Deep Learning-Based Degradation Detection and Inpainting of Tamil Nadu Temple Murals Using Transformer Networks and Diffusion Models

Ashwini Ramachandran
Department of Electronics and Communication Engineering, SRM Institute of Science and Technology
Ramapuram, Chennai, India
ar8673@srmist.edu.in

Vani Rajamanickam
Department of Electronics and Communication Engineering, SRM Institute of Science and Technology
Ramapuram, Chennai, India
vanir@srmist.edu.in

Abstract: *Restoring historical murals for cultural heritage conservation remains difficult because these intricate artworks deteriorate severely throughout their existence. Preserving historical artwork requires precise techniques to both detect and digitally restore its deteriorated areas. The combination of manual restoration approaches and basic digital tools fails to deliver precise and efficient results, leading to inconsistent outcomes and the loss of original artistic details. This study develops an enhanced approach that integrates swin transformer for detecting damage with stable diffusion for performing the restoration process to achieve superior precision and reliability in restoration work. A dataset consisting of 1,200 high-resolution tamil nadu temple mural images which contains annotated degradation types for training and evaluating proposed models. Experimental results show that using swin transformer and stable diffusion produces superior results with an Intersection over Union (IoU) of 0.92 along with Peak Signal-to-Noise Ratio (PSNR) of 32.5 dB, Structural Similarity Index (SSIM) of 0.95 and Frechet Inception Distance (FID) of 12.3. The study validates the proposed methodology as an efficient system for enhancing the quality and mechanical fidelity in mural restoration applications which indicates its potential use as a solution for digital art restoration projects.*

Keywords: *Art preservation, cultural heritage conservation, historical mural restoration, degradation detection, digital inpainting, swin transformer, stable diffusion.*

Received March 27, 2025; accepted November 30, 2025
https://doi.org/10.34028/iajit/23/2/13

1. Introduction

Historical mural restoration stands as an essential process in cultural heritage preservation since it helps us to understand both past artistic techniques and societal beliefs of previous cultures. These murals, which are several centuries old, are experiencing deterioration because of environmental threats and human damage thus demanding complex restoration processes which protect their genuine expressions simultaneously ensuring long-term survival for posterity [8, 17].

The traditional restoration technique is mainly relied on manual process which required trained professionals to perform detailed work. The traditional restorative methods are expensive and consume lengthy time period. They produce inconsistent results through human involvement therefore causing permanent alteration to original artwork. The manual restoration method proves insufficient for the large scale requirements of worldwide heritage protection since thousands of artworks need conservation treatment [3, 10].

The conventional digital process has added automatic features through basic digital tools which perform color modifications and clean up broken elements from artwork images. Digital photography and photoshop software's have become standard techniques for digital documentation and post-processing of artworks. These conventional methods fail to deliver precise restorations because they fail to identify and address the distinct degradation patterns exist in artworks. These systems fail to integrate advanced Artificial Intelligence (AI) technology which enables them to learn from extensive art restoration datasets thereby reducing their practicality in highly difficult restoration tasks [9, 13].

To address these challenges, this study creates a new computational system which uses Swin transformer to detect degradation precisely and stable diffusion models to restore damaged areas effectively. The utilization of deep learning algorithms strengthens restoration workflows by effectively capturing historical mural details and textures which results in improved efficiency and consistency.

The research contributes significantly by:

- Utilizing cutting-edge AI to detect degradation with high accuracy and restore murals while maintaining artistic authenticity.
- Offering a scalable solution applicable to large datasets of historical artworks, significantly reducing time and cost compared to manual methods.

- Ensuring that restorations are reversible and do not permanently alter the original artwork, adhering to ethical standards of art conservation.

This study is organized in the following manner: section 2 delivers an extensive study of art restoration methods and artificial intelligence applications in cultural heritage contexts. The section 3 elaborates the proposed methodology which encompasses data preparation steps, model training and metric evaluation. The experimental results appear in section 4 followed by detailed discussion in section 5 that analyzes the obtained results in relation to existing restoration challenges. The study concludes in section 6 which summarizes the research findings with the present study's limitations and future work.

2. Related Work

Advanced computational methods have emerged as a vital role in restoring ancient wall paintings during the last several years. Nasri and Huang [8] utilized the mahalanobis distance method for both mural image augmentation and lacuna detection but the technique depended on manual features without universal applicability. Xiao *et al.* [17] employed deep learning technology to restore high-resolution murals but this approach requires large databases and high capacity computational systems. Xu *et al.* [19] developed MuralDiff which incorporated diffusion models to restore murals through efficient fracture detection processing but the effectiveness reduced when dealing with large pre-training processes. Xu and Fu [18] built an AI system for color recovery that performed well regarding restoration tasks but the system failed to handle noise effectively from low-resolution artwork images. Shi and Meng [12] presented a deep learning methodology based on sampling which achieved better accuracy but required human involvement for implementation during complex deterioration situations. The research demonstrates outstanding potential of advanced methods alongside hindrances related to widespread adoption across different cases and use of substantial computing capabilities. Qian *et al.* [10] developed a deep learning analysis system to monitor wall damage with outstanding results yet it could not be applied to undocumented murals because of its requirement for dataset labels. Cao *et al.* [3] developed an Inception-v3 model variant for dynasty-based mural classification which achieved precise results even though it encountered challenges while processing murals from transitional times because of artistic style convergence. Sun *et al.* [14] determined hyperspectral imaging produced successful mural restoration outcomes by dealing with scratched walls yet required advanced equipment which proved difficult to use. Xi *et al.* [16] developed a framework for old mural restoration through advanced methods that delivered visually satisfactory results although they failed to establish quantitative accuracy measures for restoration quality. Pawar *et al.* [9] developed deep learning technology for noise detection in ancient images yet this technique remained restrained by the wide variety of historical image noise patterns. Singh *et al.* [13] developed a partial convolutional network for mural restoration yet the method required high-quality reference images for successful execution and thus made it unfit for severely damaged murals. Yu *et al.* [21] implemented a new technology to obtain refined sketches from ancient murals for better study of artistic methods and design structures. The method showed exceptional precision in sketch recovery yet required high-quality input pictures as a necessary condition which makes it ineffective for deteriorated mural restorations. Choi *et al.* [4] developed a deep learning approach to segment 3D inscriptions when extracting them from rough stelae surfaces. The method succeeded at interpreting rough surfaces alongside shape variances however the sophisticated 3D data processing flows stood as a barrier for live system applications. Yadav *et al.* [20] developed image inpainting methods to restore murals in shekhawati region. The authors developed restorative techniques combining damage detection with mask generation while proving efficient results but needed powerful computational systems for testing. Cai *et al.* [2] created an inpainting system for ancient murals using adversarial generative networks with bidirectional feature adaptation. The method achieved restoring content with high-quality results which maintained fine details yet remained sensitive to changes in training parameters due to its adversarial training system.

