CONTENT CREATION AND ARTISTIC STYLE INJECTION: REVIVAL OF KERALA MURAL ART THROUGH NEURAL NETWORKS

REVATHY PRASANNAN
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
revathymp96@gmail.com

AMRUTHA MURALIDHARAN
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
nair.amrutha3191@gmail.com

ACHAL AJITH K
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
achal.ajith@gmail.com

SABITHA M G
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
sabithamg@gmail.com

VIJI VARGHESE MANJALY
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
vijimanjaly@gmail.com

ALEENA KURIAN GEORGE
Department of CSE-AI
Adi Shankara Institute of Engineering and Technology
Kalady, Kerala
aleenakurian123@gmail.com

Abstract**—** Kerala mural art is a vibrant combination of bright colours and intricate brushstrokes. It was used as a way to depict the stories of Hindu mythology. As time has passed, the popularity of these paintings have diminished. This project aims to promote the artistic beauty and style of these paintings, through the utilisation of advancements in artificial intelligence, primarily neural networks. The process of artistic style transfer involves addressing the challenge of synthesising images, where the content of one image is replicated with the stylistic characteristics of another. This innovative art generation application revolutionises the realm of digital art, seamlessly converting photographs into captivating masterpieces inspired by iconic art styles like cubism, surrealism etc. Users can also effortlessly apply these styles to their photos, witnessing a remarkable transformation into digital canvases that echo the distinctive brushstrokes and colour palettes of legendary artists like Pablo Picasso, Vincent van Gogh, Salvador Dalí, and Claude Monet. The interactive and user-friendly interface ensures accessibility for both seasoned artists and casual enthusiasts. Prioritising high-resolution and visually stunning results, the system employs advanced neural network architectures to uphold the authenticity of chosen art styles while preserving the details of the original image. If the user has no image to work with, they are also allowed to leverage the freedom of generating brand new images through the use of a generative model that creates images based on user prompts. This tool could in turn be used to generate high quality, detailed images that match all the specifications mentioned in the prompt typed out by the user.

Keywords**—** Mural Art, Latent Diffusion Models, Neural Style Transfer, Convolutional Neural Networks, Generative AI

I. INTRODUCTION

Kerala mural art is a lovely mixture of vibrant colours namely, orange, yellow, red, and even green and black. These enormous frescoes were utilised as a means to depict stories of Hindu mythology such as the epics like Ramayana and Mahabharatha, or even scenes of deities like Vishnu, Krishna and Shiva. These lively murals date back to the 9th to 12th centuries.

These vibrant and eye-catching paintings are used as décor, but their uses go beyond simply bringing colour into a space. These paintings, which are a beautiful blend of natural pigments such as charcoal, ochre, and vegetable dyes, that go hand in hand with evergreen hues, are a window into the ancient art forms and religious beliefs centred around Kerala. Over the years, it has managed to take inspiration from other regional areas and even Buddhist cave paintings. Being mostly found in temples and palaces, these narratively effervescent paintings, comprising its scale, composition and colours are designed to complement the building's form, creating a unified work of art.

The motivation of the project is to promote mural art paintings and art styles which are popular only amongst people of particular regions. However, the process, technique and even the canvas have now changed. From the floor to the wall, onto thick cotton canvas, and now onto designer sarees: the journey of mural art has seen a shift from handcrafted beauty to low-quality digital prints, from natural to artificial colours, and from handmade brushes to synthetic ones. Somewhere along the way, the original essence and reverence for this art form seem to have been diminished. This project aims to reignite interest in mural art and preserve its old techniques, cultural heritage, and traditional values. To achieve this, we are combining image generation using AI with neural style transfer. Image generation entails using deep learning algorithms to translate textual descriptions into images. This process entails analysing the semantics of the text input, identifying key concepts and attributes, and constructing a visual representation that aligns with the provided description. By leveraging the power of AI, users can transform their imagination into tangible visual masterpieces, crafting images that embody their creative vision.

Neural style transfer on the other hand, mainly entails transferring the stylistic elements of a piece of artwork onto another image. This technique involves extracting the stylistic features of a reference image, such as brushstrokes, colour palettes, and textures, and applying them to a target image. Through the smooth integration of one image's content and another's style, users can explore a vast array of artistic possibilities, transforming their creations into captivating works of art.

