Damage Segmentation and Restoration of Ancient Wall Paintings for Preserving Cultural Heritage

Hardaat Singh Baath[1], Soham Shinde[1],Jinam Keniya[1],Priyanshu Ranjan Mishra[1]
Anil Saini[2,3], Dhiraj[2,3] *
[1] Birla Institute of Technology & Science, Pilani
f20212662@goa.bits-pilani.ac.in
f20212927@goa.bits-pilani.ac.in
f20211622@goa.bits-pilani.ac.in
f20212259@hyderabad.bits-pilani.ac.in
[2] Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, India
[3] CSIR-Central Electronics Engineering Research Institute, Pilani, India
anil@ceeri.res.in
*dhiraj@ceeri.res.in

Abstract. Rajasthani paintings are unique and intriguing art forms representing India's rich cultural heritage. However, these ancient master pieces are deteriorating due to the passage of time, environmental factors, and human actions. Preserving and Restoring these delicate artworks is crucial. One approach to aid their digital restoration is leveraging advanced technologies like deep learning. This study applies image segmentation and restoration techniques to restore the Rajasthani murals in the Mandawa region of rural rajasthan, India. The main objective is to segment the damaged murals, generate their corresponding binary masks and restore the corresponding areas of the damaged image. The research aims to achieve robust and accurate predicted masks for the murals by utilizing state-of-the-art deep learning models and using their outputs as inputs for image restoration as the final restored output image. Extensive comparisons with different segmentation models show that the proposed approach outperforms the rest with an mIOU of 0.892. The proposed method also demonstrates remarkable inpainting results with an SSIM score of 0.9812 on test images. Results show that the method achieves promising restoration of damaged ancient Indian Wall Paintings.

Keywords: Image segmentation, Ancient Indian murals, Restoration, Image Inpainting, Ensemble approach, STAPLE.

1 Introduction

Rajasthani murals, a significant part of India's cultural heritage, depict vibrant narratives and are known for their exquisite craftsmanship. However, they have suffered various damages over time, including color fading, cracks, peeling, and vandalism. Restoring these murals poses unique challenges due to their complexity and fragility, requiring specialized techniques and expertise. Traditional restoration methods are time consuming, subjective, and may not capture the original nuances effectively.

Deep learning models have emerged as powerful tools in computer vision, offering promising solutions for mural restoration. Multiple deep learning models such as U-Net++ [1], DeepLabV3+ [2], PSPNet [3], and FPN [4] were used to accurately and comprehensively segment the murals. By combining the predictions of these models, segmentation accuracy was enhanced and addressed the challenges of damaged murals. An ensemble approach using the STAPLE [5] algorithm has been proposed to fuse the predictions from multiple models, resulting in a more robust and accurate segmentation outcome. Cluster wise training of PConvNet [6] for inpainting, specifically designed to handle the diverse patterns observed in Rajasthani murals, is proposed. To evaluate the performance of the proposed approach, the dataset was curated from the Havelis in the Mandawa region of rural Rajasthan, which serves as expansive galleries of Rajasthani art. The dataset comprises high resolution images capturing intricate details, style variations, and the observed damage of the murals.

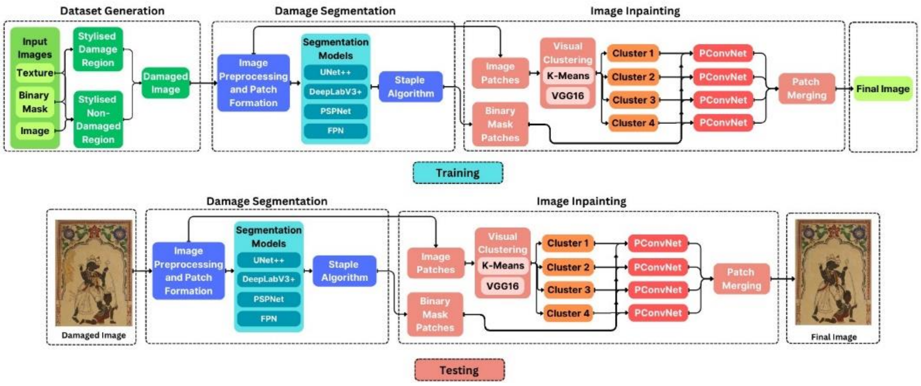

Fig. 1. Complete Pipeline of Proposed Work.

