

Enhancing Neural Style To Multi-Style and Semantic Preservation

Shikha Dwivedi

Department of Computer Engineering
JSPM's Jaywantrao Sawant College
of Engineering, Hadapsar, Pune

Om Shinde

Department of Computer Engineering
JSPM,s Jaywantrao Sawant College of
Engineering, Hadapsar, Pune.

Gaurav Sunil Taralkar

Department of Computer Engineering
JSPM's Jaywantrao Sawant College of
Engineering, Hadapsar, Pune

Rohit Thube

Department of Computer Engineering
JSPM's Jaywantrao Sawant College of
Engineering, Hadapsar, Pune

Tanmay Sinare

Department of Computer Engineering
JSPM's Jayw Sawant College of
Engineering, Hadapsar, Pune

ABSTRACT

Neural Style Transfer (NST) is a deep learning technique that combines the content of one image with the artistic style of another image to generate visually appealing results. Traditional NST methods primarily focus on single style transfer and often fail to preserve important semantic information, resulting in distortion of object structures and image details. This paper presents an enhanced Neural Style Transfer framework that supports multi style blending while maintaining semantic consistency. The proposed approach integrates semantic segmentation and attention mechanisms to apply different artistic styles to semantically relevant regions of an image. A multi style fusion module is used to combine multiple style representations in a balanced and coherent manner. The framework utilizes a pre trained VGG19 network for feature extraction and optimization of content and style features. Experimental results demonstrate that the proposed system produces high quality stylized images with improved content preservation, smoother style transitions, and better semantic integrity. The proposed method offers a flexible and effective solution for artistic image generation and can be applied in digital art, image editing, multimedia design, and creative content generation.

INTRODUCTION

Neural Style Transfer (NST) is a deep learning technique that combines the content of one image with the artistic style of another image to generate visually appealing results. The concept was first introduced by Gatys et al. using deep convolutional neural networks to extract content and style representations from images [1]. Although traditional NST methods produce impressive stylized images, they are generally limited to single-style transfer and often fail to preserve important semantic information. Recent studies have focused on semantic-aware style transfer and multi-style blending to improve content preservation and visual consistency [3], [13]. To address these limitations, this work proposes an enhanced neural style transfer framework that integrates multi-style fusion and semantic preservation. By utilizing semantic segmentation and attention mechanisms, the proposed system applies multiple artistic styles effectively while maintaining the structural integrity of objects and improving the overall visual quality of the generated images [3].

Table1 Traditional Techniques Used for Neural Style Transfer

TECHNIQUE	APPLICATION IN NEURAL STYLE TRANSFER	LIMITATIONS
BACKGROUND SUBTRACTION	PRE-PROCESSING STEP TO ISOLATE FOREGROUND CONTENT FOR STYLE APPLICATION	INEFFECTIVE WITH COMPLEX OR DYNAMIC BACKGROUNDS; SENSITIVE TO LIGHTING VARIATIONS AND SHADOWS
OPTICAL FLOW	AS IN TEMPORAL CONSISTENCY FOR VIDEO STYLE TRANSFER	COMPUTATIONALLY EXPENSIVE; SENSITIVE TO NOISE AND RAPID OBJECT MOVEMENTS
KALMAN FILTER	PREDICTS OBJECT POSITION ACROSS FRAMES FOR CONSISTENT STYLE TRACKING	INEFFECTIVE FOR NON-LINEAR OR ABRUPT STYLE CHANGES; ASSUMES LINEAR MOTION AND GAUSSIAN NOISE
MEAN-SHIFT / CAMSHIFT	SUPPORTS REGION-BASED TRACKING FOR LOCALIZED STYLE APPLICATION	STRUGGLES WITH BACKGROUND COLOR SIMILARITY; PRONE TO FAILURE UNDER OCCLUSION
DEEP LEARNING (CNN-BASED)	EXTRACTS HIERARCHICAL FEATURES FOR DETAILED STYLE AND SEMANTIC UNDERSTANDING	REQUIRES LARGE DATASETS; COMPUTATIONALLY INTENSIVE; MAY NOT SUPPORT REAL-TIME PROCESSING
YOLO / SSD (REAL-TIME DETECTORS)	DETECTS SEMANTIC REGIONS FOR TARGETED MULTI-STYLE TRANSFER	LIMITED PRECISION AT OBJECT BOUNDARIES; FOCUSES ON BOUNDING BOXES RATHER THAN DETAILED CONTOURS
SIAMESE NETWORK TRACKERS	Matches semantic features across frames for style CONTINUITY	Performs well for short durations; less effective under long occlusions or APPEARANCE CHANGES
TRANSFORMER-BASED MODELS	ENSURES LONG-RANGE SEMANTIC CONSISTENCY AND MULTI-STYLE BLENDING	HIGH COMPUTATIONAL COST; REAL-TIME APPLICATION REMAINS CHALLENGING
MULTI-OBJECT TRACKING (MOT)	MANAGES SEMANTIC REGIONS WHEN MULTIPLE OBJECTS ARE PRESENT FOR CONSISTENT STYLE APPLICATION	IDENTITY SWITCHING IN DENSE SCENES; HIGHLY DEPENDENT ON INITIAL DETECTION ACCURACY

