7th INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING & COMMUNICATION (ICICC-2024)

Organised by :

SHAHEED SUKHDEV COLLEGE OF BUSINESS STUDIES, UNIVERSITY OF NEW DELHI

IN ASSOCIATION WITH

NATIONAL INSTITUTE OF TECHNOLOGY PATNA & UNIVERSITY OF VALLADOID SPAIN

16-17th FEBRUARY 2024

1. TITLE OF THE PAPER – Review of Various Neural Style Transfer Methods: A Comparative Study

NAME ALL THE AUTHOR – Palak Singh
Ragini Rani
Kalash Jain
Akash Goel

2. INDEX:

3. ABSTRACT:

Neural Style Transfer (NST) is a computational technique in image processing that leverages convolutional neural networks (CNNs) to combine the content of a base image with the style of another, creating a visually enhanced output. By utilizing pre-trained models and deep learning algorithms, NST transforms ordinary images by extracting and synthesizing features from different artistic styles.

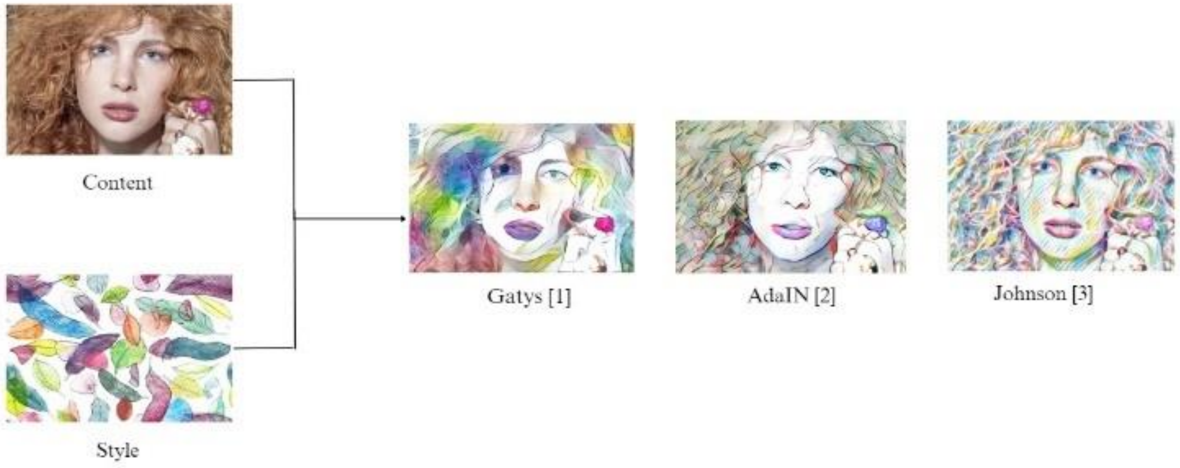
Gatys et al. pioneered neural style transfer with their "A Neural Algorithm of Artistic Style," introducing a method to blend the content of one image with the style of another. They devised a loss function encompassing both content and style, optimizing it to generate artistic images. Johnson et al. proposed improvements in "Perceptual Losses for Real-Time Style Transfer and Super-Resolution" by utilizing pretrained convolutional neural networks for faster computation. AdaIN (Adaptive Instance Normalization), introduced by Huang and Belongie in "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization," further enhanced the technique by enabling arbitrary style transfer in real-time through adaptive normalization of image activations. These advancements have significantly accelerated style transfer algorithms, facilitating the creation of visually striking images by merging content and style in various applications.

4. INTRODUCTION:

Neural Style Transfer, or NST, is like giving your photos a super cool makeover. It's like mixing different art styles with your pictures to make them look awesome and totally unique. It's like turning your regular photos into something really special. NST is like magic for pictures!.
Neural style transfer in Gatys et al. "A Neural Algorithm of Artistic Style." Their method involves defining a loss function that captures both the content and style of an image. The content loss measures the difference in content between the generated image and a content image, while the style loss measures the difference in style between the generated image and a style image. By minimizing these losses, the algorithm can generate an image that combines the content of one image with the style of another.
Johnson et al. proposed an improvement to neural style transfer in their paper "Perceptual Losses for Real-Time Style Transfer and Super-Resolution." They introduced the idea of using a pretrained convolutional neural network (CNN), such as VGG-19, to compute the content and style losses. This approach allows for faster training and inference compared to Gatys et al.'s method.