The proposed work focuses on the development of pipelines to enhance the training and performance of models in the context of damaged image processing as shown in Figure 1:

– Firstly, a novel pipeline is presented for creating accurate synthetic damaged images from healthy image samples. Training models on these artificially generated damaged images can improve performance in handling real world damaged images.

– Secondly, a robust, non-redundant, and accurate damage detection and segmentation pipeline is introduced. By incorporating multiple models and algorithms, the reliability and effectiveness of the segmentation process are enhanced, ensuring its applicability across diverse scenarios.

– Lastly, an inpainting pipeline that utilizes edge based clustering to enhance the accuracy and specialization of the missing region restoration process is presented. This approach enables the generation of visually coherent and semantically consistent inpainted images, contributing to the overall improvement of the image restoration process.

2 Related Works

Segmentation and restoration of paintings are an active research field. In the following sections, recent development in the field is discussed.

Conventional Methods: Traditional methods for picture segmentation include unsupervised and semi-supervised machine learning methods such as Region Growing [7], which begins with a seed point or region and iteratively expands by adding neighboring pixels that fulfil predefined similarity criteria. It gradually builds coherent regions based on local similarity metrics, allowing picture regions with comparable properties to be segmented.
Watershed Transform [8], a region based segmentation algorithm that considers picture pixels as a topographic surface, segments regions by filling basins with water, is another frequently used method.

Deep Learning Techniques: Semantic image segmentation is a widely used technique in which an image is pixel wise labeled into its constituent categories. It involves an encoder decoder structure, where the encoder captures high level features of the input image, and the decoder up samples and refines these features to produce a segmentation map. Popular approaches involve training models like U-Net with binary masks [9,10] or modifying existing models to suit specific requirements. For example, attention mechanisms can be integrated to selectively emphasize relevant features like cracks, during segmentation, improving the model's ability to capture essential information. Additionally, specialized masks, such as dusk like and jelly like masks [11], are used for texture synthesis testing and restoration in e-Heritage conservation, representing different types of damage.
Advanced methods combine image processing and artificial intelligence based methods such as such as Delaunay triangulation based interpolation and exemplar based inpainting [12]. These techniques drive innovation in the field and enhance the digital restoration of historical artefacts by improving the restoration of damaged areas.

3 Methodology

This section describes the pipeline used for the image segmentation process. A dataset of 488 high resolution clean images (6000x6000 pixels) was carefully collected for effective model training. These images were sourced from the ancient murals in the Havelis of the Mandawa region in rural Rajasthan. Additionally, damaged images and their corresponding binary masks were included in the dataset to facilitate the training process. The following sections will present further details on the data collection and preparation process, providing a comprehensive overview of the methodology employed.

3.1 Dataset Generation

Creating an adequate dataset is crucial for training and evaluating damage segmentation algorithms. This section focuses on two key aspects: masks which depict the regions of damage and textures which define the appearance of damage, and using them with clean images to create a synthesized damaged image. These processes were vital in generating the necessary data to train and test the proposed damage segmentation framework.

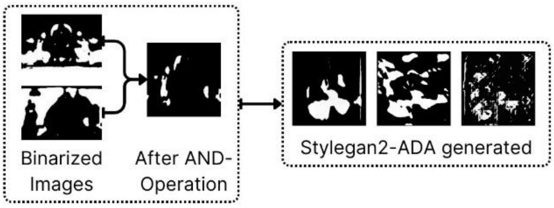

Fig. 2. Examples of masks generated.