Jiang *et al.* [6] implemented eXplainable Artificial Intelligence (XAI) to study restoration model choices in ancient architecture and lacquer art while giving explanations of decision-making processes. The enhanced transparency in XAI techniques faced difficulties because of complex deep learning models. Sharma and Kukreja [11] developed a faster Region-based Convolutional Neural Network (R-CNN) framework to enhance comic panel detection strategies in cultural heritage preservation. The technique revealed effectiveness in locating historical mural fragments yet struggled when applied to different artistic variations and various forms of damage.

Valdes *et al.* [15] developed a machine learning detection system which uses Generative Adversarial Networks (GANs) along with U-Net and K-Nearest Neighbors (KNN) and Convolutional Neural Networks (CNNs) to identify mural deteriorations. The method produced successful results for detecting mural deterioration but encountered limitations when trying to apply it to various mural styles under different environmental conditions. Gerasimiuk *et al.* [5] established MURAL as an unsupervised random forest-based embedding tool for electronic health record data analysis. The unsupervised learning methods from healthcare research stand as potential models for mural restoration because these methods excelled at treating untagged and under-annotated data sets. The research investigation did not explicitly discuss cultural heritage applications even though it created conditions for adaptations in mural restoration work. The research by Lianji and Haihong [7] presented deep learning strategies for Dunhuang mural restoration by implementing CNNs and GANs. The applied method delivered substantial progress in restorative imaging while successfully protecting both detailed features and colors of dunhuang murals. Due to their substantial network size requirements, GANs faced limitations in real-time applications. Agarwal and Dixit [1] propose a Style-Consistent Generative Adversarial Network (SCGAN) based framework to recognize restored images, using a caffe-driven Principal Component Analysis (PCA) filtration step to suppress irrelevant variations and stabilize the learned features. Experiments on restored imagery indicate improved recognition consistency and robustness, suggesting that PCA-guided filtering can complement GAN outputs for downstream classification tasks. Extensive research demonstrates the utility of machine learning and deep learning approaches in restoration and deterioration identification tasks for cultural artifacts but still requires more efficient and flexible models which will address cultural heritage conservation requirements as shown in Table 1.

Table 1. Summary of related works in mural restoration and degradation detection.

Reference	Dataset	Application	Approach	Results	Disadvantages
Nasri and Huang [8]	Bey's Palace mural images	Lacuna extraction	Mahalanobis distance classification	Accurate lacuna detection in small datasets	Manual feature extraction, limited generalization
Xu *et al*. [19]	Large mural dataset	Mural restoration	Diffusion model (MuralDiff)	High-quality restoration with fracture detection	Computationally intensive, large pre-training needed
Xu and Fu [18]	Low-res relic images	Colour restoration	DL-based color recovery	Good color fidelity	Poor noise handling
Shi and Meng [12]	Complex deterioration murals	Sample-based Deep Learning (DL) restoration	High restoration accuracy	Requires human intervention for complex cases	-
Qian *et al.* [10]	Ancient murals with labels	Damage grading	CNN-based grading	Excellent classification accuracy	Cannot handle unlabeled data
Cai *et al.* [2]	Ancient murals	Inpainting	GAN with bidirectional feature adaptation	Preserves fine details	Sensitive to training parameters
Lianji and Haihong [7]	Dunhuang murals	Restoration	CNN+GAN hybrid	Retains color and details	Large network size, not real-time capable

2.1. Problem Formulation

The restoration of ancient murals requires complex procedures because they encounter different types of damage which includes cracking as well as fading and peeling and biological corruption. Traditional restoration techniques demand extensive manual labor that leads to long time-consuming operations and subjectivity errors during the process. but The automated restoration process has shown promising results through deep learning and GAN model but also faces several difficulties in real time implementation. The system faces three main drawbacks in terms of its operational requirements: significant data labeling necessities, expensive computational resources and adaptation challenges for different mural art compositions and forms of deterioration. Cultural and historical authenticity of mural artwork should be preserved throughout the restoration process. This study develops a cutting-edge pipeline structure for restoration which integrates modern degeneration detection systems with text-to-image processing techniques based on Swin transformer and stable diffusion to address existing challenges. The proposed restoration model aims to achieve outstanding outcomes by using minimal resources and maintaining the original artistic value of the murals.

3. System Methodology

The proposed research implements sophisticated deep learning approaches to detect degradation symptoms in historical murals and performing restoration processes, as shown in Figure 1. The methodology implements a complete pipeline architecture that consists of three primary steps which include data preprocessing, model training and evaluation processing. The section details all methodological steps in their entirety.

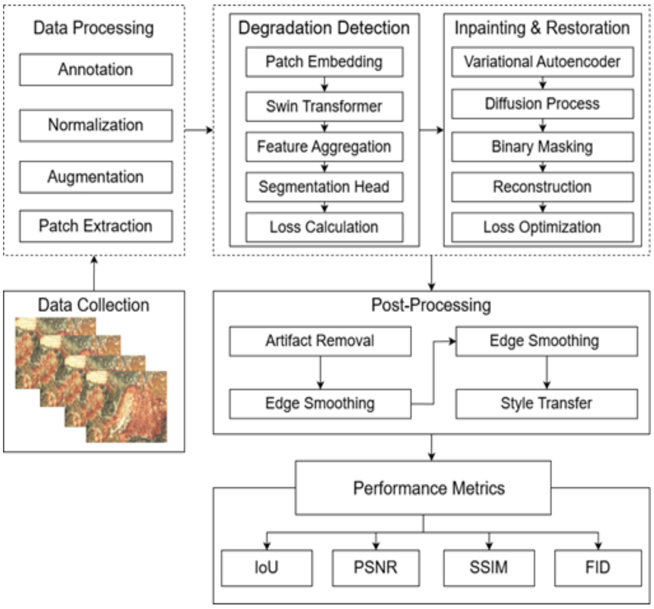

Figure 1. Proposed architecture diagram.