AdaIN is a technique introduced by Huang and Belongie in their paper "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization." AdaIN allows for arbitrary style transfer by adaptively normalizing the activations of one image to match the statistics (mean and standard deviation) of another image. This technique enables real-time style transfer and provides more flexibility in choosing style images. This approach is remarkably faster than the previous method by Gatys et al., with no loss of flexibility in transferring to new styles.

Overall, these approaches and techniques have significantly advanced the field of neural style transfer, making it possible to generate artistic images that combine the content of one image with the style of another in real-time or near-real-time.



5. LITERATURE REVIEW :

Sr. No	Reference	Author(s)	Year	Research Focus	Methodology	Key Findings
1.	Wei and Levoy, 2000	L.-Y. Wei, M. Levoy	2000	Fast Texture Synthesis using Tree Structured VQ	Tree Structured Vector Quantization (VQ)	Proposed a fast texture synthesis algorithm using tree structured VQ
2.	Mahendran and Vedaldi, 2014	A. Mahendran, A. Vedaldi	2014	Understanding Deep Image Representations by Inverting	Convolutional Neural Networks (CNNs)	Explored the inversion of deep image representations for understanding
3.	Simonyan and Zisserman, 2014	Karen Simonyan, Andrew Zisserman	2014	Very Deep Convolutional Networks for Image Recognition	Deep Convolutional Networks	Introduced a deep architecture for large-scale image recognition
4.	Yang et al., 2014	C.-Y. Yang, C. Ma, M.-H. Yang	2014	Single-Image Super-Resolution: A Benchmark	Super-Resolution Algorithms	Established a benchmark for single-image super-resolution algorithms
5.	Liu et al., 2015	F. Liu, C. Shen, G. Lin	2015	Deep Convolutional Neural Fields for Depth Estimation	Deep Convolutional Neural Networks (CNNs)	Proposed deep CNNs for depth estimation from a single image

Sr. No	Reference	Author(s)	Year	Research Focus	Methodology	Key Findings
6.	Gatys et al., 2016	L. A. Gatys, A. S. Ecker, M. Bethge	2016	Image Style Transfer using CNNs	Convolutional Neural Networks (CNNs)	Developed an approach for artistic style transfer using CNNs
7.	Johnson et al., 2016	J. Johnson, A. Alahi, L. Fei-Fei	2016	Perceptual Losses for Real-Time Style Transfer	Convolutional Neural Networks (CNNs)	Introduced perceptual loss for real-time style transfer and super-resolution
8.	Dosovitskiy and Brox, 2016	Alexey Dosovitskiy, Thomas Brox	2016	Generating Images with Perceptual Similarity Metrics	Deep Networks, Perceptual Metrics	Proposed a method for image generation based on perceptual similarity metrics
9.	Li and Wand, 2016	C. Li, M. Wand	2016	Precomputed Real-time Texture Synthesis with M-GANs	Markovian Generative Adversarial Networks (M-GANs)	Introduced precomputed real-time texture synthesis using M-GANs
10.	Huang and Belongie, 2017	X. Huang, S. Belongie	2017	Arbitrary Style Transfer in Real-Time	Adaptive Instance Normalization (AdaIN)	Achieved real-time arbitrary style transfer with adaptive normalization

Sr. No.	Reference	Author(s)	Year	Research Focus	Methodology	Key Findings
11.	Elad and Milanfar, 2017	Michael Elad, Peyman Milanfar	2017	Style Transfer via Texture Synthesis	Image Texture Synthesis	Explored style transfer through texture synthesis, achieving realistic results
12.	Hou et al., 2017	Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, P. Torr	2017	Deeply Supervised Salient Object Detection with Short Connections	Deep Learning, Short Connections	Proposed a deeply supervised approach for salient object detection with short connections
13.	Jing et al., 2020	Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song	2020	Neural Style Transfer: A Review	Neural Style Transfer Techniques	Presented a comprehensive review of neural style transfer methods
14.	Khatter et al., 2021	H. Khatter, S. Arif, U. Singh, S. Mathur, S. Jain	2021	Product Recommendation System for E-Commerce using Collaborative Filtering	Collaborative Filtering, Textual Clustering	Developed a recommendation system integrating collaborative filtering and textual clustering for E-Commerce
15.	Sharma et al., 2022	S. Sharma, S. Gupta, D. Gupta, S. Juneja, G. Singal, G. Dhiman, S. Kautish	2022	Recognition of Gurmukhi Handwritten City Names Using Deep Learning	Deep Learning, Cloud Computing	Implemented a system for recognizing handwritten Gurmukhi city names using deep learning and cloud computing