Mask Generation: The mask generation process involves applying pixel wise AND operations on the initial mask set created using binarization of random images, creating random white patches resembling damage patterns. From these images, 897 binary masks were crafted. These binary masks were used as input for the Stylegan2-ADA [13] framework to generate the final set of binary masks, as shown in Figure 2. In total, 5,000 masks were generated for further processing, of which 90% was used as the training set and 10% was used as the test set.

Texture Generation: A total of 875 texture images were collected to generate the final textures. This collection included 205 diverse texture images from internet stock repositories, 220 AI-generated images for unique textures, and 450 images extracted from damaged sections of real murals for authentic and relevant samples. These texture images were then used as input data for the Stylegan2-ADA [13] framework to generate customized textures as shown in Figure 3.

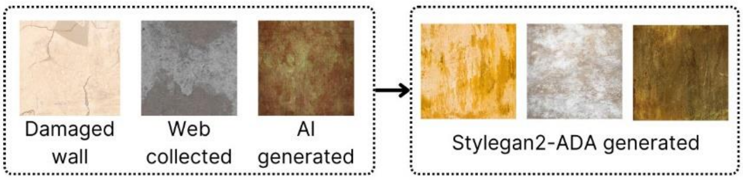

Fig. 3. Example of textures generated.

Damage Image Creation: Originally captured clean images (6000x6000 pixels) were divided into smaller grids (512x512 pixels) to augment the in-house dataset. To imitate real damage on clean images, dataset preparation pipeline includes the following steps: The following steps outline the pipeline as shown in Figure 4:

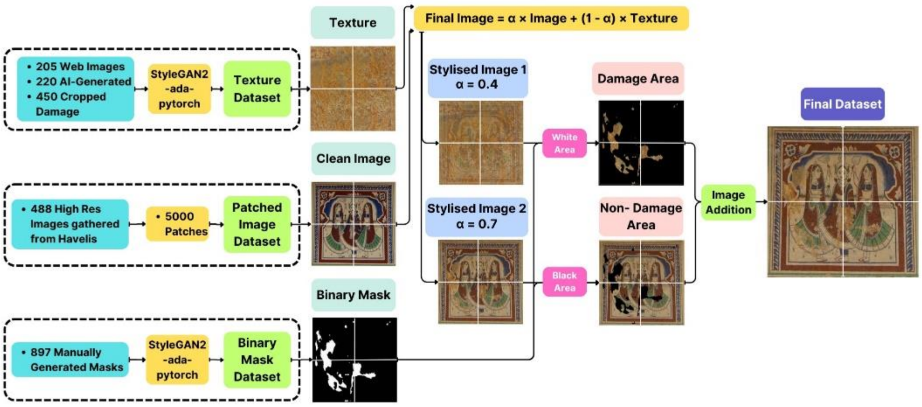

Fig. 4. Complete pipeline of Damage Generation

– **Stylized Image Generation:** Two images with artistic filters or transformations were generated using clean images and a texture source. One image represents damaged patches, while the other represents non-damaged parts.

– **Segmentation and Separation:** By overlaying a binary mask, damaged and non-damaged regions in both stylized images were segmented and separated, ensuring an accurate representation of each area.

– **Result Combination:** Finally, the results were merged from the damaged and non-damaged stylized images, creating a synthesized damaged painting that mimics real instances of damaged artwork. This process captures the characteristics and appearance of actual damaged images.

Using these steps, 5,000 synthesized damaged paintings were created a dataset that accurately represent the desired characteristics for the research. These additional samples were used to train the models further.

3.2 Damage Segmentation

Next step in the pipeline is segmenting damaged regions from the damaged images and generate a binary mask using the proposed method. The steps followed is discussed in the subsequent sections.

