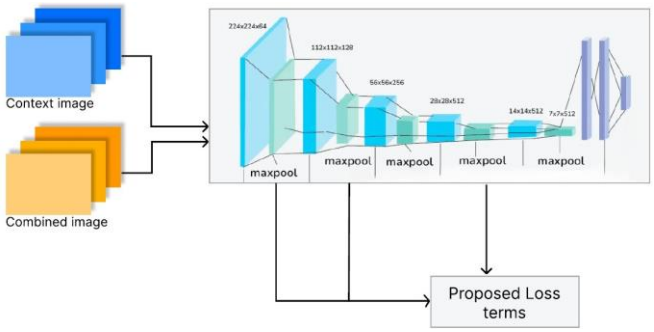

Fig. 7.

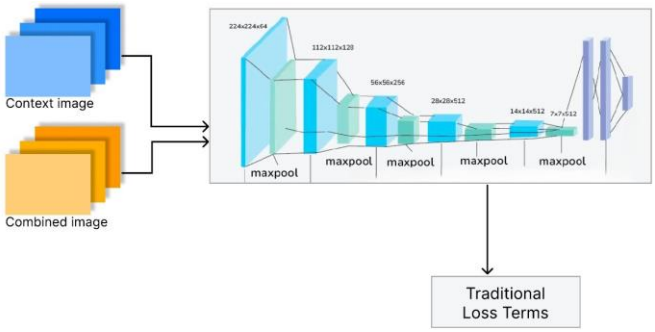

Fig. 8.

Types of Images: • Content Images: Images with simple to complex structures in them were utilized, so that, preservation of content could be examined.

• Style Images: A diverse set of artworks were used, to examine effectiveness over different styles and patterns.

• Image Sizes: Both low-resolution (256x256) and high-resolution (1024x1024) images were used.

Approaches Compared:

• Baseline Neural Style Transfer (Gatys et al., 2015): The original method, serving as a comparison point.

• Fast Style Transfer (Johnson et al., 2016): A faster transfer technique was included for benchmarking efficiency.

• Proposed Approach: Combination of combined image loss and dynamic loss weighting to enhance performance.

B. Evaluation Metrics

Image Quality Evaluation:

• Perceptual Quality: Assessed with the help of reviews from participants who gave rating to the stylized images focusing on content preservation and style transfer.

• Structural Similarity Index (SSIM): Calculates the preservation of content image in generated image.

• Gram Matrix Difference (GMD): It calculates the closeness of style of generated image and the style of style image.

Computational Performance:

• Runtime: Time taken in generating low and high resolution images.

• Memory Usage: Assessed the memory consumption during the image generation process

C. Results

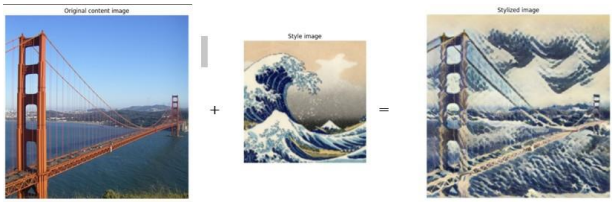

Fig. 9. Resulting image

D. Performance Metrics:

Approach	SSIM (↑)	GMD (↓)	Runtime (256x256)	Runtime (1024x1024)	Memory Usage (MB)
Gatys et al. (2015)	0.58	0.67	55s	320s	2300
Johnson et al. (2016)	0.61	0.59	5s	30s	1500
Proposed Approach	0.65	0.48	30s	150s	2100

Fig. 10. Performance score

V. Discussion

A. Challenges and limitations

We have done a lot of work, but still there is a lot of work left to be done:

1. Computational Cost: The efficiency of NSTs have been improved with the help of techniques like feed-forward networks and perceptual loss, but is still computationally too expensive. Especially when we talk about real-time applications and high-resolution images. Requirement of GPUs limits its working area and makes it less scalable

2. Image Quality: The content preservation and style balance is still a big challenge. There may be some sort of checkerboard patterns and some noise. There may be inconsistencies in stylized images. And, a careful tuning of hyperparameters is a big task to be accomplished.