3.1. Data Pre-Processing

The initial work entails compiling a detailed collection of high-definition photos from Tamil Nadu temple murals. A proper annotation process takes place on these images to identify different degradation types like cracks, fading, peeling and biological growth. The data preprocessing method follows these successive operations for every extracted image:

- Normalization: normalization is a vital preprocessing operation to minimize intra-dataset variability that may affect model training. The normalization step transforms color variations and lighting differences between images to make models focus on structural degradation patterns instead of light intensity or hue variations. Each pixel intensity requires modification through a standardization process during normalization as in Equation (1).

$$I' = \frac{I - \mu}{\sigma} \tag{1}$$

where I is the original image, μ is the mean pixel value, and σ is the standard deviation of pixel values.

- Augmentation: Augmentation techniques are very important to make machine learning models more robust and generalizable because these methods create artificial dataset expansions through modifications of original images. The modification process applies different conditions to murals, which help the model understand degradation perception from any viewing perspective through angular rotations, allow scaling for mural dimension flexibility, and conduct horizontal or vertical mirroring to check adaptability. The diversified training samples provide more illustrations without any extra expense to collect actual data, thus strengthening the model’s consistency when performing under varying environmental conditions.

3.2. Degradation Detection Using Swin Transformer

The Swin transformer provides innovative vision capabilities through its adaptation of transformer architecture to successfully mark and outline degraded areas in mural images.

3.2.1. Image Patch Processing

The initial processing divides the input image I into separate $P \times P$ non-overlapping blocks. Linear projection transforms flattened vectors representing image patches into feature vectors x_i which exist as a result of their linear projection method. The initial step prepares information for transformer models through its patch embedding structure which forms a basic structured data representation, as in Equation (2).

$$z_0 = PatchEmbed(I) \tag{2}$$

Here, z_0 represents the initial feature representations derived from the input image patches.

3.2.2. Swin Transformer Architecture

The Swin transformer hierarchical design works with shifted window procedures to process image patches as its core operational method. The model architecture successfully captures smaller details and larger context elements which represent essential characteristics for identifying advanced damages in mural paintings. Figure 2 depicts the Swin transformer architecture which is primarily composed of the window and shifted Window-based Multi-head Self-Attention Mechanisms (W-MSA and SW-MSA). Each block embeds layers for Layer Normalization (LN), the multi-head self-attention mechanism, and a Multi-Layer Perceptron (MLP), with residual connections around each sub-layer.

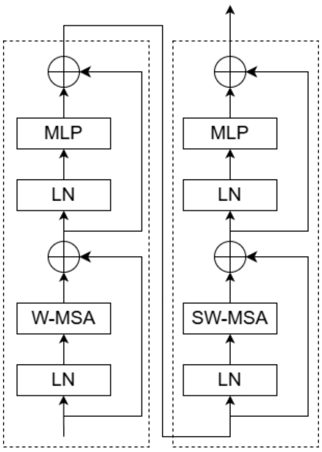

Figure 2. Swin transformer block diagram.

1) Window-based Multi-head Self-Attention (W-MSA)

The model operates through attention mechanisms within localized windows which enable it to focus on image areas one at a time.

The attention mechanism can be expressed as in Equation (3):

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

Where Q, K, and V represent the query, key, and value matrices derived from the feature vectors, and d_k is the dimensionality of the key vectors.

2) Shifted Window-based Multi-Head Self-Attention (SW-MSA)

The attention calculation area extends across multiple layers to help the model process contextual data. The shifted window-based multi-head self-attention mechanism is described in Equation (4):

$$SW - MSA(\cdot) = WindowPartition(Shift(W - MSA(Unshift(WindowReverse(\cdot))))) \tag{4}$$

Through the shifting mechanism, users can prevent their focus from becoming isolated in separate windows while developing a better image comprehension by enabling window-to-window information sharing.

3.2.3. Output Segmentation

Next, the processed feature maps z_L from the last layer of the Swin transformer are passed through a segmentation head. This head has special training to produce a binary mask M that shows degraded versus non-degraded areas, as in Equation (5).

$$M = SegmentationHead(z_L) \tag{5}$$

3.2.4. Loss Function

The model uses the binary cross-entropy loss function [8, 17] which calculates the difference between predicted degradation probabilities and actual manually labeled degradation labels, as in Equation (6):

$$L_{det} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\ \log(1 - \hat{y}_i)] \tag{6}$$

Here, y_i denotes the ground truth label for the i-th pixel, indicating whether it is degraded (1) or non-degraded (0), and $\hat{y}_i$ is the predicted probability that the i-th pixel is degraded.

The combination of advanced computational techniques using the Swin transformer creates a strong detection system for historical murals deterioration assessment and delivering superior accuracy along with efficiency.

3.3. Inpainting Using Stable Diffusion

Stable diffusion is a generative model that employs knowledge of data distribution within a latent space to treat damaged or incomplete areas. Figure 3 illustrates the stable diffusion process used in inpainting.

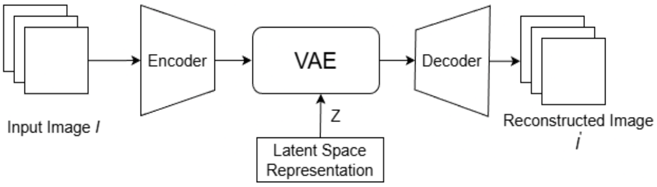

Figure 3. Stable diffusion inpainting process.

3.3.1. Latent Space Representation

The input image I is encoded into a latent representation z using a Variational Autoencoder (VAE), as in Equation (7):

$$z = Encoder(I) \tag{7}$$

The latent space reduces the computational complexity of the diffusion process.

3.3.2. Diffusion Process

The diffusion process consists of two phases:

1) **Forward Process**

Noise is gradually added to the latent representation z over T time steps, as in Equation (8):

$$q(z_t \mid z_t - 1) = N(z_t; \sqrt{1 - \beta_t}\ z_{t-1}\ ,\beta_t I) \tag{8}$$

Where β_t is the noise schedule.

2) **Reverse Process**

The model learns to denoise the latent representation by predicting the noise ε_θ at each time step, as in Equation (9):

$$P_\theta(z_t - 1 \mid z_t) = N(z_t - 1; \mu_\theta(z_t, t), \Sigma_\theta(z_t, t)) \tag{9}$$

where μ_θ and Σ_θ are the mean and variance predicted by the model.

3.3.3. Inpainting with Mask Guidance

The degraded regions are inpainted by conditioning the diffusion process on the surrounding context, as in Equation (10):

$$\hat{I} = Decoder(Denoise(z_T, M)) \tag{10}$$

Where M is the binary mask indicating degraded regions, and *Denoise* is the reverse diffusion process.