6. RESEARCH GAPS :

References	Research Gaps
Gatys et al., 2016	Limited exploration of general image style transfer beyond artistic applications.
Huang and Belongie, 2017	Need for addressing depth estimation challenges, especially in complex scenes.
Johnson et al., 2016	Research gap in assessing the generalizability and limitations of image generation based on perceptual similarity metrics.
Wei and Levoy, 2000	Exploration of salient object detection with short connections in complex visual environments.
Elad and Milanfar, 2017	Limited exploration of collaborative filtering and textual clustering integration for recommendation systems in diverse E-Commerce scenarios.
Simonyan and Zisserman, 2014	Lack of in-depth analysis on the scalability and efficiency of very deep convolutional networks.
Dosovitskiy and Brox, 2016	Research gap in standardized metrics for evaluating and comparing the performance of neural style transfer techniques.

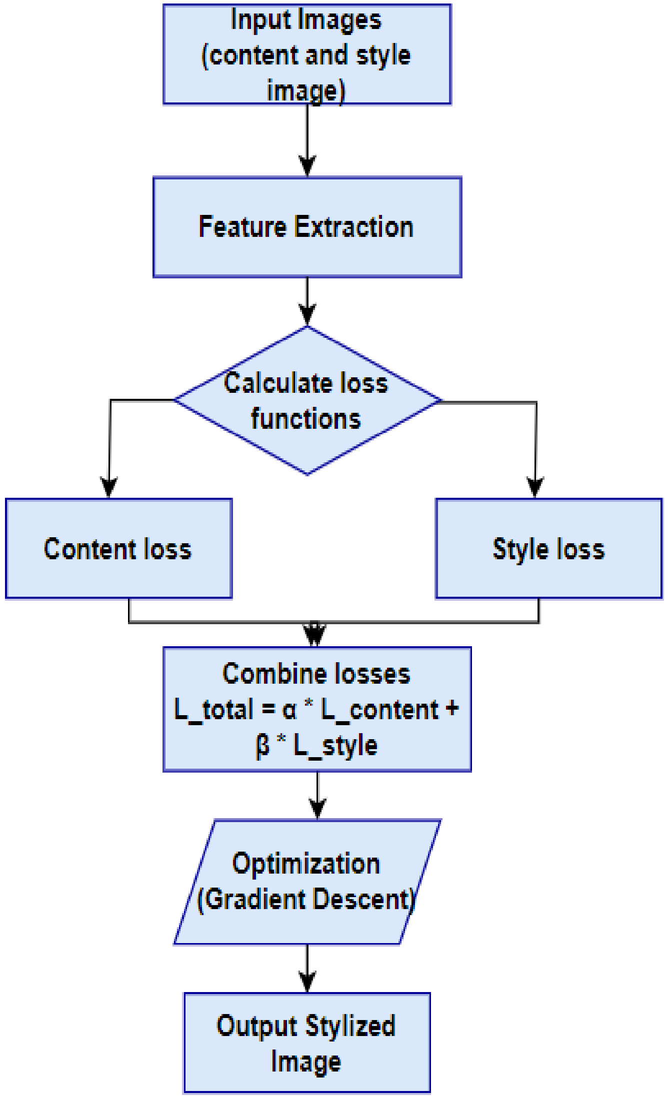
References	**Research Gaps**
Mahendran and Vedaldi, 2014	Limited understanding of the interpretability and practical implications of inverted features.
Liu et al., 2015	Research gap in addressing depth estimation challenges, especially in complex scenes.
Li and Wand, 2016	Limited exploration of practical applications and challenges in inverting visual representations with CNNs.
Yang et al., 2014	Need for a more comprehensive benchmark that covers diverse scenarios in single-image super-resolution.
Jing et al., 2020	Lack of standardized metrics for evaluating and comparing the performance of neural style transfer techniques.
Hou et al., 2017	Limited exploration of salient object detection with short connections in complex visual environments.
Mahendran and Vedaldi, 2014	Limited understanding of the interpretability and practical implications of inverted features.

