

*School of Computer Science Engineering and Information Systems*

*Department of Computer Applications*

*Fall Semester 2023 - 2024*

**SET Conference**

First Review (11th October, 2023)

Reg. No: 23MCA0131, 23MCA0133

Name: Srijan Dutta, Shaon Ghosh

Guide Name: Dr. R.K Nadesh

Paper Title: From Real to Artistic: Semantic Constraints in Style Transfer

***Problem Description:***

In the field of image processing, style transfer has become really popular. With the help of these techniques we can transform ordinary images into a mesmerizing piece of art while retaining both the features of the original image and the style of the artist or the era that is being copied. However, maintaining the semantic meaning and context of the original image is the key issue in this domain. This paper intends to provide a detailed analysis of three most popular semantic constraints and how they affect and help in the preservation of the meaning and context of objects and structures within the original image while applying a chosen artistic style.

***Literature Review:***

[1]In this method we have three input images: a content image (C), a style image (S), and an initial generated image (G). A pre-trained neural network like VGG-19 is used to dissect those images and extract their content object and style features. This algorithm calculates two distinct losses:

* Content Loss: Measures how different G is from C in terms of content.
* Style Loss: Measures the variance in artistic style between G and S.

To achieve a balance between preserving content and transferring style, weighted sum of style features and content object is used(known as total loss).Then optimization is done by adjusting pixel values in G to minimize total loss so that G can look like C in terms of content and more like S in terms of style. Finally, the output image is generated which miraculously balances the content structures in C and artistic qualities of S. The main advantage is the utilization of pre-trained deep neural networks minimizes the need for extensive training and users having precise control over the degree of stylization by adjusting the weights of content and style losses. Drawbacks include relying on the quality of selected image for the quality of output image and overfitting may cause a mimicry of the original image.

[2] In this paper a pre-trained CNN instead of the slow iterative optimization process is used. The network uses a perpetual loss function to measure the difference between the content and style elements. The perceptual loss between the generated output and a dataset of paired content and style images is minimized during the training phase. The feed forward method of analyzing the image through various layers helps it in operating in real time or near real time.

It is best suited for interactive applications like video games and other dynamic contents. It is not as accurate as the traditional style transfer methods, but it creates a fine balance between rapid processing and aesthetically appealing results. It has limited user control over stylization, and it fails to perform well on extreme styles or abstract artwork.

[3] In this paper, we present a simple adaptive instance normalization (AdaIN) layer that for the first time enables arbitrary style transfer in real-time. Instance Normalization is a type of normalization layer commonly used in deep neural networks. AdaIN normalizes each channel (feature map) of an image separately. It calculates the mean and standard deviation of each channel within an image and scales and shifts the values so that they have a standard distribution. Replacing it with correlation alignment or histogram matching could further improve quality by transferring higher-order statistics. We can further increase this domain in the direction of texture synthesis.

[4] This method differs from the feed forward approach in the fact that it gives high speed results while maintaining the quality of the image. Here the CNN architecture being used consist of both convolutional as well as deconvolutional layers for feature upscaling and sampling. A layer of instance normalization is added between the layers of convolution and linear activation functions. The loss function used for training combines content loss and style loss, similar to traditional neural style transfer. Optimization is done by minimizing the content and style loss. Instance Normalization plays a vital role in quickly analyzing and applying a specific artistic style.

[5] This study focuses on the application of a fundamental architecture to three distinct computer vision tasks, namely depth prediction, surface normal estimation, and semantic labeling. The proposed approach employs a multiscale convolutional network that can be conveniently adapted to each task through minor modifications, enabling direct regression from the input image to the output map. This method improves predictions by gradually using different scales and effectively captures many image details without using superpixels or low-level segmentation. The proposed approach achieves remarkable performance on benchmarks for all three tasks, as demonstrated by the experimental results.

[6] In the paper titled "Texture synthesis using convolutional neural networks" by Gatys, Ecker, and Bethge, the authors introduces an approach to texture synthesis utilizing deep convolutional neural networks (CNNs). It is built upon the idea that deep neural networks trained on image classification extracts hierarchical representations of visual textures. The authors propose a method to generate textures by iteratively modifying an initial random image to match the statistical properties of a target texture. They achieve this by minimizing the difference in features between the generated image and the required texture at multiple layers of the CNN. This approach presents remarkable results in bringing together diverse textures while preserving high-level structural information, opening up possibilities for various applications in computer vision, graphics, and texture generation. Gatys et al.'s paper has had a significant impact on the field of neural style transfer and texture synthesis, inspiring subsequent research in artistic style transfer, image generation, and deep learning-based texture manipulation.