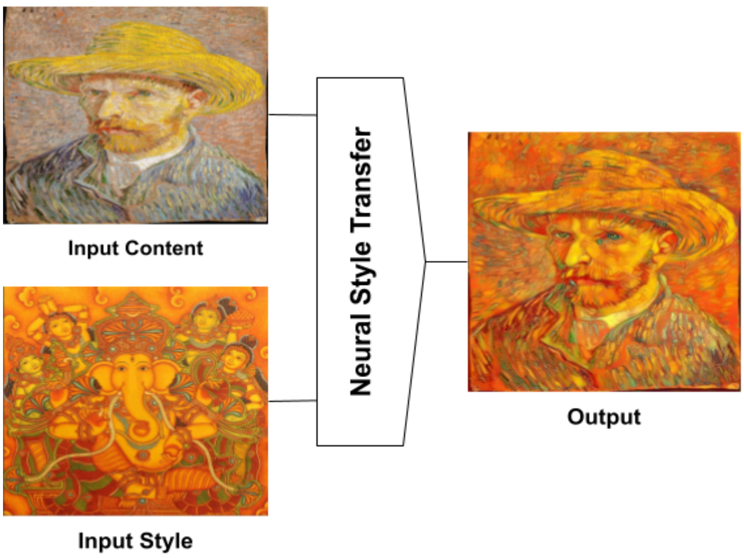


Fig. 1: Neural style transfer of a mural painting on a self portrait of Vincent Van Gogh

II. LITERATURE REVIEW

An approach named Hierarchical Style Transfer Network (HSTN) is presented by Sunder Ali[1] and is used for image stylization that allows users to regulate the level of style that is applied to a picture. A denoising CNN (DnCNN) and a fixpoint control loss are integrated by the HSTN, to preserve content details while enabling user control over stylization intensity. The network is composed of a loss network block, a DnCNN block, and an encoder-decoder block. Extensive experiments demonstrate HSTN's effectiveness in producing unique stylization results that maintain content image detail. The paper emphasises the network's flexibility in fusing multiple styles and creating new styles, highlighting its potential for user-controlled, detail-preserving image stylization. As for the limitations, the above mentioned algorithm fails to capture the intricate details within less-known and complicated paintings, such as mural art. This is because of its limited qualitative study. Moreover, it lacks user control with respect to its labelled regions.

The article[3] by Robin Rombach introduces Latent Diffusion Models (LDMs), a novel approach that utilises cross-attention layers to improve the computational efficiency and performance of diffusion models for synthesising high-resolution images. LDMs achieve very competitive outcomes in unconditional image production, text-to-image synthesis, and super-resolution, as well as in image inpainting and class-conditional picture synthesis. When comparing LDMs to pixel-based diffusion models, training and inference times are notably shortened. While LDMs struggle with its own limitations such as the challenge of generating high-resolution images from lower-resolution inputs, and might have limitations in precisely controlling specific aspects of the generated image, like pose, lighting, or background elements, Stable Diffusion helps tackle both of these issues by incorporating techniques to handle missing details and add new features during image generation, and offering more control over these aspects through its architecture or training data.

In response to the scarcity of existing embroidery designs or pre-drawn ethnic patterns, an enhanced deep learning model[2] by Yong Zheng proposes a novel method titled Fast Style Transfer for Ethnic Pattern Innovation (FST-EPI). FST-EPI empowers the creation of visually engaging elements for these areas within an artwork. FST-EPI consists of three main components: background generation, background fusion, and fast transfer, which together aim to preserve cultural elements while integrating unique styles. The focus of this reference paper being intricate local chinese art was taken as inspiration by our team for our paper which deals with and promotes mural art similarly, aiming to preserve and revive the untapped potential of Kerala mural art.

A novel approach to creating images from textual descriptions of the desired style is presented in the research article "DiffStyler: Controllable Dual Diffusion for Text-Driven Image Stylization"[4] by Nisha Huang. Text-driven stylization offers more intuitive control by allowing users to describe the desired style with text. However, bridging the gap between text descriptions and image outputs is difficult for traditional image processing methods. The method utilises learnable noise based on the content image. Text descriptions for less well-known art genres may be few or may not have the precise vocabulary needed to convey the subtleties of it. Because DiffStyler relies so heavily on text instructions, it might not be able to accurately depict the fine details and techniques that are frequently seen in mural art. There might be subjectivity in the way that artistic styles are seen and interpreted.

Xinying Han[5] makes use of DE-GAN to transform real images or photographs into artwork utilising the concept of style mitigation. The algorithm is used for its higher subjective image quality as well as its consistency in delivering images with the perceptual qualities of art. Due to exposure to lack of variety of art styles during the training phase, it struggles to capture the nuances of lesser known styles. This results in inaccurate generated images. It also focuses mainly on depth which is not the most crucial in mural art which depends mostly on its vibrant colours and patterns.

III. PROPOSED METHODOLOGY

A. Datasets

1) WikiArt: WikiArt It features artwork by 195 distinct artists. Paintings by 195 different painters are included in the folders that each provide information about the dataset used for each job (Style, Artist, and Genre classification). Over 50,000 photos make up the dataset. where the dataset information (style, artist, and genre classification) for each job is contained in a folder. Mainly there are wikiart dataset data fields as images (containing images) and labels (represent various emotions).This dataset has a training set composed of 81433 images. Dataset is available at https://paperswithcode.com/dataset/wikiart.