7. PROPOSED METHODOLOGY :

- The Gatys Method utilizes VGG-Network with 16 convolutional and 5 pooling layers, employing average pooling for gradient flow. It assesses feature maps at each layer, calculating content loss via squared error, and style loss through statistical property differences in mean and variance.

- The AdaIN Method employs an encoder-decoder structure, utilizing AdaIN layers to align mean and variance of content and style feature maps. It employs pre-trained VGG-19 for content and style loss calculation, emphasizing mean and variance disparity.

- Johnson's Method involves an image transformation network (fv) and a loss network (ϕ) using pre-trained VGG with 16 layers. It optimizes via gradient descent, employing perceptual loss functions for content and style reconstruction, emphasizing feature maps and Gram matrices.



Flowchart of Gayts Method

Gayts Method:-

$$\mathbf{L}_{\text{total}}(\vec{b},\vec{q},\vec{z}) = \alpha\mathbf{L}_{\text{content}}(\vec{b},\vec{z}) + \beta\mathbf{L}_{\text{style}}(\vec{q},\vec{z})$$

This objective is seeking to minimize a total loss function $\mathbf{L}_{\text{total}}$ which combines two components: content loss($\mathbf{L}_{\text{content}}$) and style loss ($\mathbf{L}_{\text{style}}$).The coefficients α and β determine the relative importance of content and style in the final generated image.

Content Loss:

$$\mathbf{L}_{\text{content}}(\vec{b},\vec{z},l) = 1/2\sum_{m,n}\left(F^{l}_{mn} - P^{l}_{mn}\right)^2$$

The content loss measures the difference between the feature maps ($\mathbf{F}_{mn}$) of the original image ($\vec{b}$) and the generated image ($\vec{z}$). It uses the squared error between the two sets of feature maps.

Style Loss:

$$\mathbf{L}_{\text{style}}(\vec{q},\vec{z}) = \sum_{l=0}^{L} v_l S_l$$

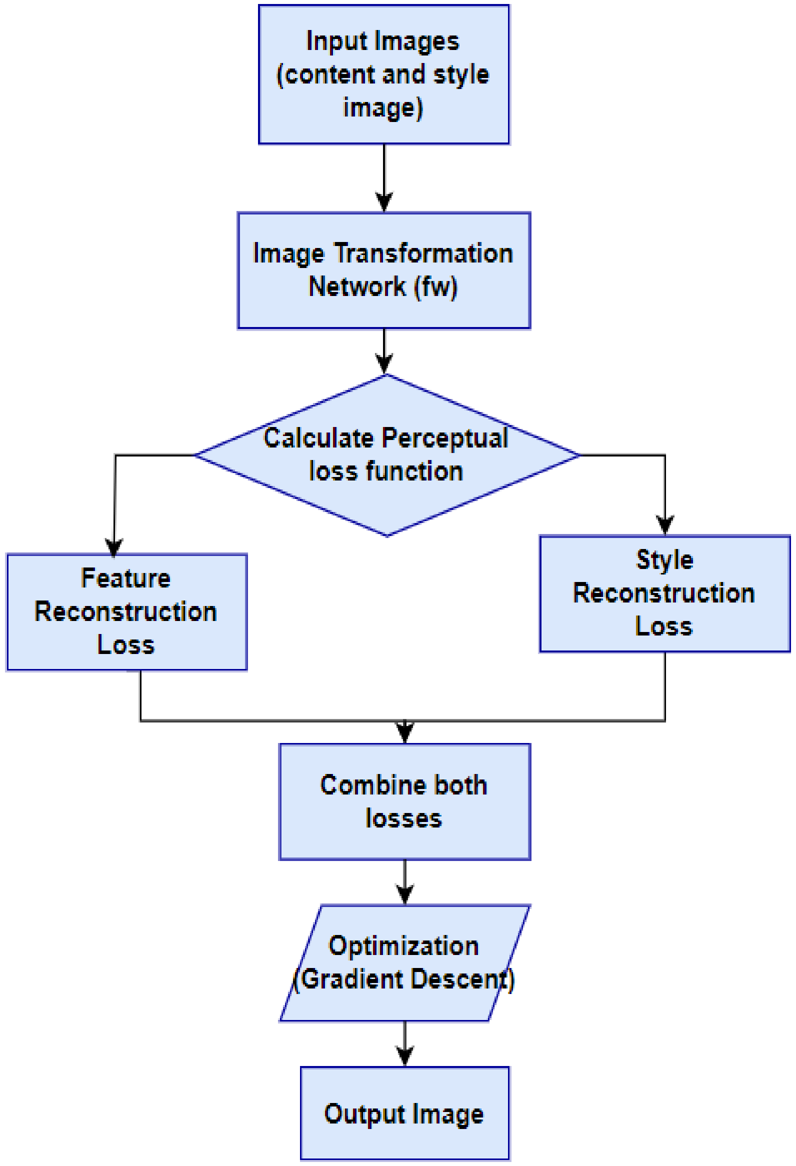
The style loss is the sum of style losses across multiple layers (S_l), weighted by coefficients v_l. Style loss is calculated based on the differences in Gram matrices (Gmn) between the style image ($\vec{q}$) and the generated image ($\vec{z}$).