[7] In the paper titled "A parametric texture model based on joint statistics of complex wavelet coefficients" by Portilla and Simoncelli , the authors present a pioneering method for modeling and synthesizing textures by analysing complex wavelet coefficients. They propose a statistical model that captures the relationships among these coefficients, enabling the generation of highly realistic textures with the same statistical properties as a given sample. This approach advances the field of texture analysis and synthesis by offering a principled framework for understanding and recreating complex textures, and it has found applications in computer vision, image processing, and texture-based image manipulation. Portilla and Simoncelli's work has served as a foundation for subsequent research in texture modeling and has contributed significantly to the development of algorithms and techniques for generating and analysing textures in various domains.

[8]In the paper "Learning transferable visual models from natural language supervision" by Radford et al., presented at the International Conference on Machine Learning in July 2021, the authors introduce an approach to training visual models using natural language supervision. They address the challenge of transferring knowledge from large text datasets to enhance the performance of computer vision models. This paper explores how pre-trained language models, such as GPT-3, can be leveraged to supervise the learning of visual models, ultimately achieving impressive results in various visual recognition tasks. This paper sheds light on the exciting potential of cross-modal learning and offers valuable insights into the intersection of natural language processing and computer vision, with implications for a wide range of applications in AI and machine learning.

[9]In the seminal paper "Generative adversarial nets" authored by Goodfellow et al. and published in Advances in Neural Information Processing Systems (NeurIPS) in 2014, the authors introduced the concept of Generative Adversarial Networks (GANs). GANs have since become a foundational framework in the field of deep learning and generative modeling. This groundbreaking work proposed a novel approach to generative modeling by training two neural networks, a generator and a discriminator, in a adversarial fashion. The generator aims to generate data that is indistinguishable from real data, while the discriminator aims to differentiate between real and generated data. The interplay between these networks leads to the generation of highly realistic and diverse samples, with applications spanning image synthesis, style transfer, data augmentation, and more. This paper marked a significant milestone in the development of generative models and continues to inspire innovations in the field of machine learning and artificial intelligence.

[10]In the paper "Learning to discover cross-domain relations with generative adversarial networks" authored by Kim et al. and presented at the International Conference on Machine Learning in July 2017, the authors introduce a novel application of Generative Adversarial Networks (GANs) for discovering cross-domain relations. This research addresses the challenge of learning meaningful relationships between data from different domains. By employing GANs, the proposed method learns to map data from one domain to another while preserving semantic consistency. The paper presents a valuable contribution to the field of domain adaptation and transfer learning, as it allows for the discovery of mappings between diverse data domains, with its application in image-to-image translation, style transfer, and more. This paper focuses on the versatility of GANs and their potential to facilitate knowledge transfer across domains, opening up new avenues for cross-domain data analysis and synthesis.

[11]In their paper titled "Unpaired image-to-image translation using cycle-consistent adversarial networks," authored by Zhu et al. and presented at the IEEE International Conference on Computer Vision, the authors propose a novel approach to image-to-image translation that does not require paired data. This work leverages Cycle-Consistent Adversarial Networks (CycleGANs) to learn mappings between two domains in an unsupervised manner. By enforcing cycle consistency, where translating an image from one domain to another and back should yield the original image, the model learns meaningful mappings between domains such as turning paintings into photos or horses into zebras. This technique has found wide applications in style transfer, domain adaptation, and image synthesis, showcasing the power of adversarial networks for unsupervised cross-domain tasks. The paper provides a significant contribution to the field of computer vision and image processing, offering a practical solution for unpaired image translation tasks.

[12]In their paper titled "Clipstyler: Image style transfer with a single text condition," authored by Kwon and Ye and presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition in 2022, the authors introduce an innovative method for image style transfer. The key novelty of this approach is the use of a single text condition to control the style of the generated image, leveraging the CLIP (Contrastive Language–Image Pre-training) model. By conditioning on text descriptions, this technique allows users to specify the desired style, making it highly flexible and interpretable. The paper's contributions lie at the intersection of natural language processing and computer vision, providing a novel and practical solution for generating stylized images guided by textual descriptions. This work holds great promise for applications in content creation, design, and artistic expression, as it simplifies and democratizes the process of image style manipulation.