Fig. 2: Sample images from WikiArt

2) ImageNet: ImageNet is a dataset organised according to the WordNet hierarchy. It consists of over 14 million images of everyday objects , scenes organised into over 21,000 classes, 50,000 validation images and 100,000 test images. The models trained on this dataset are used for object detection and classification.The project has played a pivotal role in pushing forward research in computer vision and deep learning. Dataset is available at https://www.image-net.org/download.php.

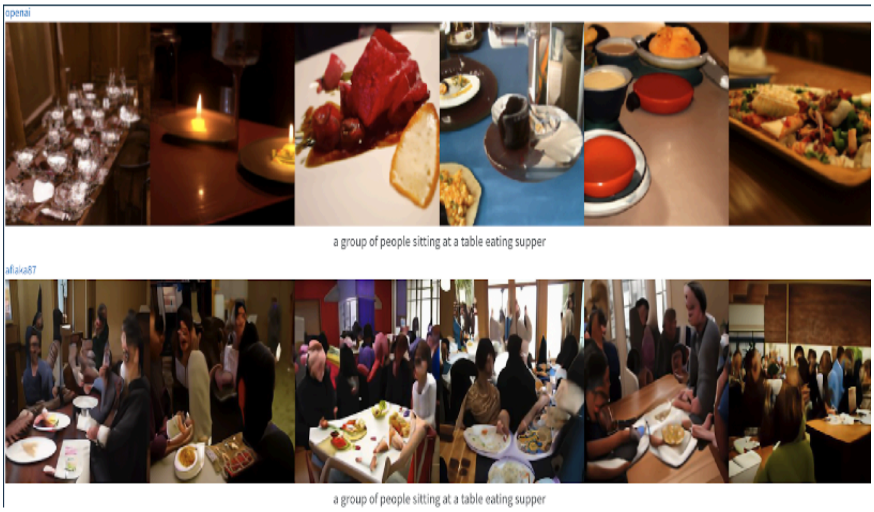


Fig. 3: Sample images from ImageNet

3) LAION-5B: One can use the extensive data of LAION-5B, which has 5,85B CLIP-filtered image-text pairs. 2,3B samples have English language, 2,2B samples have more than 100 different languages, and 1B sample has texts that cannot have a specific language assigned to them. The dataset can be found at https://laion.ai/blog/laion-5b/.

Fig. 4: Sample images generated by model trained on a subset of LAION-5B

B. Proposed Model

The primary objective of this project is to develop an image generator that responds to detailed user prompts, generating images tailored to the specified criteria. The user is allowed to either use the image generator tool or pick a content image of preference. In addition, the system offers a unique feature allowing users to select from a wide array of art styles, including both renowned and lesser-known styles. This feature aims to broaden users' artistic horizons beyond their current knowledge, fostering an appreciation for diverse art forms. The selected art style is applied to the generated image using the "Neural Style Transfer" technique, enabling users to visualise their input in different artistic styles.