Style Representation:

$$S_l = \frac{1}{4N_l^2 M_l^2} \sum_{m,n} (G_{mn}^l - B_{mn}^l)^2$$

Each style loss (Sl) for a specific layer (l) is computed as the mean squared difference between the Gram matrix of the generated image(Gmn) and the Gram matrix of the style image (Bmn).

The goal of optimization is to find the optimal values for the generated image (z) by adjusting its features to minimize the total loss. This involves iteratively updating the generated image using techniques like gradient descent, considering both content and style components with specified weighting coefficients α and β. The result is an image that combines the content of the original image with the artistic style of the reference image.



Flowchart of AdaIN method

AdaIN Method:

The AdaIN method facilitates neural style transfer by combining content and style from two input images to create a stylized output. It utilizes an encoder-decoder structure, where the encoder captures essential features, and the decoder generates the final stylized image.

Encoder:

The encoder is a segment of the pre-trained VGG19 model, extracting features from both content and style images. The AdaIN layer aligns the mean and standard deviation of content and style feature maps without introducing additional parameters.

AdaIN Layer Formula:

$$\mathbf{AdaIn(p,q)=\sigma(q)((p-\mu(p))/\sigma(p))+\mu(q)}$$

This formula adjusts the mean and standard deviation of the content feature map to match those of the style feature map

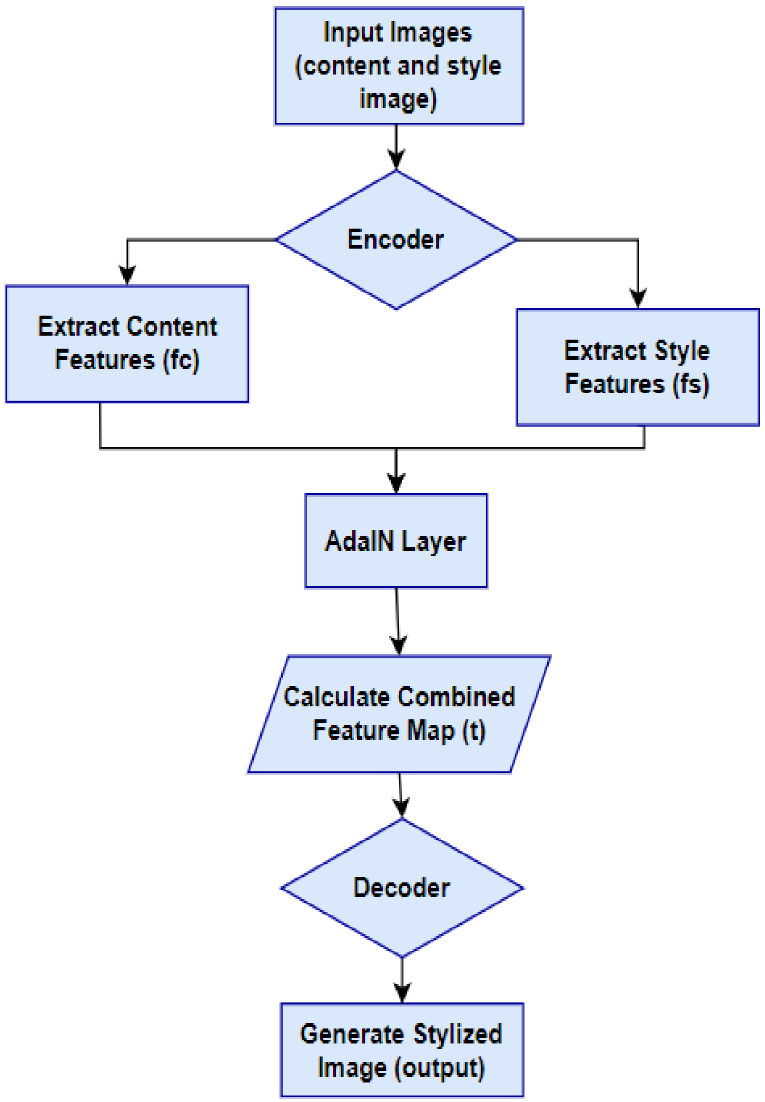
Decoder:

The decoder mirrors the encoder's architecture and is configured symmetrically. It avoids normalization layers within the decoder to maintain optimal performance.

Loss Functions:

The loss functions for neural style transfer involve content loss ($\boldsymbol{L_c}$) and style loss ($\boldsymbol{L_s}$), combined to form the total loss ($\boldsymbol{L_t}$) with a weighting parameter (**λ**).

$$\boldsymbol{L_t = L_c + \lambda L_s}$$



Flowchart of Johnson method

Johnson method:-

The Johnson method employs an image transformation network (fv) and a loss network (ϕ) for neural style transfer. The primary goal is to transform input images (p) into output images (q) by minimizing a combination of loss functions through gradient descent.

$$V^* = \arg min_v \, E_{p,\{qi\}}[\sum_{i=1} \lambda_i \, l_i(\mathrm{fv}(p), q_i)]$$

To enhance the capture of perceptual and semantic details, a pre-trained classification network (ϕ) is employed as a constant loss network, determining loss functions for the image transformation network.

Perceptual Loss Functions:

1- Feature Reconstruction Loss

Evaluates high-level differences by examining activations of the jth layer (ϕ_j) during processing of image p. It calculates the squared Euclidean distance normalized by the size of the feature map.

$$l_{feat}^{\phi,j}(\hat{q},\mathrm{q})= \frac{1}{C_J H_J V_J}||\phi_j(\hat{q})- \phi_j(\mathrm{q})||_2^2$$

2- Style Reconstruction Loss

Measures the difference between the Gram matrix of the generated image and the target image using the squared Frobenius norm.

$$l_{style}^{\phi,j}(\hat{q},\mathrm{q})= ||\ G_j^{\phi}(\hat{q})- G_j^{\phi}(\mathrm{q})\ ||_F^2$$

3- Pixel Loss

Measures dissimilarity between the output image $\hat{q}$ and the target image q using squared Euclidean distance normalized by their size C×H×V. Applicable when a ground-truth target q is available.

$$l_{pixel}(\hat{q},\mathrm{q}) = \frac{||\ \hat{\mathrm{q}},\mathrm{q}\ ||_2^2}{CHV}$$

The Johnson method combines these loss functions to optimize the image transformation network for effective neural style transfer. The approach considers both style and content targets, making it adaptable for various applications

8. RESULTS & DISCUSSION:

Approach	***Efficiency***	***Real-time Performance***	***Quality***
Gatys	*Moderate*(50-70%)	*No*	*High*(80-90%)
AdaIN	*High*(70-90%)	*Yes*	*Moderate*(70-80%)
Johnson	*High*(70-90%)	*Yes*	*Moderate*(60-70%)

Comparative Analysis of Neural Style Transfer Approaches

In our research, we compared three methods for changing the style of images: Gatys neural transfer, AdaIN (real-time arbitrary transfer), and Justin Johnson's Fast Neural Style Transfer. We looked at three important things: efficiency, how quickly they work in real-time, and the quality of the changed images.

Gatys neural transfer is great for making images look artistic but exhibits lower efficiency due to its optimization-based approach , because it needs a lot of computational resources. It gives beautiful results, but it may not be suitable for real-time performance applications.