3.4. Loss Function

The inpainting model is trained using a combination of reconstruction loss and perceptual loss, as in Equation (11):

$$L_{inpaint} = \lambda_1\ L_{rec} + \lambda_2 L_{perc} \tag{11}$$

Where $L_{rec}=\|I-\hat{I}\|_1$ is the $L1$ reconstruction loss, and L_{rec} is the perceptual loss computed using a pre-trained Visual Geometry Group (VGG) network to ensure that the inpainted regions are visually realistic.

3.5. End-to-End Pipeline Process

The end-to-end pipeline designed for mural restoration seamlessly integrates degradation detection and inpainting processes to efficiently restore degraded mural images. The system begins with the input of degraded mural image I through which the Swin transformer analyzes and detects degraded areas before producing binary mask M. The mask provides visual indications for damaged areas which need complex fixation while assisting the following analytical process. The stable diffusion model performs intelligent reconstruction on the damaged or faded regions from M through usage of detailed information in M. The binary mask operating with the inpainting algorithm protects the original mural aesthetics so the system produces a restored version of the mural through $\hat{I}$. The complete approach both enables automated restoration operations along with improved exactness and better quality in the final image result which maintains the cultural history of every mural.

3.6. Evaluation Metrics

The pipeline performance is evaluated by:

- **Intersection over Union (IoU)**: this evaluation method calculates the intersection between estimated and genuine degraded regions through the IoU calculation, as in Equation (12).

$$IoU = \frac{TP}{FP + FN + TP} \tag{12}$$

- **Peak Signal-to-Noise Ratio (PSNR)**: PSNR evaluates the image restoration quality by comparing the restored image against its original version, as in Equation (13).

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{13}$$

- **Structural Similarity Index (SSIM)**: measures the structural similarity between the restored and original images as in Equation (14).

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 C_2)} \tag{14}$$

- **Fréchet Inception Distance (FID)**: measures the perceptual quality of the restored image, as in Equation (15).

$$FID = \| \mu_r - \mu_g \|^2 + Tr\left(\Sigma_r + \Sigma_g - 2(\Sigma_r + \Sigma_g)^{\frac{1}{2}}\right) \tag{15}$$

3.7. Performance Measurements

The performance of the restoration process is evaluated through qualitative measurements including training time data, inference time data, and Graphics Processing Unit (GPU) memory requirements. Training duration is defined by the time needed for complete training of models across a set of datasets, an important factor for projects wherein model adaptability upon new datasets is primarily concerned. Inference time represents the time taken for processing a single input image to the output image, basically it shows the feasibility of the system for real-time applications. Lastly, GPU memory usage indicates the memory utilized during both training and inference runs, revealing many resources the pipeline needs. In combination, these metrics provide a complete view of the pipeline’s performance, trading speed, accuracy, and utilization of assets to optimize the restoration process.

4. Experimental Results

4.1. Dataset Overview

The study makes use of a 1,200-image database with typical pixel resolution at 2,000 x 3,000 pixels for each image. The collection contains temple mural artwork ranging from various historical dynasties which was sourced from both temples and museums and archival records in Tamil Nadu, as shown in Figure 4.

A variety of art from the Pallava and Chola and Nayak and Maratha dynasties is included in this dataset which displays multiple artistic styles and thematic elements. The murals in this dataset exhibit various signs of aging and wear, such as cracks, fading, missing parts, and discoloration. The images contain binary masks which have documented all deteriorated areas through manual annotation. The thorough annotation method provides exact information for our model training and evaluating stages. A methodical process splits the dataset into three sets with training taking up 70% and validation along with testing each comprising 15% of the entire dataset. This split enables a thorough evaluation of the model performance along the entire machine learning workflow, where one can ensure that the models are robust and have good generalization power to unseen real-world data.

Figure 4. Sample images represent multiple vibrant historic temple wall art pieces found in Tamil Nadu India.

4.2. Training Parameters and Configurations

The proposed Swin transformer system and stable Diffusion models achieved best results through training with specific hyperparameter settings. The Swin transformer was trained for 100 epochs, with AdamW optimizing the learning process by controlling the decay of the initial learning rate (1×10^{-4}) using cosine annealing. The stable diffusion model underwent training for 200 epochs with a batch size of 8, utilizing the Adam optimizer and a learning rate of 5×10^{-5}. Training was conducted on NVIDIA A100 GPUs in mixed precision mode to reduce memory usage and accelerate computation. The training procedure employed different augmentation techniques, including random cropping, flipping, and rotation to enhance model performance across various conditions as detailed in Table 2. Early stopping was implemented to prevent overfitting, ensuring the selection of the best checkpoints based on validation performance.

Table 2. Training parameters and configurations.

Parameter	Swin transformer	Stable diffusion
Training epochs	100	200
Batch size	16	8
Optimizer	AdamW	Adam
Initial learning rate	1×10^{-4}	5×10^{-5}
Learning rate schedule	Cosine annealing	Fixed
GPU	NVIDIA A100 (all models were trained using NVIDIA A100 GPUs.)	NVIDIA A100 (all models were trained using NVIDIA A100 GPUs.)
Mixed precision	Yes	Yes
Data augmentation	Random cropping, flipping, rotation	Random cropping, flipping, rotation
Early stopping	Yes	Yes

4.3. Model Performance Comparison

The Table 3 shows the comparison of four different degradation detection methods assessed on the test set with a series of metrics that is IoU, precision, recall, and F1-score. The Swin transformer model outperforms all other models in achieving every evaluation metric with an IoU score of 0.92 and precision value of 0.94, recall value of 0.91 and an F1-score of 0.93, as depicted in Figure 5. The model demonstrates high ability for precise detection and segmentation of areas affected by degradation in the mural samples.

Table 3. Comparison of degradation detection models on test set.

Model	IoU	Precision	Recall	F1-Score
Swin transformer	0.92	0.94	0.91	0.93
Vision transformer	0.89	0.91	0.88	0.90
U-Net	0.85	0.87	0.84	0.86
ResNet-50	0.82	0.84	0.81	0.83

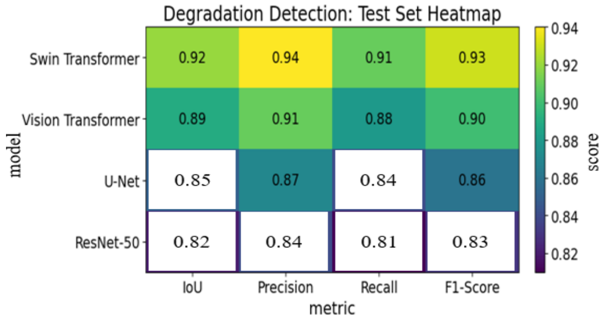

Figure 5. Performance evaluation of degradation detection algorithms.