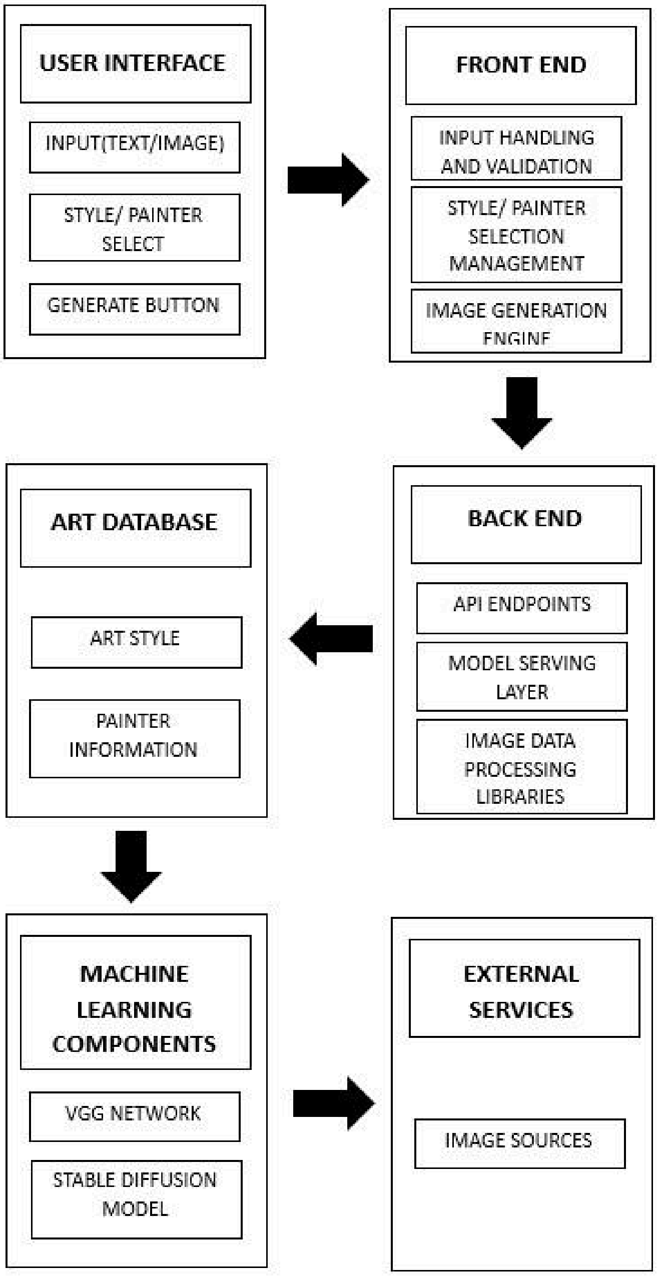


Fig. 5: Architecture Diagram

Our project, while drawing inspiration from and facilitating the transfer of diverse art styles from around the globe, centres primarily on the preservation and evolution of a time-honoured, traditional art form known as mural art. Characterised by its vivid palette, including hues of red, yellow, green, black, and white, mural art is traditionally created using a technique known as tempera, wherein these colours are intricately blended to narrate stories, often rooted in Hindu mythology. This art form is distinguished by its rich symbolism and abundant use of elaborate floral motifs that adorn its expansive canvases.

User Interface

In the UI for our model, the user must upload their own images or photos after clicking on the "Browse" button and once the image has been uploaded, the user moves on to select a style from a range of pre-existing styles which can further be detailed upon or their style of choice further amplified by selecting an image from the style chosen. The UI features a clean and innovative design, putting an emphasis on usability and simplicity so that users of all skill levels may utilise it.

Fig. 6: User Interface

Front End

The frontend consists of a user interface (UI) that allows users to upload images, choose a style for transfer, and adjust various parameters. Users can upload their images directly from their device using a file upload button or by dragging and dropping images onto the interface. The system supports various image formats such as JPEG, PNG, etc. Users can select a style for transfer from a list of predefined styles or upload their own style image. Once the user is satisfied, they can initiate the processing of the image. The frontend sends the image and style information to the backend for processing. After processing is complete, the frontend provides feedback to the user, indicating that the stylized image is ready. Users can then download the stylized image to their device or share it.The frontend includes error handling mechanisms to handle issues such as invalid inputs, failed uploads, or processing errors. It provides clear error messages to guide users on how to resolve these issues. The frontend is designed to be responsive, ensuring that it works well on a variety of devices and screen sizes, including desktops, tablets, and smartphones.Overall, the frontend of an image generation and style transfer system is designed to provide a seamless and enjoyable user experience, allowing users to easily create stylized images with just a few clicks.

Art Database

The art database provides a centralised and organised way to store and access information about artworks and artists, and the interaction between the frontend and backend allows users to easily search, browse, and explore art-related content. The interaction between the frontend and backend of an art database is mediated through an application programming interface (API). The frontend sends requests to the backend via the API, specifying the data or functionality it needs. The backend processes these requests, retrieves the requested data from the database, and sends it back to the frontend in a format that can be displayed to the user.

Backend

The backend contains the core algorithms for image generation and style transfer. These algorithms use deep learning techniques to analyse and transform images based on the selected style. It includes trained models for image generation and style transfer. These models have been trained on large datasets of images to learn the underlying patterns and styles. The trained models are deployed in the backend to process user inputs and generate stylized images. It also implements the logic for image generation and style transfer, which involves applying the style of a reference image to a target image. This process requires complex computations to adjust the features of the target image while preserving its content and natural language processing (NLP) techniques to convert textual descriptions into images. The backend manages the storage and retrieval of images, styles, and other related data. It may use a database or other storage systems to store and organise this data for efficient processing. It includes mechanisms for error handling and logging to track and manage errors that occur during image processing. This helps in identifying and resolving issues to ensure the system runs smoothly.

Machine Learning Components

Neural Style Transfer

The VGG-Network, a Convolutional Neural Network established and fully detailed in, provides the foundation for the results reported in this paper. Its performance on a standard visual object identification benchmark challenge is comparable to that of humans. We made use of the feature space that the 19 layer VGG-Network's 16 convolutional and 5 pooling layers provided. The fully connected layers of the VGG19 architecture were discarded for this application, since classification is not performed in this case.

The VGG19 model was trained on the ImageNet dataset, consisting of over 14 million images belonging to over 21,000 classes. With its extensive ImageNet training, VGG19 functions as a potent feature extractor. ImageNet's millions of annotated photos have "taught" VGG19 how to identify and distinguish between items in images.

Upon employing the trained VGG19 model, its performance revealed a deficiency in handling style transfer tasks for paintings boasting a plethora of vibrant hues and intricate patterns, notably exemplified in mural art. To rectify this shortfall, a decision was made to refine the model's capabilities. Through a process of fine-tuning, the model underwent training on an eclectic array of artistic genres, ranging from various paintings and sculptures to engravings, iconography, and, notably, mural art. Impressively, this refined model achieved an outstanding accuracy rate of 97.46% in the classification of these diverse art forms. This enhanced model demonstrated a remarkable ability to discern intricate details such as multiple colours, rich patterns, and brushstrokes within the style images, leading to significantly improved accuracy in style transfer tasks. The model's performance exemplifies the intricate dance between art and technology, where computational methods enhance our understanding and appreciation of artistic expression in all its diverse forms.