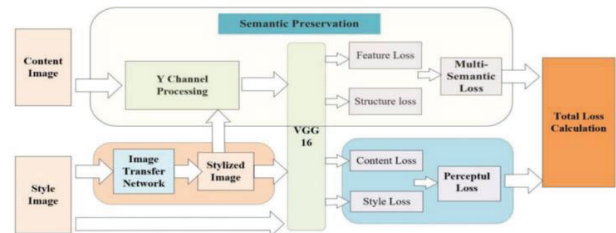
LITERATURE REVIEW

Several researchers have proposed different approaches to improve Neural Style Transfer (NST) by enhancing style quality and preserving content information. AdaAttN

introduced Adaptive Attention Normalization to establish fine grained correspondence between content and style features, resulting in improved stylization quality and content preservation [1]. StyTr² employed a transformer based architecture to capture long range dependencies and enhance semantic consistency during style transfer [2]. Style Mixer proposed a semantic aware multi style transfer framework that automatically applies different artistic styles to semantically meaningful regions, achieving better style diversity and smoother transitions [3].

Recent studies have focused on semantic preservation and attention mechanisms. MAST aligned content and style features based on semantic manifolds to improve structural fidelity [9]. CAST utilized contrastive learning to learn richer style representations and reduce visual artifacts [8]. Zhao et al. introduced semantic masks to guide the style transfer process, preserving semantic boundaries and reducing distortions [13]. Although these methods improve stylization quality, challenges such as efficient multi style fusion, semantic consistency, and real time performance still remain. Therefore, this work proposes an enhanced neural style transfer framework that integrates multi style blending and semantic preservation to generate visually coherent and semantically meaningful stylized images.

ARCHITECTURE DIAGRAM



1. Input Layer (Image Loading and Preprocessing)

The system takes a content image and a style image as input. Both images are resized, converted into tensors, and normalized to make them compatible with the VGG19 network.

2. Feature Extraction Layer

A pre-trained VGG19 model is used to extract content and style features from different convolutional layers. These features capture the structure and texture information of the images.

3. Style and Content Loss Computation Layer

Content loss measures how well the generated image preserves the original content, while style loss measures the similarity to the style image. Both losses are combined to calculate the total loss.

4. Optimization Layer

The generated image is optimized using the Adam optimizer. Pixel values are updated iteratively to minimize the total loss and produce the final stylized image.

TABLE2 HYPER PARAMETER SET

Hyperparameter	Table 2 - Hyper Parameter Set		
	Value/Range	Purpose	Consideration
Image Size	512 × 512 pixels	Controls input resolution and processing speed	Larger images increase computational cost
Learning Rate	0.01	Step size for Adam optimizer	Can be adjusted for faster or smoother convergence
Total Steps	300	Number of optimization iterations	Higher steps improve quality but slow process
Style Weight	1e5	Balances the influence of style loss	Higher values prioritize style over content
Content Weight	1	Balances the influence of content loss	Lower values preserve less content structure
Optimizer	Adam	Optimization algorithm	Suitable for direct image optimization
Style Layers	conv1_1, conv2_1, conv3_1, conv4_1, conv5_1	Capture style at different scales	Multi-scale style extraction
Content Layer	conv4_2	Captures detailed content structure	Best preserves object integrity
Normalization Mean	[0.485, 0.456, 0.406]	ImageNet mean for input normalization	Standard for pre-trained VGG models
Normalization Std	[0.229, 0.224, 0.225]	ImageNet std for input normalization	Matches the VGG model's training distribution