Image Preprocessing: Before segmentation, images underwent crucial pre-processing. This included enhancing color saturation, brightness, and sharpness while introducing controlled noise levels as shown in Figure 5. These improvements made the images visually appealing, representative of the scene, and captured finer details, thus facilitating better performance of the segmentation model. Controlled noise levels introduced variations that mimicked real world scenarios. The pre-processing workflow improved the dataset and segmentation model performance. Images were finally divided into 512x512 patches to preserve details during segmentation and inpainting.

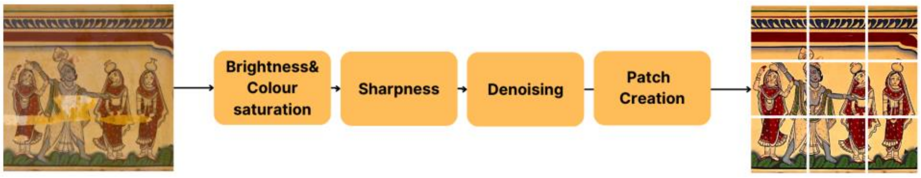

Fig. 5. Steps for data preprocessing.

Segmentation Models: The Segmentation model divides an input image into meaningful regions or objects and labels pixels or areas based on their category. In this study, four segmentation models were employed from Segmentation Models which are U-Net++ [1], DeepLabV3+ [2], PSP Net [3] and FPN[4], and transfer learning was applied to harness the capabilities of pre-trained models by adapting them to new tasks or datasets.

– **UNet++:** UNet++ [1] extends UNet with nested and dense skip connections, improving segmentation performance. It captures multiscale information and promotes feature fusion.

– **DeepLabV3+:** It is an advanced segmentation model that enhances accuracy with dilated convolutions and atrous spatial pyramid pooling. It uses an encoder decoder architecture with skip connections to capture multiscale information and achieve precise delineation.

– **PSPNet:** It is a powerful image segmentation model that captures rich contextual information at different scales. It improves scene understanding and segmentation accuracy by incorporating comprehensive contextual details.

– **FPN:** FPN (Feature Pyramid Network) [4] is an image segmentation model that leverages feature pyramids to incorporate contextual information and enhance segmentation accuracy. By integrating it, multiscale features can be extracted, and precise segmentation of complex visual scenes can be achieved.

Loss Functions: In our scenario, the objective is to segment damaged areas within an image; two specific loss functions were employed: – ***Cross-Entropy Loss (LCE), Dice Coefficient Loss (LDC)***.

Final Loss: The same final loss function is used for all the models for ease of implementation. The loss term weights were determined by performing a hyper parameter search for the custom developed training dataset.

$$L_{final} = 0.4L_{CE} + 0.6L_{DC}$$

STAPLE Algorithm: STAPLE [5] (Simultaneous Truth and Performance Level Estimation) [5] is an unsupervised learning expectation maximization algorithm, it computes a probabilistic estimate of the accurate segmentation and measures the performance level of each segmentation in a collection. The algorithm considers segmentation masks combining them optimally to form a probabilistic estimate of the accurate segmentation.

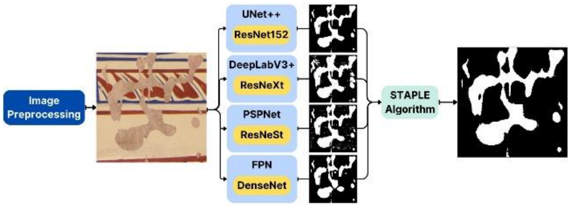

Fig. 6. The collected masks undergo further refinement using the STAPLE algorithm.