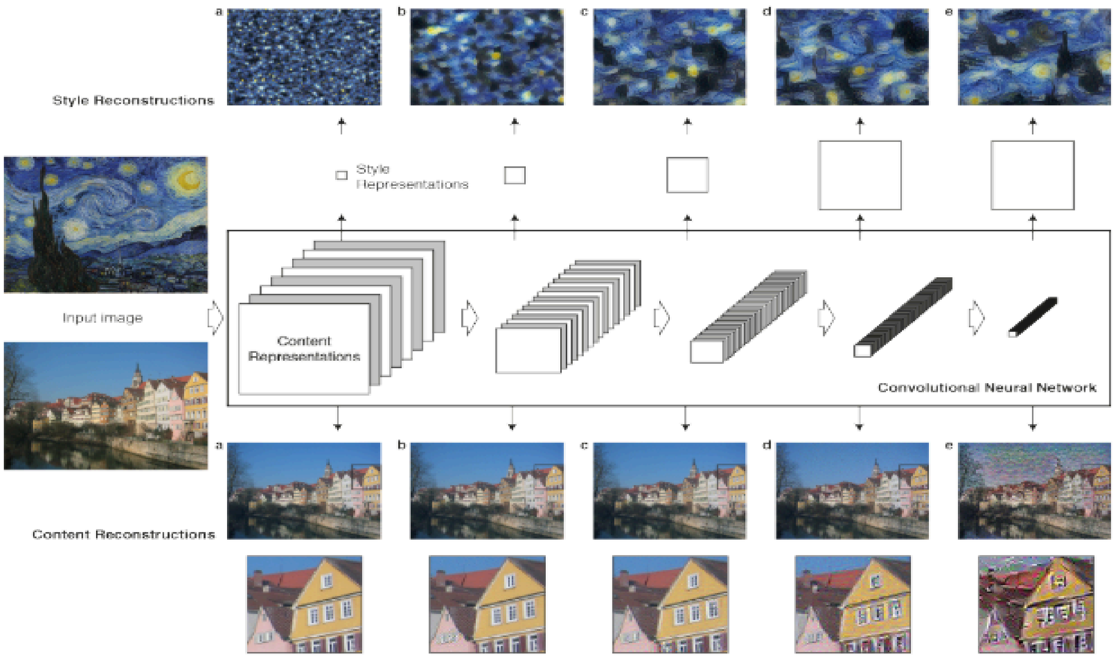


Fig. 7: A **Convolutional Neural Network (CNN).** With each iteration, the size of the image decreases using a downsampling method (eg. max pooling). At each layer in the CNN, a reconstruction of the input gets created.

Source: [10]

Generally speaking, every network layer specifies a non-linear filter bank, the complexity of which rises as the number of layers increases. As a result, the filter responses to an input image that is provided $\vec{x}$ are encoded in every CNN layer. Each of the N_l feature maps in a layer with N_l distinct filters has a size of M_l, which is the feature map's height times width. Thus, a matrix $F^l \in R^{N_l \times M_l}$, where F^l_{ij} is the activation of the ith filter at location j in layer *l*, can be used to store the answers in a layer *l*. Let $\vec{p}$ and $\vec{x}$ represent the original and generated images, respectively, and P^l and F^l represent the corresponding feature representations in layer *l*. The squared-error loss between the two feature representations is then defined.

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \Sigma_{i,j} (F^l_{ij} - P^l_{ij})^2 \qquad (1)$$

The derivative of this loss relative to layer l activations is equal to

$$\frac{\partial \mathcal{L}_{content}}{\partial F^l_{ij}} = \begin{cases} \left(F^l - P^l\right)_{ij} & if\ F^l_{ij} > 0 \\ 0 & if\ F^l_{ij} < 0 \end{cases} \qquad (2)$$

from which standard error back-propagation can be used to calculate the gradient with respect to the image $\vec{x}$. As a result, we can modify the output image $\vec{x}$ until it produces the same response as the original image $\vec{p}$ in a particular CNN layer. The five content reconstructions originate from the layers "conv1_1" (a), "conv2_1" (b), "conv3_1" (c), "conv4_1" (d), and "conv5_1" (e) of the original VGG-Network.