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide for their valuable guidance, continuous support, and encouragement throughout the development of this project titled **"Enhancing Neural Style to Multi-Style and Semantic Preservation."** Their expert advice, constructive suggestions, and technical insights helped us understand the concepts of Neural Style Transfer, semantic preservation, and multi-style blending in depth. We are also thankful to the Head of the Department and all faculty members of the Department of Computer Engineering for their constant motivation and academic support during the course of this work.

We extend our heartfelt appreciation to our institution for providing the necessary infrastructure, laboratory facilities, and resources required for the successful completion of this project. We also acknowledge the contributions of researchers whose work in the fields of Deep Learning, Computer Vision, Neural Style Transfer, and Semantic Segmentation served as valuable

references for our study. Finally, we thank our friends, classmates, and family members for their encouragement, cooperation, and support throughout the project, which greatly contributed to its successful completion.

CONCLUSION

The implemented Neural Style Transfer system effectively transfers artistic styles onto content images while preserving structural integrity. By using the VGG19 network as a fixed feature extractor, the framework accurately captures both content and style representations. The combination of content and style loss functions ensures balanced stylization. The iterative optimization process successfully refines the generated image over multiple steps. Hyperparameters such as style weight, content weight, and learning rate significantly influence the stylization quality. The system shows flexibility for various styles and can be adapted for real-time applications with proper tuning. Additionally, the approach maintains semantic details of objects during the style transfer process. The Gram matrix-based style representation provides reliable texture matching. The architecture can be further extended to multi-style blending and video style transfer. Overall, this work lays a strong foundation for future enhancements in semantic-aware and region-specific style transfer systems.

REFERENCES

- [1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2414–2423.
- [2] Chen, D., Liao, J., Yuan, L., Yu, N., & Hua, G. (2017). Coherent online video style transfer. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1105–1114.
- [3] Huang, H., Zhang, H., Zhang, Y., Li, P., & Lin, W. (2019). Style mixer: Semantic-aware multistyle transfer network. IEEE Transactions on Image Processing, 28(12), 5960–5971.
- [4] Sechenov, A., Chebotar, Y., & Lempitsky, V. (2021). Depth-aware neural style transfers for videos. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1660–1669.
- [5] Hua, Z., & Zhang, Y. (2021). Multi-attention network for arbitrary style transfer. International Conference on Neural Information Processing (ICONIP), 147–158.
- [6] Zhang, M., Yang, X., Xu, Y., Yan, L., & Gao, X. (2022). CAST: Contrastive arbitrary style transfer. ACM Transactions on Graphics (TOG), 41(4), 1–10.
- [7] Guo, X., & Hao, X. (2021). 3S-Net: Arbitrary semantic-aware style transfer with controllable ROI. 2021 IEEE International Conference on Image Processing (ICIP), 2783–2787.
- [8] Hu, R., Zhao, L., Zhou, T., Wu, D., & Zhang, L. (2022). Latent style: Multi-style image transfer via latent style coding and skip connections. Signal, Image and Video Processing, 16(6), 1467–1475.
- [9] Huo, Y., Gao, M., Li, W., & Qiao, Y. (2020). MAST: Manifold alignment for semantically aligned style transfer. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 4462–4471.
- [10] Zhao, S., Gallo, O., Frosio, I., & Kautz, J. (2017). Automatic semantic style transfer using deep convolutional neural networks and soft masks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4160–4168.
- [11] Huo, Y., Gao, M., Li, W., & Qiao, Y. (2020). MAST: Manifold alignment for semantically aligned style transfer. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 4462–4471.
- [12] Yao, Y., Hu, Y., Zhang, C., Wang, Z., Wang, B., & Fu, Y. (2019). Attention-aware multistroke style transfer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1467–1475.

- [13] Xiang, T., Feng, W., Liu, D., & Zhang, W. (2023). NCCNet: Arbitrary neural style transfer with multi-channel conversion. International Conference on Image and Graphics (ICIG), 140–151.