Implementation (Segmentation): Unet++[1], DeeplabV3+[2], PSPNet[3], and FPN[4] were selected as the preferred models to detect and segment damaged regions in wall paintings. Paired with ResNet152[14], ResNext10132x8d[15], timm-ResNeSt269e[16], and DenseNet161[17] encoders, these models offer deep and efficient architectures capable of effectively capturing intricate visual patterns. Pre-trained weights from Imagenet[18] were utilized for transfer learning, leveraging their learned feature representations.
During training, damaged images and corresponding binary masks were used. The Adam optimizer, known for its versatility and efficiency, was employed, along with custom loss functions combining Cross Entropy Loss and Dice Loss. The training was performed on patches of size 512 x 512, with 100 epochs and a batch size of 8, adjusted as needed for different models. The STAPLE [5] algorithm, operating as an unsupervised procedure, iteratively refined the segmentation masks for improved accuracy.
The damaged image was fed into the four models in the testing phase, resulting in segmentation masks. These masks were processed through the STAPLE [5] algorithm to obtain a refined and improved final mask as shown in Figure 6. The refined mask, along with the original image, was used for the inpainting procedure, allowing for intelligent restoration of the damaged regions based on the precise information provided by the refined mask

3.3 Image Inpainting

For inpainting., the PConvNet [6] was utilized as the preferred model. PConvNet [6] has shown strong performance in inpainting tasks involving irregular and variable damage patterns, making it highly suitable for the in-house dataset and research objectives. By integrating PConvNet [6] into the framework as shown in Figure 7, our goal is to leverage its inpainting capabilities to restore the damaged regions of the murals based on the segmentation masks generated by the damage segmentation model. In the proposed approach, multiple PConvNet [6] models were trained separately on different clusters of datasets, where clusters are formed based on the type of patterns and style present in each patch. Further details on the clustering process are discussed in the subsequent section.

Visual Clustering: In the visual clustering step, the proposed approach uses VGG16[19] for feature extraction and K-Means clustering to group similar patches.

K-Means [20] clustering partitions the dataset into distinct clusters, K=4 yielding the best results. These clusters inform the subsequent inpainting process, ensuring tailored treatment based on their specific characteristics. This approach enhances accuracy and produces realistic results in image inpainting. The clustered patches are then passed to the respective PConvNet [6] models for inpainting.

PConvNet: It applies partial convolution [6] on the damaged image, i.e., it only performs convolution operation on the non-missing pixels to get the contextual information of the image, which it later uses to in-paint damaged sections of the image.

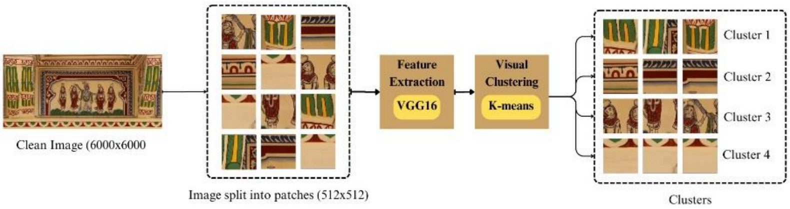

Fig. 7. Result of Visual Clustering of Patches (For k equals 4).

Loss functions: Four loss functions were utilized: Pixel Reconstruction Loss(L_{valid}, L_{hole}), Perceptual Loss($L_{perceptual}$), Style Loss(L_{style}), Total Variation Loss(L_{tv}). The total loss as adapted from [6] is the combination of all the above loss functions and is shown below:

$$L_{total} = L_{valid} + 6L_{hole} + 0.05L_{perceptual} + 120(L_{style_{out}} + L_{style\ comp}) + 0.1\ L_{tv}$$

Implementation (Inpainting): After segmenting the damaged patches, the image is divided into smaller patches of size 512 x 512 for inpainting. This ensures the preservation of finer details during the inpainting procedure. To extract subtle features, such as lines and repeating patterns, the VGG16 model is employed. The extracted features are the basis for clustering the patches into four distinct clusters using the K-Means algorithm. This clustering approach allows the inpainting model to specialize in recovering damages specific to each cluster. The PConvNet [6] model, known for its ability to address irregular and variable damage patterns, is adopted for the inpainting task. Four individual PConvNet [6] models are trained, each focused on recovering damages associated with its respective cluster.