On top of the CNN output in each layer of the network, we built a style representation that computes the correlations between the different filter responses, where the expectation is taken over the spatial range of the input image. The Gram matrix $G^l \in R^{N_l \times N_l}$, where G^l_{ij} is the inner product between the vectorized feature map, provides these feature correlations. In layer l, i and j:

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk} \qquad (3)$$

To create a texture that matches the style of a given image, gradient descent is computed onto a copy of the original content image to find another image that fits the style representation of the original image. Thus, let $\vec{a}$ and $\vec{x}$ represent the original and generated images, respectively, and let A^l and G^l denote the corresponding style representations in layer l. Next, the layer's portion of the overall loss is

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} \left(G^l_{ij} - A^l_{ij}\right)^2 \qquad (4)$$

and the total loss would be

$$\mathcal{L}_{style}\left(\vec{a}, \vec{x}\right) = \sum_{1=0}^{L} \omega_l E_l \qquad (5)$$

where w_l are weighting factors that indicate how much each layer contributed to the overall loss. It is possible to calculate analytically the derivative of E_l in relation to the layer l activations:

$$\frac{\partial E_l}{\partial F^l_{ij}} = \begin{cases} \frac{1}{N_l^2 M_l^2} ((F^l)^T \left(G^l - A^l\right))_{ji} & if\ F^l_{ij} > 0 \\ 0 & if\ F^l_{ij} < 0 \end{cases} \qquad (6)$$

To create images that combine the content of an image with the style of a painting, we jointly reduce the distance between the image being generated from the content representation of the image in one layer of the network and the style representation of the painting in several layers of the CNN. Thus, let $\vec{a}$ represent the artwork and $\vec{p}$ the image. The loss that we minimise is

$$\mathcal{L}_{total}\left(\vec{p}, \vec{a}, \vec{x}\right) = \alpha \mathcal{L}_{content}\left(\vec{p}, \vec{x}\right) + \beta \mathcal{L}_{style}\left(\vec{a}, \vec{x}\right) \qquad (7)$$

where the weighting variables for content and style reconstruction are denoted by α and β, respectively.

Text-to-image Generation using Stable Diffusion

Often, when we think of the content image, we picture a particular scene or item. Taking your own pictures or browsing stock photo libraries may not produce images that exactly match your ideal lighting, composition, or detail level. Finding an image that perfectly captures the desired subject matter can be difficult. This is where Stable Diffusion's text-to-image generation capabilities offer a powerful solution. With detailed textual instructions, Stable Diffusion gives you complete control over every aspect of the content image, unlike searching for pre-existing photos. Indicate the items, how they are arranged, the lighting, and even the atmosphere in general. This degree of control guarantees that the created content precisely matches your intended final style transfer outcome. This generated image can hence be used as the content image to perform neural style transfer using any preferred style image, allowing the user to have even more control over what specific art style the final output image should be.

Stable Diffusion is a latent text-to-image diffusion model. A diffusion model that is trained in the autoencoder's latent space is combined with an autoencoder to create the latent diffusion model known as Stable Diffusion v2. U-Net, a text encoder, and the variational autoencoder (VAE) make up stable diffusion. Training diffusion models, which may be viewed as a series of denoising autoencoders, aims to eliminate successive Gaussian noise applications on training images.

While in training,

- An encoder transforms images into latent representations by encoding them. The autoencoder transfers pictures of shape H x W x 3 to latents of shape H/f x W/f x 4 using a relative downsampling ratio of 8.
- The OpenCLIP-ViT/H text-encoder is used to encode text prompts.
- Through cross-attention, the text encoder's output is fed into the latent diffusion model's UNet block, which comprises a ResNet backbone.
- The loss serves as a reconstruction goal between the UNet's prediction and the noise that was introduced to the latent.

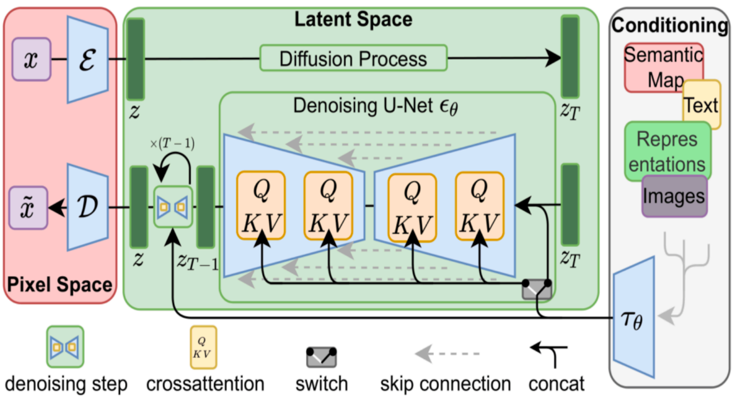


Fig. 8: Conditioning LDMs using a more general cross-attention technique or concatenation is desired. Source: [3]

5 billion image-text pairs were categorised based on language and divided into different datasets by resolution, predicted likelihood of having a watermark, and predicted "aesthetic" score (i.e., subjective visual quality) in the publicly available LAION-5B dataset, which was created from Common Crawl data scraped from the internet. This is where Stable Diffusion was trained using pairs of images and captions.