The vision transformer shows an overall successful performance through its metrics which reached an IoU of 0.89 and precision of 0.91 and recall of 0.88 and F1-Score of 0.90 though slightly less than the Swin transformer results. U-Net demonstrates reliable performance as an image segmentation tool but shows lower accuracy than other models by obtaining an IoU measurement of 0.85 and scores of precision 0.87 and recall 0.84 and F1-score 0.86. The evaluation of ResNet-50 reveals its least accurate performance when applied to visual recognition tasks because it obtains an IoU of 0.82, precision of 0.84, recall of 0.81 and an F1-score of 0.83. The findings indicate ResNet-50 demonstrates effectiveness yet needs additional adjustments and augmentation methods to raise its performance rates on historical mural degradation analysis.

4.4. Inpainting Model Evaluation

The Table 4 reports the evaluation of different inpainting models on a test set in terms of PSNR, SSIM, and FID. The proposed stable diffusion model dominates all metrics with 32.5 dB PSNR measurement as well as 0.95 SSIM and 12.3 FID score, as illustrated in Figure 6. The satisfactory performance metrics demonstrate the model’s outstanding ability to produce high-quality restorations which show close similarities to original images. The Denoising Diffusion Probabilistic Model (DDPM) model achieved a PSNR rate of 31.8 dB plus an SSIM number of 0.94 and an FID score of 13.5 to establish its efficacy within image inpainting applications while trailing behind the proposed model. The GAN-Based Inpainting model exhibits efficient restoration abilities with a PSNR of 30.2 dB while achieving an SSIM of 0.92 and an FID of 15.8 yet produces images that stand somewhat apart from the original images compared to leading models. PatchMatch, which employs a different technique, registers a PSNR of 28.7 dB, an SSIM of 0.89, and an FID of 18.4. The metrics indicate this model successfully performs but does not achieve the same detailed quality that the other models show in texture replication. This comparative study presents the distinct advantages of each methodology, as well as the most appropriate ones to be used depending on the type of inpainting tasks to be performed.

Table 4. Comparison of inpainting models on test set.

Model	PSNR (dB)	SSIM	FID
Stable diffusion	32.5	0.95	12.3
DDPM	31.8	0.94	13.5
GAN-based inpainting	30.2	0.92	15.8
PatchMatch	28.7	0.89	18.4

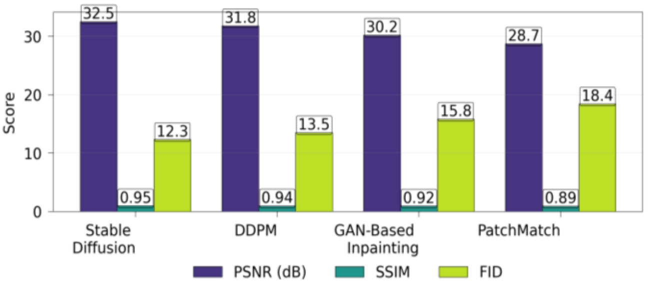

Figure 6. Performance assessment of different inpainting models.

4.5. Comprehensive Pipeline Performance Analysis

The performance evaluation of end-to-end pipelines consisting of degradation detectors and inpainting models appears in Table 5 through multiple assessment criteria including IoU, PSNR, SSIM, and FID. The proposed Swin transformer and stable Diffusion combination stands out as the superior system by achieving an IoU score of 0.92 together with PSNR measurement of 32.5 dB, SSIM at 0.95 and minimal FID score of 12.3, as displayed in Figure 7. The joint operation between these models creates superior detection along with better restoration quality through their excellent alignment. The Swin transformer and DDPM combination detects degradation at an IoU of 0.92 while the system produces PSNR at 31.8 dB and SSIM at 0.94 but yields an FID of 13.5. The proposed degradation detector reaches identification results matching its top-performing competitors but shows a slight decrease in restoration quality. The combination of vision transformer with stable diffusion yields an IoU score of 0.89 and PSNR of 32.0 dB and SSIM of 0.94 and FID of 13.0 which trails the best results across the proposed pipeline. Different inpainting methods demonstrate equal results when used with the same detection model under DDPM conditions compared to Swin-recorded statistics. The U-Net and GAN-based inpainting setup exhibits effective yet slightly less precise restoration capabilities through its results of an IoU 0.85 and PSNR 30.2 dB as well as SSIM 0.92 and FID 15.8. The ResNet-50 model connected to a PatchMatch system produces the lowest restoration capabilities evidenced by 0.82 IoU score combined with 28.7 dB PSNR reading alongside 0.89 SSIM value and 18.4 FID value. The analysis reveals significant differences between detection and inpainting tools so the Swin transformer along with stable diffusion emerges as an outstanding performer in dual detection and restoration operations.

Table 5. End-to-end pipeline performance.

Pipeline	IoU	PSNR (dB)	SSIM	FID
Swin transformer+Stable diffusion	0.92	32.5	0.95	12.3
Swin transformer+DDPM	0.92	31.8	0.94	13.5
Vision transformer+Stable diffusion	0.89	32.0	0.94	13.0
Vision transformer+DDPM	0.89	31.8	0.94	13.5
U-Net+GAN-based inpainting	0.85	30.2	0.92	15.8
ResNet-50+PatchMatch	0.82	28.7	0.89	18.4

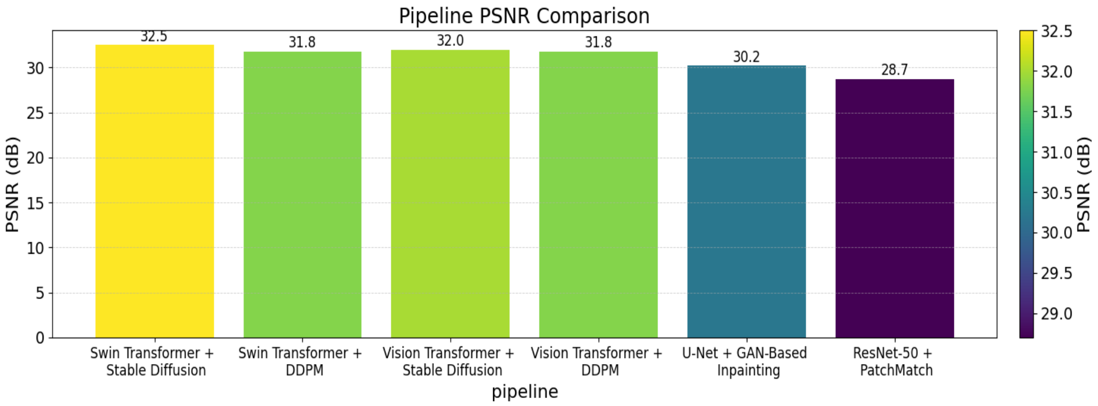

Figure 7. Various end-to-end pipeline combination over different essential metrics.

