

High-Resolution U-Net Restoration of Degraded Tamil Palm-Leaf Manuscripts

Aravindhana R¹, Dr. P. Durgadevi²

¹M.Tech Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamil Nadu, India

²Assistant Professor (Senior Grade), Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamil Nadu, India

Abstract

Palm-leaf manuscripts are a primary carrier of Indic literary and scientific heritage, yet ink erosion, worm holes, and missing character blocks make them difficult to digitise. This paper presents a restoration-centric pipeline for Tamil palm-leaf images. Clean scans of 262 leaves from the Naladiyar, Tholkappiyam, and Thirukadugam collections are paired with synthetically damaged counterparts generated by documented OpenCV operators. A 31.04-million-parameter U-Net is trained at a resolution of 384 by 1280 with a hybrid loss of L1 plus 0.5 times the mean squared error, for 30 epochs, on an AMD Instinct MI210 accelerator. The best checkpoint reaches a validation loss of 0.0314 in 28.04 minutes, with a peak allocation of 42.54 GB of graphics memory. Qualitative inference recovers stroke continuity and inverts simulated holes and missing blocks while preserving leaf geometry. A Tesseract Tamil baseline on the printed Mozhi corpus yields a mean character error rate of approximately 0.34 and is treated as a transcription control, not as a palm-leaf reading score. Translation via IndicTrans2 is specified as a planned, not-yet-trained module. The implemented contribution of this work is supervised high-resolution restoration under an explicit, reproducible protocol.

Keywords: Palm-Leaf Manuscripts, Document Restoration, U-Net, Synthetic Degradation, Tamil Heritage, Computer Vision, Cultural Heritage Digitisation

1. Introduction

Palm-leaf manuscripts (olaichuvadi) remain among the oldest portable writing media in South and Southeast Asia. In the Tamil tradition they preserve grammar, ethics, and medicine, yet the physical leaves degrade faster than they can be transcribed by hand: ink fades, fibres crack, and cord holes enlarge into lacunae. Large-scale photography produces images that are still hard to read. Pixel-level restoration is therefore a useful first stage in any digitisation pipeline that later attempts transcription or translation.

Classical enhancement, including median filtering, morphological closing, and global binarisation, helps on uniform noise but fails when the damage is semantic. A worm hole is not Gaussian noise, and a missing akshara cannot be inpainted by a local filter that has no model of script shape. Learning-based restorers treat the problem as image-to-image translation [1, 2, 3]. A generative adversarial network (GAN) can produce sharp pages, but it can also invent high-frequency texture. For heritage leaves, invented strokes are dangerous, because they may look like characters to a later reader [1].

This work takes a conservative position. Restoration is trained as a paired, supervised mapping from a synthetically damaged leaf to its known clean original, using an encoder-decoder U-Net with skip connections [3] and an L1 and mean squared error (MSE) objective that penalises deviation from the ground-truth pixels rather than encouraging adversarial sharpness. The canvas is 384 by 1280 pixels, which respects the long aspect ratio of a palm leaf instead of collapsing it to a square. Training uses an institutional AMD Instinct MI210 graphics processing unit (GPU) so that this geometry can be retained at a batch size of 16.

The official project title also names transcription and translation. Those modules are not claimed as completed contributions. A Tesseract Tamil baseline on Mozhi printed pairs [4] and a character-component visualisation on restored leaves are reported as a partial proof of concept. IndicTrans2 is the planned offline translation backend [5]. The hypothesis of this paper is narrower: a high-resolution supervised U-Net, trained on documented synthetic pairs, can invert simulated palm-leaf damage well enough to serve as a reproducible restoration front-end for later reading modules.

The contributions of this work are:

- a paired Tamil palm-leaf restoration dataset of 262 leaves with three documented damage operators;
- a 31.04-million-parameter U-Net trained at 384 by 1280 pixels, with logged epoch metrics;
- a quantitative validation loss of 0.0314, together with qualitative galleries; and
- an explicit interface to future optical character recognition, handwritten text recognition, and neural machine translation, aligned with United Nations Sustainable Development Goal 11, Target 11.4, on safeguarding cultural heritage [6].