Latent Diffusion Models

Diffusion models are probabilistic models that are intended to teach the reverse process of a fixed Markov Chain of length T by gradually denoising a normally distributed variable. This process results in the learning of a data distribution p(x). The most effective models for picture synthesis use a reweighted version of the variational lower bound on p(x), which is similar to denoising score-matching. The denoising autoencoders $\varepsilon_\theta(x_t, t)$; t = 1... T, which are trained to predict a denoised variant of their input x_t, where x_t is a noisy version of the input x, can be understood as an evenly weighted sequence of these models. This objective is simplified as

$$L_{DM} = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[\,\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2\right] \qquad (8)$$

with t uniformly sampled from {1, . . . , T}.

The perceptual compression models containing ε and $\mathcal{D}$ are chosen so that it is possible to be able to access a low-dimensional, effective latent space where high-frequency, undetectable information are abstracted out, making it suitable for likelihood-based generative models. The underlying UNet is built primarily from 2D convolutional layers, and the objective is further focused on the perceptually most relevant bits using the reweighted bound, which is now

$$L_{DM} = \mathbb{E}_{\varepsilon(x), \epsilon \sim \mathcal{N}(0,1), t} [\|\epsilon - \epsilon_\theta (z_t, t)\|_2^2] \qquad (9)$$

As the forward process is fixed, samples from p(z) may be decoded to image space with a single pass through $\mathcal{D}$, and z_t can be effectively acquired from ε during training.

Conditioning Mechanisms

Diffusion models (DMs) hold immense potential for generating images based on additional information beyond simple categories or blurry starting points. This research explores a novel approach to unlock this potential by leveraging the power of cross-attention mechanisms. These likelihood-based models are capable of modelling conditional distributions of the form $p(z|y)$, where y denotes any type of input such as text, semantic maps etc. This can be done by using a conditional denoising autoencoder $\varepsilon_\theta(z_t, t, y)$. By adding the cross-attention mechanism to the basic UNet backbone of DMs, we make them more versatile conditional image generators. This technique works well for teaching attention-based models with a range of input modalities.

Fig. 9: Samples of images generated using Stable Diffusion v2

IV. RESULT

Here's a comparison of how our model has learned to identify the complexity of mural paintings and generate art using it:

Fig 10: Content image

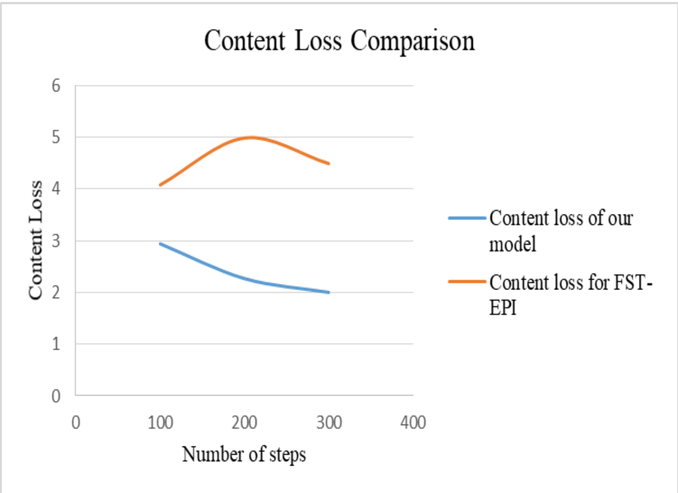
Fig 11: Style image

Fig 12: Image generated by FST-EPI

Fig 13: Image generated by our model

From the above comparison, it is evident that our model was able to encapsulate the spirit of mural painting, while FST-EPI has also performed style transfer. These murals are elaborate tapestries with vivid colours that depict mythology and nature. Our model does a great job of reproducing these features. It recognizes the significance of crisp lines for motifs and figures, the harmony of earthy tones with colourful accents, and the distinctive use of perspective.

The graphs given below demonstrate and compare the losses between our model and FST-EPI, taking into consideration the results generated above.



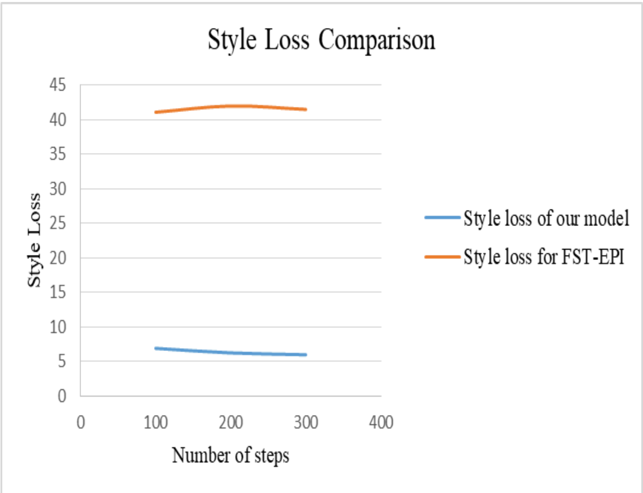


Fig 14: The drastic difference comes in the style loss, evident from the images generated by each model. Our model was able to transfer the style onto the resulting image with very little style loss, also ensuring that the contents of the original image remain preserved.