AdaIN is very efficient especially for things that need to happen in real-time. It balances quality and speed, so it's good for practical uses where you want both.

Justin Johnson's method is also effective. It strikes a good balance between speed and producing high-quality images. While it may not achieve the same level of artistic results as Gatys' method, it is versatile and suitable for various tasks.

9. CONCLUSION & FUTURE WORK :

This paper explores the fascinating world of transforming images using neural networks, specifically focusing on a technique called neural style transfer. We looked at Gatys and his colleagues' pioneering work, which combines artistic styles with photos . While their approach is robust, it can be slow and time-consuming due to iterative optimization.

To overcome this challenge, we discussed the Adaptive Instance Normalization (AdaIN) method. AdaIN offers a faster alternative, allowing real-time style switching without the need for extensive training. It achieves this by efficiently combining content and style using version normalization. This makes AdaIN a convenient and effective tool for neural style transfer, giving users control in real-time.

Furthermore, we delved into the Johnson method, which employs perceptual loss functions and loss networks to measure differences in images . By using pre-trained image classification networks, this method improves the quality of image transformation and super-resolution tasks. The paper also explores image conversion networks, shedding light on their design principles and effectiveness.

FUTURE WORK

Potential future work in neural style transfer and image transformation includes improving the speed of methods like Gatys' neural transfer for real-time applications. Researchers could explore hybrid approaches, combining the artistic quality of Gatys' method with the speed of AdaIN. Another idea is to let people control the style of images while they're happening, like in live videos. Some researchers are looking into ways to make these methods work without needing lots of pictures for training.

Furthermore, user studies and evaluation metrics can provide insights into preferences and quality assessment. Finally, exploring dynamic style adaptation based on changing input conditions is a potential direction for future research. These efforts aim to enhance the flexibility, speed, and overall quality of neural style transfer methods.

10. REFERENCES :

[1] Gatys, L. A., Ecker, A. S., Bethge, M. (2016). "Image Style Transfer using CNNs." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[2] Huang, X., Belongie, S. (2017). "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization." In IEEE International Conference on Computer Vision (ICCV), 2017.
[3] Johnson, J., Alahi, A., Fei-Fei, L. (2016). "Perceptual Losses for Real-Time Style Transfer and Super-Resolution." In European Conference on Computer Vision, 2016.
[4] Wei, L.-Y., Levoy, M. (2000). "Fast Texture Synthesis using Tree Structured Vector Quantization." In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, 2000.
[5] Elad, M., Milanfar, P. (2017). "Style Transfer via Texture Synthesis." In IEEE Transactions on Image Processing, 2017.
[6] Simonyan, K., Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." In International Conference on Learning Representation, 2014.
[7] Dosovitskiy, A., Brox, T. (2016). "Generating Images with Perceptual Similarity Metrics." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[8] Khatter, H., Arif, S., Singh, U., Mathur, S., Jain, S. (2021). "Product Recommendation System for E-Commerce using Collaborative Filtering and Textual Clustering." In Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021.
[9] Sharma, S., Gupta, S., Gupta, D., Ju neja, S., Singal, G., Dhiman, G., Kautish, S. (2022). "Recognition of Gurmukhi Handwritten City Names Using Deep Learning." In Hindawi Limited, 2022.
[10] Mahendran, A., Vedaldi, A. (2014). "Understanding Deep Image Representations by Inverting." CVPR, 2014.
[11] Liu, F., Shen, C., Lin, G. (2015). "Deep Convolutional Neural Fields for Depth Estimation from a Single Image." In CVPR, 2015.
[12] Li, C., Wand, M. (2016). "Precomputed Real-time Texture Synthesis with Markovian Generative Adversarial Networks." In European Conference on Computer Vision, 2016.
[13] Yang, C.-Y., Ma, C., Yang, M.-H. (2014). "Single-Image Super-Resolution: A Benchmark." In European Conference on Computer Vision, 2014.
[14] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M. (2020). "Neural Style Transfer: A Review." In IEEE Transactions on Visualization and Computer Graphics, 26(11), 3365-3385, 2020.
[15] Hou, Q., Cheng, M.-M., Hu, X., Borji, A., Tu, Z., Torr, P. (2017). "Deeply Supervised Salient Object Detection with Short Connections." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.