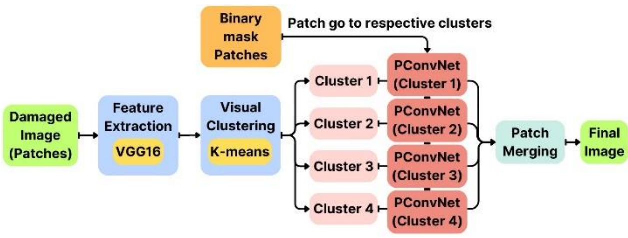

Fig. 8. Inpainting pipeline.

This cluster specific approach results in exceptional accuracy in the inpainting process. The restored patches are seamlessly stitched together, generating the final restored image with a resolution of 6000x6000 pixels as shown in Figure 8.

4 Results

The algorithm development for the complete pipeline is performed in Python on an ubuntu machine with i7 processor, 32GB RAM, and Nvidia GTX 1080Ti GPU. The proposed method represents a breakthrough in artificial dataset generation in the form of synthetic textures and masks for simulating the damaged regions from healthy images. As presented in the methodology proposed in Figure 2, diverse and realistic binary masks were generated (Comparison shown in Figure 9), including jelly like masks [13] resembling authentic damage patterns. The generated masks and their quality in terms of different performance indices have been compared in Table 1, with earlier published results to mark their effectiveness in inpainting techniques. This innovation enhances the dataset's comprehensiveness and accuracy. Integrating the alpha factor and generated textures, as shown in Figure 4, is crucial in replicating realistic damage with fidelity, closely mimicking actual damage scenarios. To evaluate and assess the effectiveness of the proposed methodology for generating damage segmented masks, a comparison was conducted with other existing methods.

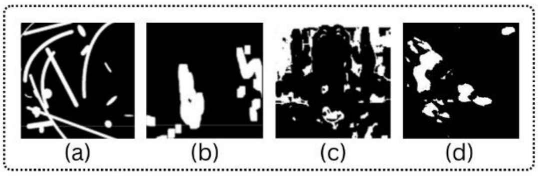

Fig. 9. Comparison: (a) Dynamic (b) Gated Convolution (c) cGAN [21] (d) Proposed

Table 1. Comparison of Mean Intersection over Union (mIOU) scores for different segmentation models.

Segmentation Model	**mIOU**
UNet++[1]	0.851
DeepLabV3+[2]	0.873
PSPNet[3]	0.820
FPN[4]	0.794
Proposed Approach	0.892

The proposed methodology for generating damage segmented masks achieved a mIoU (mean Intersection over Union) value of 0.892 as shown in Table 1, surpassing all other methods in the comparison. The mIoU value was calculated using the 500 images of the test dataset. This indicates remarkable efficiency and exceptional performance in accurately identifying damaged regions. The proposed approach's high accuracy and consistency make it suitable for practical applications where accurate damage identification is crucial. The generated inpainted regions seamlessly integrate with the surrounding undamaged areas while preserving the artistic attributes of the original mural. The proposed method demonstrates impressive performance with an SSIM score of 0.9812 when tested with 500 images of the test dataset.

Fig. 10. (Left to Right) Damaged Image, Damage Segmentation, Inpainting Result

As depicted in Figure 10, the end to end pipeline excels in dataset generation, damage segmentation, and image inpainting, representing a significant advancement in mural restoration.

5 Conclusion

Art forms always played an essential role in human history. The murals present on walls of the old Havelis of Rajasthan represent a significant vibrant form of open art which are now in a deserted state. The proposed work presents a framework for the digital restoration of this art form through deep learning-based methods. The framework consists of an array of modules comprising synthetic damaged data generation, visual clustering, and inpainting as a whole. The proposed method advances the field by providing a more sophisticated and effective approach to artificial dataset generation. The utilization of diverse binary masks and the incorporation of the alpha factor and generated textures contribute to a higher level of realism in the simulated damage. Incorporating results of various segmentation models using the STAPLE algorithm ensures the robustness and accuracy of damage segmentation. Clustering the images and training PConvNet separately on each cluster helps us get tailored results for each cluster, ensuring better restoration. The future work includes the creation of a VR-based 3D model for walk-throughs and storytelling by using the restored mural images for presenting an enhanced 3D version of the Havelis targeting heritage preservation for coming generations.