[13] Motion style transfer is a technique used to create realistic motions in near real-time. The authors of this paper explore the ways in which motion style transfer can be achieved. They classify their approaches in two categories: Direct and Indirect. Direct approach directly learns the mapping between input and output motions, while indirect approach first learns the patterns of the input motion and then uses that representation to generate output motion based on the desired prompt. Direct approach is easier to train but can result in less accuracy whereas indirect approach is more accurate but takes time. The main challenges faced in this style transfer technique are:

* Preserving the naturalness and realism of the output
* Handling challenging motion styles
* Generalizing unseen motion data

[14] Neural Style Transfer (NST) is a deep learning technique that helps with the style transfer of one image to another. In this paper the authors provide details about the latest developments in the field of NST. They go through its history, applications, and how far NST has come. They also highlight methods that might help NST to be made more accessible to non-experts.

[15] In the paper, Wang et al. propose a new way of implementing NST which is way faster than the already existing algorithms. Their algorithm is based on a hierarchical deep convolutional neural network that learns to transfer artistic style at multiple scales. According to the authors, the previous methods fail to capture all the intricate details mainly because they work on a single scale and fail to consider the relationship between different scales. They address this problem in this new hierarchical NST algorithm where all features are first extracted in multiple scales and then fed to a CNN that learns to transfer the artistic style at every scale. The outputs of the CNN at every scale are then combined to produce the final stylized image.

[16] Data Augmentation is the process that artificially increases the amount of data by generating new data points from existing dataset. This is achieved by making minor changes to the data or by using machine learning techniques to generate new data points. In the paper Zheng et al. propose a new method for data augmentation using NST. They propose generating new data using NST that will be similar to original images in terms of context but have different styles. The authors evaluate their method on a variety of classification tasks, and they show a significant improvement in performance of machine learning models for small datasets. For example, on the CIFAR-10 dataset, STaDa improved the accuracy of the ResNet-18 model from 91.5% to 93.1% using only 10% of the original training dataset.

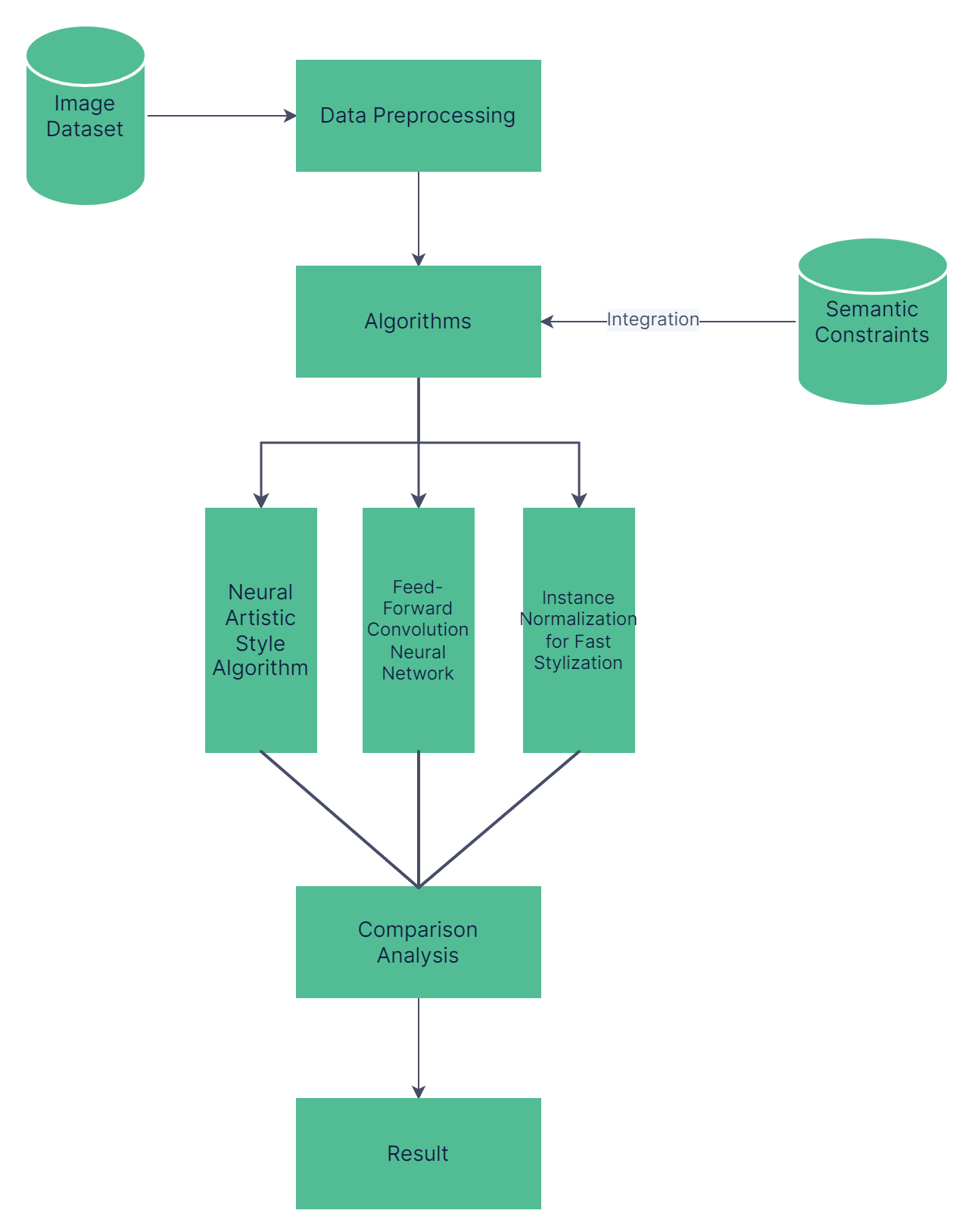
[17] In this paper Gatys et al. explores a deep-learning approach to photographic style transfer, a process that aims to transfer the style of a reference image onto another input image while maintaining photorealism The primary challenges addressed in this paper are structure preservation and semantic accuracy. The method involves constraining the transformation applied to the input image to be locally affine in color space to prevent distortions. The paper provides a detailed explanation of the method, including photorealism regularization and the integration of semantic segmentation into the style transfer process.