4.6. Computational Efficiency

The analysis in Table 6 evaluates computational efficiency by reviewing three measures which include training time, inference time per image and GPU memory usage across different models. The Swin transformer model operates efficiently due to its 12-hour training period while consuming 0.5 seconds to process each image and demanding only 8 GB GPU memory, as shown in Figure 8.

Table 6. Computational efficiency of the proposed model with baseline methods.

Model	Training time (hrs)	Inference time (sec/img)	GPU memory (GB)
Swin transformer	12	0.5	8
Vision transformer	15	0.7	10
Stable diffusion	20	2.0	12
DDPM	25	5.0	15
GAN-based inpainting	18	1.5	10
PatchMatch	22	10.0	14

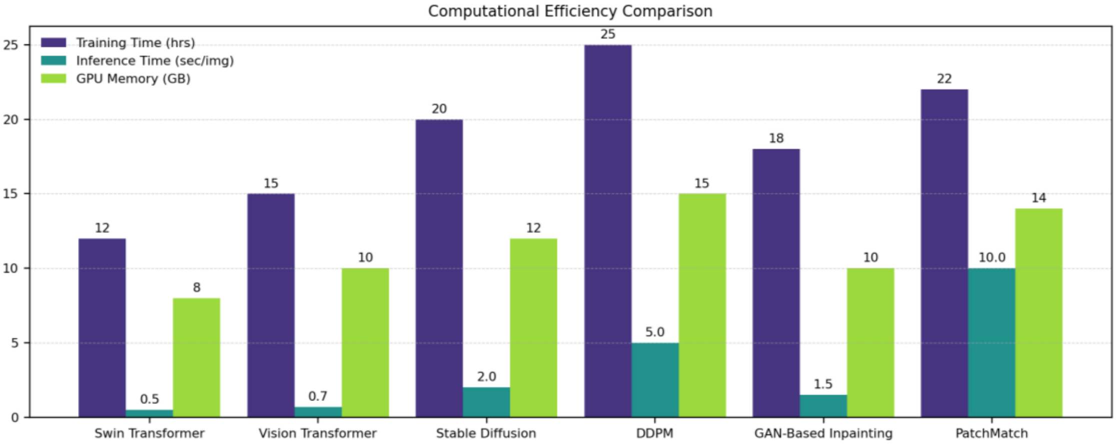

Figure 8. Computational efficiency of the different models.

The system shows effectiveness in rapid image processing while demanding minimal computer resources for its operations. the stable diffusion model needs 20 hours to train which exceeds the amount of time needed for the Swin transformer process due to its longer training duration. The high-quality product from this model becomes possible at the cost of 2.0 seconds per image processing time and 12 GB required GPU memory. The vision transformer requires fifteen hours for training yet performs slower than Swin transformer since it detects images with an inference time of 0.7 seconds per image at a GPU memory usage of 10 GB. This makes it less efficient than Swin transformer in all observed metrics. DDPM stands as the model requiring maximum computational resources since it needs 25 hours for training together with 5.0 seconds for inference time and 15 GB GPU memory usage. The performance of GAN-based inpainting surpasses DDPM when evaluating efficiency because it needs 18 hours for training and 1.5 seconds for inference as well as 10 GB GPU memory. The PatchMatch model demonstrates the lowest performance levels based on the stated metrics because it requires 22 hours of training time and consumes 10 seconds per image for inference along with using 14 GB of GPU memory. Other applications that need fast delivery or require minimal GPU memory would not find this model adequate. The Swin transformer stands out for its effective computational performance during training and operation due to the studied models varying substantially in terms of their computational requirements.

4.7. Art Historian's Qualitative Evaluation

Art historians made qualitative evaluations of the restored murals as shown in Table 6 through assessments of cultural accuracy together with artistic style retention as well as the visual quality across multiple pipelines. The Swin transformer and stable diffusion combination earned positive marks with critics awarding it 9.5 out of 10 for cultural accuracy and 9.7 out of 10 for artistic style preservation and 9.6 out of 10 for overall visual quality, as illustrated in Figure 9. The pipeline achieved remarkable marks which demonstrate its superior capacity to maintain original artwork integrity by using contemporary restoration methods. The Vision transformer system that works with DDPM obtained similar results to previous pipelines with scores reaching 9.0 for cultural accuracy and 9.2 for artistic style preservation and 9.1 for overall visual quality. The restoration capabilities match well with the authentic reproduction of artworks even though the top pipeline maintains a small advantage over this variant. The U-Net in combination with GAN-Based Inpainting scored lower than other methods according to the evaluation with 8.5 for cultural accuracy, 8.7 for artistic style preservation and 8.6 for overall visual quality. The scoring results indicate that the precision in retaining original artistic intentions and cultural elements falls slightly short of the leading pipeline. ResNet-50 along with PatchMatch achieved the most inferior rankings out of all examined pipelines since evaluators assigned ratings of 7.8 for cultural accuracy together with 8.0 for artistic style preservation and 7.9 for overall visual quality. The analysis revealed significant areas of improvement for detection and inpainting systems used together because their output demonstrated noticeable deviations from original art authenticity as well as aesthetic features. The assessment results demonstrate that superior recovery results at high precision levels depend heavily on selecting appropriate technological combinations which also respect historical art values.

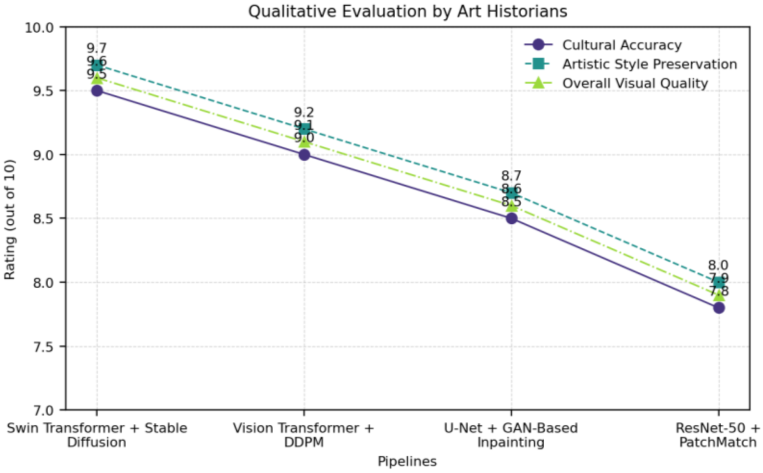

Figure 9. Qualitative evaluation by art historians.