6 Acknowledgement

The authors gratefully acknowledge the support of the Department of Science and Technology(DST), Government of India, for funding this project under grant No: DST/TDT/SHRI/48/2021.

References

1. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang," UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," in IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1856-1867, June 2020, doi: 10.1109/TMI.2019.2959609.
2. Chen, Liang-Chieh, et al." Encoder-decoder with atrous separable convolution for semantic image segmentation." Proceedings of the European conference on computer vision (ECCV). 2018.
3. Zhao, Hengshuang, et al." Pyramid scene parsing network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
4. Lin, Tsung-Yi, et al." Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
5. S. K. Warfield, K. H. Zou and W. M. Wells," Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," in IEEE Transactions on Medical Imaging, vol. 23, no. 7, pp. 903-921, July 2004, doi: 10.1109/TMI.2004.828354.
6. Liu, G., Reda, F.A., Shih, K.J., Wang, TC., Tao, A., Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11215. Springer, Cham. https://doi.org/10.1007/978-3-030-01252-6_6.
7. S. A. Hojjatoleslami and J. Kittler," Region growing: a new approach," in IEEE Transactions on Image Processing, vol. 7, no. 7, pp. 1079-1084, July 1998, Doi: 10.1109/83.701170.
8. Roerdink, Jos BTM, and Arnold Meijster." The watershed transform: Definitions, algorithms and parallelization strategies." Fundamenta informaticae 41.1-2 (2000): 187-228.
9. Al-shaikhli, Saif & Bahrani, Ahmed. (2010). Segmentation of Tumor Tissue in Gray Medical Images Using Watershed Transformation Method. Int. J. Adv. Comp. Techn.. 2. 123-127.
10. U. Singh, S. Maiti, A. Saini and Dhiraj," Ancient Indian Murals Digital Restoration through Image InPainting," 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2023, pp. 635-640, Doi: 10.1109/SPIN57001.2023.10116111.
11. Yu, Tianxiu, et al." End-to-end partial convolutions neural networks for Dunhuang grottoes wall-painting restoration." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
12. Mol, V. Rakhi, and P. Uma Maheswari." The digital reconstruction of degraded ancient temple murals using dynamic mask generation and an extended exemplar based region-filling algorithm." Heritage Science 9 (2021): 1-18.

13. Karras, Tero, et al.” Training generative adversarial networks with limited data.” Advances in neural information processing systems 33 (2020): 12104-12114.
14. K. He, X. Zhang, S. Ren and J. Sun,” Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, Doi: 10.1109/CVPR.2016.90
15. S. Xie, R. Girshick, P. Doll´ar, Z. Tu and K. He,” Aggregated Residual Transformations for Deep Neural Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5987-5995, Doi:10.1109/CVPR.2017.634.
16. Ross Wightman, Hugo Touvron, Herv´e J´egou: ResNet strikes back: An improved training procedure in timm. CoRR abs/2110.00476 (2021)
17. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger,” Densely Connected Convolutional Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, Doi:10.1109/CVPR.2017.243.
18. J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei,” ImageNet: A large-scale hierarchical image database,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, Doi: 10.1109/CVPR.2009.5206848.
19. H. Qassim, A. Verma and D. Feinzimer,” Compressed residual-VGG16 CNN model for big data places image recognition,” 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2018, pp. 169-175, Doi: 10.1109/CCWC.2018.8301729.
20. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100–108. https://doi.org/10.2307/2346830
21. Rakhimol, V., and P. Uma Maheswari. “Restoration of Ancient Temple Murals Using cGAN and PConv Networks.” Computers & Graphics, vol. 109, 2022, pp. 100–110, https://doi.org/10.1016/j.cag.2022.11.001.