**Proposed Work**

The proposed project looks to conduct an extensive analysis of the three popular semantic constraints within the context of style transfer techniques. These semantic constraints will be examined in-depth to understand their impact on the preservation of the features and context of the original image and structures present within the original image while simultaneously applying a chosen artistic style.

***Detailed Design***

***Work Plan:***



1.

2.

Signature(Students) Signature(Guide)

***Reference:***

* [1] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," arXiv preprint arXiv:1508.06576, 2015.
* [2] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 694-711.
* [3] X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-Time With Adaptive Instance Normalization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1501-1510.
* [4] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," in Proceedings of the British Machine Vision Conference (BMVC), 2016.
* [5] D. Eigen and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2650-2658.
* [6] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in Advances in neural information processing systems, 2015.
* [7] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," International Journal of Computer Vision, vol. 40, no. 1, pp. 49-70, 2000.
* [8] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proceedings of the International Conference on Machine Learning (ICML), 2021, pp. 8748-8763.
* [9] I. Goodfellow et al., "Generative adversarial nets," in Advances in neural information processing systems, 2014.
* [10] T. Kim et al., "Learning to discover cross-domain relations with generative adversarial networks," in International Conference on Machine Learning (ICML), 2017, pp. 1857-1865.
* [11] J. Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2223-2232.
* [12] G. Kwon and J. C. Ye, "Clipstyler: Image style transfer with a single text condition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 18062-18071.

[13] S. M. A. Akber, S. N. Kazmi, S. M. Mohsin, and A. Szcz ˛esna, “Review Deep Learning-Based Motion Style Transfer Tools, Techniques and Future Challenges,” Sensors, vol. 23, no. 5, p. 2940, 2023.

[14] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite, and K. Kotecha, "Neural Style Transfer: A Critical Review," IEEE Access, vol. 10, pp. 9539183-9539213, 2022.

[15] X. Wang, G. Oxholm, D. Zhang, and Y.-F. Wang, "Multimodal Transfer: A Hierarchical Deep Convolutional Neural Network for Fast Artistic Style Transfer," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6337-6345, 2017.

[16]X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, and A. Smolic, "STaDA: Style Transfer as Data Augmentation," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9778-9787, 2020.

[17] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414-2422, 2016.