Table 6. Qualitative evaluation by art historians.

Pipeline	Cultural accuracy	Artistic style preservation	Overall visual quality
Swin transformer+Stable diffusion	9.5	9.7	9.6
Vision transformer+DDPM	9.0	9.2	9.1
U-Net+GAN-based inpainting	8.5	8.7	8.6
ResNet-50+PatchMatch	7.8	8.0	7.9

5. Experimental Results

The testing of different models for image restoration, especialy, as example, for preserving of old murals, gives possibily the results that express quality of degradation detection and inpainting. Each model has been measured on numerous performance indices, such as IoU, precision, recall, F1-score for the degradation detection, and PSNR, SSIM, FID for inpainting. During degradation detection the Swin transformer performs best by reaching an IoU score of 0.92 together with precision of 0.94, recall at 0.91 and an F1-score of 0.93. The model shows impressive detecting capabilities due to its IoU score of 0.92 together with outstanding precision score of 0.94 and recall score of 0.91 and F1-score score of 0.93 which indicate its potential role as a useful tool in early digital restoration workflows. The vision transformer achieves a set of evaluation metrics including an IoU at 0.89 and precision at 0.91 together with recall at 0.88 and an F1-score at 0.90. Even though less effective than the Swin transformer, the vision transformer is a promising design offers excellent processing of highly complex visual information which is crucial in maintaining the integrity of fine detail artwork. The U-Net model shows effectiveness in segmentation tasks based on its IoU result of 0.85 together with precision level of 0.87 and recall level of 0.84 and F1-score of 0.86. The provided accuracy measurements demonstrate an acceptable margin of error compared to three other models yet represent a minor drawback when it comes to preserving intricate murals. The recognition expert system ResNet-50 displays the least performance since it achieves an IoU of 0.82 and precision of 0.84 and recall of 0.81 alongside an F1-score of 0.83. ResNet-50 demonstrates usability in detecting historical mural degradation although additional modifications might be necessary to optimize its performance for this particular field.

Stable diffusion delivers outstanding results in inpainting evaluations because of its 32.5 dB PSNR rating plus 0.95 SSIM value and 12.3 FID score. The model demonstrates top-level image restoration capabilities through its metrics which prove its capacity to restore both intricate textures and original artistic tone together with historical accuracy. The DDPM model achieves excellent results comparable to stable diffusion through metrics that show a PSNR of 31.8 dB and SSIM of 0.94 accompanied by an FID score of 13.5. The restoration ability of this method provides strong performance to reconstruct art pieces that faithfully replicate original artistic designs. GAN-based inpainting demonstrates professional performance but produces substandard evaluation results at 30.2 dB PSNR along with 0.92 SSIM and 15.8 FID scores. The comparative analysis shows GAN-based inpainting has capabilities but its restoration quality does not reach the standards needed for artwork repair which demands precise execution and accurate rendering of small elements. The PSNR rating reached 28.7 dB while the SSIM value was 0.89 and FID measurement reached 18.4 with the distinct inpainting approach in PatchMatch. The outcome shows these models have restrictions in copying delicate details because their ability to duplicate precision is deficient compared to alternative models.

The combination of Swin transformer with stable diffusion shows outstanding performance as a degradation detection and inpainting system with an IoU value of 0.92 alongside PSNR at 32.5 dB and SSIM at 0.95 and FID at 12.3, signifying the most effective results. This relation not only performs faithfully in detecting the deteriorated region but it also performs effectively after restoration, by maintaining the structural integrity, as well as the aesthetic of historical murals. The evaluation of the different model combinations including vision transformer with stable diffusion along with U-Net with GAN-based inpainting produced important findings about their relative performance levels. The selection of appropriate restorative models becomes vital because each implementation provides unique capabilities and drawbacks during restoration efforts. The thorough assessment confirms that digital restoration technology has accomplished significant strides that can significantly influence the preservation of cultural assets. Each model and combination within it possesses distinct capacities to offer, when considering the decision is crucial for affording the restoration objectives with respect of and maintenance of the original artwork's cultural import.

6. Conclusions

In summary, a thorough comparison of many algorithms to detect degradation and inpainting shows that the combination of the Swin transformer and stable diffusion provides the strongest solution for a restoration of historical murals, also succeeding in the highest metrics over all: IoU (0.92), PSNR (32.5 dB), SSIM (0.95), and FID-score (12.3). This pipeline proves its ability to identify degraded areas accurately and restore them effectively thus protecting both artistic value and cultural worth of artworks. The restoration of high-stakes projects with tiny details would benefit from the superior precision and quality of the stable diffusion model since vision transformer and U-Net models fall short of delivering equivalent results in this domain. The current study faces two main drawbacks because the selected models work most efficiently with specific types of artwork degradation and due to their processing intensity they may not work feasibly across all situations. Future work should concentrate on improving those models to be more low resource friendly without affecting the performance as well as expanding them to attack to bigger array of artistic content and variety of degradation. This could imply the inclusion of more adaptable algorithms and looking into application with less resource-intensive networks which still it holds high accuracy and faithfulness in restoration tasks.