The main philosophy of neural style transfer is to be able to adjust the weights of both the content and style image, to determine how accurately and beautifully the model is able to reproduce the intricate details in the style image, while being able to preserve the content in the content image. While style loss adds to an image's artistic flair, content loss maintains an image's subject matter. The model preserves the objects, forms, and general composition of the content image in the final output by minimising the loss of content. The model incorporates the textures, brushstrokes, and distinctive colour patterns of the style image into the generated image by reducing the style loss. With the increase in the number of steps, the content image slowly starts to incorporate the broad stylistic elements from the style image. This is explained further below:

Image number	Content weight α	Style weight β	Number of steps	Content loss	Style loss
1	1	10000	100	22.941767	3.078124
2	1	10000	200	21.269157	2.831555
3	1	10000	300	21.001907	2.547383

Table 1: Setting the content weight α and style weight β, and the number of steps taken by the VGG-19 model to achieve the final result

Here, we have generated an image of a rabbit using a simple text prompt "a cute white rabbit". This image was used as the content image for the style transfer method.

Fig. 15: Image generated using Stable Diffusion v2

Fig. 16: Style image used for this experiment

1 2 3

Fig 17: The style of the content image is evidently changing with each iteration in the above case. Our model was able to replicate the mural art style onto the image of the rabbit excellently.

V. CONCLUSION

In this paper, the main intention is to target the ongoing challenges related to transferring art styles of comparatively higher creative efficiency. This method integrates modules such as background creation, background integration, and fast transfer, aiming to both preserve and innovate pattern elements.We have presented a novel approach for text-to-image generation, combining the stability and efficiency of diffusion models with the artistic rendering capabilities of neural style transfer. Our proposed model exhibits promising results, generating high-quality images that match the input text descriptions while integrating desired artistic styles. This approach opens up new possibilities for creating visually appealing images from textual descriptions, followed by educating and expanding the existence and influence of historic and famous art styles and their artists throughout. Comparison to existing models reveal that while the generation capabilities are similar, our model exhibits exceptional performance in style transfer of not only famous art and its styles but also those of a traditional field of lesser known varieties, showcasing its effectiveness and competitiveness among existing solutions. Since the model has been trained on mural art, and even other categories of art such as sculptures, engravings and so on, along with tweaks in parameters, it shows great improvement and likeness in its style transfer capabilities compared to existing models. Further research could explore refining the model's performance and scalability, as well as investigating its applicability to other domains beyond image generation.

VI. REFERENCES

1. S.A. Khowaja et al., "Extending User Control for Image Stylization using Hierarchical Style Transfer Networks," Heliyon, 2024.
2. Y. Zheng et al., "Fast style transfer for ethnic patterns innovation," Expert Systems with Applications, vol. 249, Part B, p. 123627, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.123627
3. R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684-10695, 2022.
4. N. Huang et al., "Diffstyler: Controllable dual diffusion for text-driven image stylization," IEEE Transactions on Neural Networks and Learning Systems, 2024.
5. X. Han, Y. Wu, and R. Wan, "A Method for Style Transfer from Artistic Images Based on Depth Extraction Generative Adversarial Network," Applied Sciences, vol. 13, no. 2, p. 867, 2023.
6. Q. Cai et al., "Image Neural Style Transfer: A Review," Computers and Electrical Engineering, vol. 108, p. 108723, 2023.
7. Y. Zhang et al., "Inversion-based style transfer with diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10146-10156, 2023.
8. S. Ravi, "COLOUR CULTURE AND IDENTITY: INFLUENCE OF COLOURS ON KERALA MURAL ART," IJASOS- International E-journal of Advances in Social Sciences, vol. 1, pp. 315, 2015, doi: 10.18769/ijasos.6675.
9. D. Lee, H. Kim, and Y. Park, "Image Generation with Conditional Generative Adversarial Networks," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1234-1245, Jun. 2023.
10. L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," arXiv:1508.06576 [cs.CV], 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1508.06576, Sep 2015.
11. H. Yu, I. Chen, and J. Sun, "Attention-based Neural Networks for Artistic Image Generation," Proc. European Conference on Computer Vision (ECCV), pp. 901-917, Oct. 2023.
12. P.N. Deelaka, "Neural Artistic Style Transfer with Conditional Adversaria," arXiv preprint arXiv:2302.03875, 2023.
13. J. Miah, D.M. Cao, M.A. Sayed, and M.S. Haque, "Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks," arXiv preprint arXiv:2310.18237, 2023.
14. Z. Wang, L. Zhao, and W. Xing, "Stylediffusion: Controllable disentangled style transfer via diffusion models," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 7677-7689.
15. X. Yu and G. Zhou, "Arbitrary style transfer via content consistency and style consistency," The Visual Computer, vol. 40, no. 3, pp. 1369-1382.