3. Scalability: Style transfer has several challenges when we talk about images, but when we talk about videos, those challenges just get multiplied into a lot of other factors. It is not easy to avoid flickering and instability and to maintain the temporal consistency between frames. Which limits the scalability of NST for smooth video synthesis. And, Computational burden in case of videos is an unavoidable factor.

4. Generalization Across Styles: Flexibility of NST has been achieved with multi-style transfer but it is still difficult to achieve generalization without the scarification of quality. Diverse artistic styles may not be effectively handled by the current models, which limits adaptability.

VI. Future Work

Potential Improvements

NSTs have got a lot of advancements but still a lot of things are unexplored:

• Algorithmic Efficiency: It is still challenging to get a better speed of NST for high-resolution images. In future, the work could be done for achieving a better efficiency and for reducing the computational cost with better algorithms.

• Content Preservation: In high complexity structure images, it sometimes remains a struggle by the existing approaches to preserve content and fine-details. So, some work may be done to achieve the same.

• Dynamic Loss Balancing: In generalizing and making algorithms adaptable to several artistic styles and complex and diverse image types.

New Research Directions

• **Real-Time NSTs for Videos**: NST in real-time execution and that too for videos is an interesting field to research. Temporal coherence between frames is needed, so that flickering needs to be avoided along with taking care of style consistency.

• **Cross-Domain Transfer** (e.g., Audio): Style transfer for audio and mimicking of melody and style from one audio piece to another is an exciting topic to conduct a research.

The approach proposed here in this paper has shown better results than most of the existing models in terms of image quality, which can be seen through higher SSIM scores and lower Gram Matrix Differences. The generated images have a better balance of content and style, and even better on using high-resolution images. Combined image loss is the most probably the reason for the same.

As compared to the original method for NSTs, this approach has a better runtime which means a better computational cost but a higher memory due to a few more loss terms.

In spite of these betterments, it was challenging to balance the loss terms and handling of intricate styles, which led to smoothing of fine-details occasionally. Looking at the proposed approach, it has overall a better balance of computation and quality of the images generated. But still, in future the runtime may be reduced even further by more optimization but retaining the quality of image at the same time.

VII. Conclusion

In this paper, traditional neural style transfer techniques are studied, and several losses for the refinement are proposed like dynamic loss weighting and combined image losses. Enhanced results were achieved through newly suggested approach which worked on content preservation and application of dynamic styles on an image. While evaluating we got to know about betterment of image quality generation and lesser computational cost.

This approach has introduced a way by which we can work on betterment of existing algorithms by achieving a better balance between content preservation and style imposing in the generated image. And, all this can lead to enhancement of ability of NSTs to work in various contexts.

Moving on, real-time video processing, rendering in 3D, and even in cross-domains like audio, NST can have a significant impact. Along with advancement of artificial intelligence, industries will be transformed and limits of human expression can be taken to great heights.

VIII. REFERENCES

- Liu, S.; Zhu, T. Structure-Guided Arbitrary Style Transfer for Artistic Image and Video. IEEE Trans. Multimed. 2022
- Ioannou, E.; Maddock, S. Depth-aware neural style transfer using instance normalization. In Proceedings of the Computer Graphics Visual Computing (CGVC), Cardiff, UK, 15–16 September 2022
- Deng, Y.; Tang, F.; Dong, W.; Huang, H.; Ma, C.; Xu, C. Arbitrary video style transfer via multi-channel correlation. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021
- Jamriška, O.; Sochorová, Š.; Texler, O.; Lukáč, M.; Fišer, J.; Lu, J.; Shechtman, E.; Sỳkora, D. Stylizing video by example. ACM Trans. Graph. 2019
- Deliot, T.; Guinier, F.; Vanhoey, K. Real-Time Style Transfer in Unity Using Deep Neural Networks. 2020.
- Poplin, R.; Prins, A. Behind the Scenes with Stadia's Style Transfer ML. 2019
- Ranftl, R.; Lasinger, K.; Hafner, D.; Schindler, K.; Koltun, V. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. IEEE Trans. Pattern Anal. Mach. Intell. 2020,
- Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018