References

[1] Agarwal K. and Dixit M., "Scrupulous SCGAN Framework for Recognition of Restored Images with Caffe Based PCA Filtration," *The International Arab Journal of Information*

Technology, vol. 21, no. 1, pp. 107-116, 2024. DOI: 10.34028/iajit/21/1/10

[2] Cai X., Lu Q., Yao J., Liu Y., and Hu Y., "An Ancient Murals Inpainting Method Based on Bidirectional Feature Adaptation and Adversarial Generative Networks," *Advances in Computer Graphics*, vol. 14496, pp. 300-311, 2023. https://doi.org/10.1007/978-3-031-50072-5_24

[3] Cao J., Yan M., Jia Y., and et al., "Application of a Modified Inception-V3 Model in the Dynasty-Based Classification of Ancient Murals," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, no. 1, pp. 1-25, 2021. https://doi.org/10.1186/s13634-021-00740-8

[4] Choi Y., Murtala S., Jeong B., and Choi K., "Deep Learning-based Engraving Segmentation of 3-D Inscriptions Extracted from the Rough Surface of Ancient Stelae," *IEEE Access*, vol. 9, pp. 153199-153212, 2021. https://doi.org/10.1109/ACCESS.2021.3127229

[5] Gerasimiuk M., Shung D., Tong A., Stanley A., and et al., "MURAL: An Unsupervised Random Forest-based Embedding for Electronic Health Record Data," *in Proceedings of the International Conference on Big Data*, Orlando, pp. 4694-4704, 2021. https://doi.org/10.1109/BigData52589.2021.9672045

[6] Jiang X., Harun S., and Liu L., "Explainable Artificial Intelligence for Ancient Architecture and Lacquer Art," *Buildings*, vol. 13, no. 5, pp. 1-12, 2023. https://doi.org/10.3390/buildings13051213

[7] Lianji H. and Haihong Z., "Research and Application of Dunhuang Mural Image Restoration Technology Based on Deep Learning," *in Proceedings of the International Conference on Electronics and Devices, Computational Science*, Marseille, pp. 348-353, 2024. https://doi.org/10.1109/ICEDCS64328.2024.00068

[8] Nasri A. and Huang X., "Images Enhancement of Ancient Mural Painting of Bey's Palace Constantine, Algeria and Lacuna Extraction Using Mahalanobis Distance Classification Approach," *Sensors*, vol. 22, no. 17, pp. 1-20, 2022. https://doi.org/10.3390/s22176643

[9] Pawar P., Ainapure B., Rashid M., Ahmad N., and et al., "Deep Learning Approach for the Detection of Noise Type in Ancient Images," *Sustainability*, vol. 14, no. 18, pp. 1-19, 2022. https://doi.org/10.3390/su141811786

[10] Qian K., Zhang P., Huang R., Feng W., and Sun J., "Learning to Grade Deterioration for Ancient Murals," *Journal of Ambient Intelligence and Humanized Computing*, vol. 15, pp. 1727-1734, 2024. https://doi.org/10.1007/s12652-019-01487-9

[11] Sharma R. and Kukreja V., "Detecting the Past: Advancements in Comic Panel Detection for Cultural Heritage Preservation," *in Proceedings of the 4th International Conference on Data Analytics for Business and Industry*, Bahrain, pp. 529-532, 2023. https://doi.org/10.1109/ICDABI60145.2023.10629627

[12] Shi W. and Meng X., "Mural Restoration Research Based on Samples and Deep Learning," *in Proceedings of the IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems*, Bangalore, pp. 1-4, 2024. https://doi.org/10.1109/ICITEICS61368.2024.10625359

[13] Singh U., Maiti S., and Saini A., "Ancient Indian Murals Digital Restoration Through Image Inpainting," *in Proceedings of the 10th International Conference on Signal Processing and Integrated Networks*, Noida, pp. 635-640, 2023. https://doi.org/10.1109/SPIN57001.2023.10116111

[14] Sun P., Hou M., Lyu S., Wang W., and et al., "Enhancement and Restoration of Scratched Murals Based on Hyperspectral Imaging-A Case Study of Murals in the Baoguang Hall of Qutan Temple, Qinghai, China," *Sensors*, vol. 22, no. 24, pp. 1-19, 2022. https://doi.org/10.3390/s22249780

[15] Valdes J., Lopez C., and Lara A., "Detection of Deteriorated Sections in Murals Using Machine Learning," *in Proceedings of the IEEE 7th Congreso Internacional en Inteligencia Ambiental, Ingenieria de Software y Salud Electronica y Movil*, David, pp. 1-8, 2024. https://doi.org/10.1109/AmITIC62658.2024.10747620

[16] Xi W., Wang E., and Shen Z., "Bring Old Mural Images to Life," *in Proceedings of the IEEE 7th Information Technology and Mechatronics Engineering Conference*, Chongqing, pp. 933-937, 2023. https://doi.org/10.1109/ITOEC57671.2023.10291602

[17] Xiao H., Zheng H., and Meng Q., "Research on Deep Learning-Driven High-Resolution Image Restoration for Murals from the Perspective of Vision Sensing," *IEEE Access*, vol. 11, pp. 71472-71483, 2023. https://doi.org/10.1109/ACCESS.2023.3295253

[18] Xu W. and Fu Y., "Deep Learning Algorithm in Ancient Relics Image Colour Restoration Technology," *Multimedia Tools Applications*, vol. 82, pp. 23119-23150, 2023. https://doi.org/10.1007/s11042-022-14108-z

[19] Xu Z., Zhang X., Chen W., Liu J., and et al., "MuralDiff: Diffusion for Ancient Murals Restoration on Large-Scale Pre-Training," *IEEE Transactions on Emerging Topics in*

Computational Intelligence, vol. 8, no. 3, pp. 2169-2181, 2024. https://doi.org/10.1109/TETCI.2024.3359038

[20] Yadav A., Sharma S., Sangwan D., and Kumar R., "Revitalizing Ancient Murals in the Shekhawati Region Through Image Inpainting Techniques," *in Proceedings of the 11th International Conference on Signal Processing and Integrated Networks*, Noida, pp. 361-366, 2024. https://doi.org/10.1109/SPIN60856.2024.10511302

[21] Yu Z., Lyu S., Hou M., Sun Y., and Li L., "A New Method for Extracting Refined Sketches of Ancient Murals," *Sensors*, vol. 24, no. 7, pp. 1-16, 2024. https://doi.org/10.3390/s24072213

Ashwini Ramachandran is a research scholar in the Electronics and Communication Engineering Department of the SRM Institute of Science and Technology, Chennai. She has a B.E degree in Electronics and Communication Engineering (2007) from Anna University, Chennai, and a Master's degree in Applied Electronics (2011) from Sathyabama University, Chennai. Her research interest includes Image Processing, Deep Learning and Computer Vision.

Vani Rajamanickam is a Professor in the Department of ECE at SRMIST, Ramapuram. She Graduated in Electronics and Communication Engineering at Bharathiyar University, Coimbatore, Tamilnadu, India. She secured Master of Engineering in Communication Systems at Anna University, Chennai, India. She Pursued Ph.D. in the field of Image Processing at Anna University, Chennai, India. She is in teaching profession for more than 20 years Her area of interest includes Image Processing, Deep Learning, IoT, and Medical Data Analysis.