2. Related Work

2.1 Document Restoration

Souibgui and Kessentini (2020) proposed DE-GAN, a conditional generative adversarial network for document enhancement that produces strong visual cleanup of dirty pages [1]. Adversarial losses encourage sharpness, which is useful for binarised Western documents but risky when the output may be treated as evidence of a glyph. Yang et al. (2025) studied appearance prediction for damaged historical documents and showed that restoration can be examined without immediate optical character recognition (OCR) [2]. Paired degrade-to-clean training, including high-dynamic-range (HDR) style document restoration, supplies the supervision pattern used here [7]. The U-Net remains a robust baseline for dense prediction because skip connections copy high-resolution features from the encoder to the decoder, preserving thin boundaries [3]. This paper prefers an L1 and MSE objective over an unconstrained GAN as the first implemented restorer. Adversarial or diffusion alternatives can be compared later under an OCR-linked metric.

2.2 Historical and Palm-Leaf Datasets

Nikolaidou et al. (2022) surveyed historical document image datasets and showed that many public sets are European, binarised, or layout-oriented [8]. Indic palm-leaf resources exist but are uneven in labelling. The Tamil Handwritten Palm Leaf Manuscript Dataset provides Tamil handwritten palm-leaf images without pixel-aligned clean counterparts for restoration [9]. PALM-LAY is a multi-script layout corpus with region annotations, including a Tamil subset [10]. It is complementary for future line detection, and it is not sufficient to train a pixel restorer. Southeast Asian handwritten-text sets such as Sleukrith, for Khmer [11], and AMADI LontarSet, for Balinese [12], show that competition-grade leaf

reading is possible when expert transcription exists. They are cited as methodological neighbours, not as training data, because they are not Tamil.

2.3 Indic OCR and Translation

Multimodal large language models have been argued to help historical handwriting recognition when prompted with visual context [13]. ConvNeXt is a modern convolutional backbone for character-level classifiers [14]. Printed Indic OCR corpora such as Mozhi enable character error rate (CER) evaluation [4]. Tesseract with Tamil language data is an open, reproducible baseline that is weak on historical damage, which is precisely why restoration is a prerequisite [15]. Pracalit manuscript pages exported from Transkribus include expert Extensible Markup Language (XML) transcriptions [16]. IndicTrans2, from AI4Bharat, is a high-quality offline translator among Indian languages and English [5]. The FLORES-200 set provides a standard machine-translation evaluation benchmark [17]. These systems are surveyed so that restoration is correctly positioned as the upstream module that should eventually improve CER and translation quality. Those metrics are not claimed as restoration results.

3. Dataset and Synthetic Degradation

Clean red-green-blue (RGB) scans of 262 Tamil palm leaves were collected from three named works: Naladiyar, Tholkappiyam, and Thirikadugam. Each leaf is a long, narrow strip with punched binding holes and several lines of handwritten script. No public corpus supplies an ancient Tamil image, an aligned clean counterpart, and a transcription together, so paired supervision is constructed rather than assumed.

Synthetic damage is generated with operators from the Open Source Computer Vision Library (OpenCV) that mimic three physically motivated processes [18]:

- Ink erosion: morphological erosion and local intensity attenuation that thins and breaks strokes.
- Missing blocks: rectangular regions replaced by an estimated background colour, standing in for torn or overwritten akshara groups.
- Paper damage: elliptical holes that mimic worming and enlargement of the cord hole.

Each damaged file is paired with its clean original by filename. The split is 85 percent training and 15 percent validation at the leaf level, yielding 223 training pairs and 39 validation pairs. Images are resized to 384 by 1280 RGB tensors. A separate folder of 199 binarised pages is retained for a future line-segmentation experiment and is not used in U-Net training.

This construction has a known domain gap. Simulated holes have sharper edges than museum damage, and real unpaired degradation remains a measurement for later work. The operators are documented so that the gap can be audited rather than hidden. Figure 1 shows the four states of one leaf, from the clean original through ink erosion and missing blocks to paper damage.

Figure 1: Synthetic Damage Operators Applied to a Clean Tamil Palm-Leaf Scan.

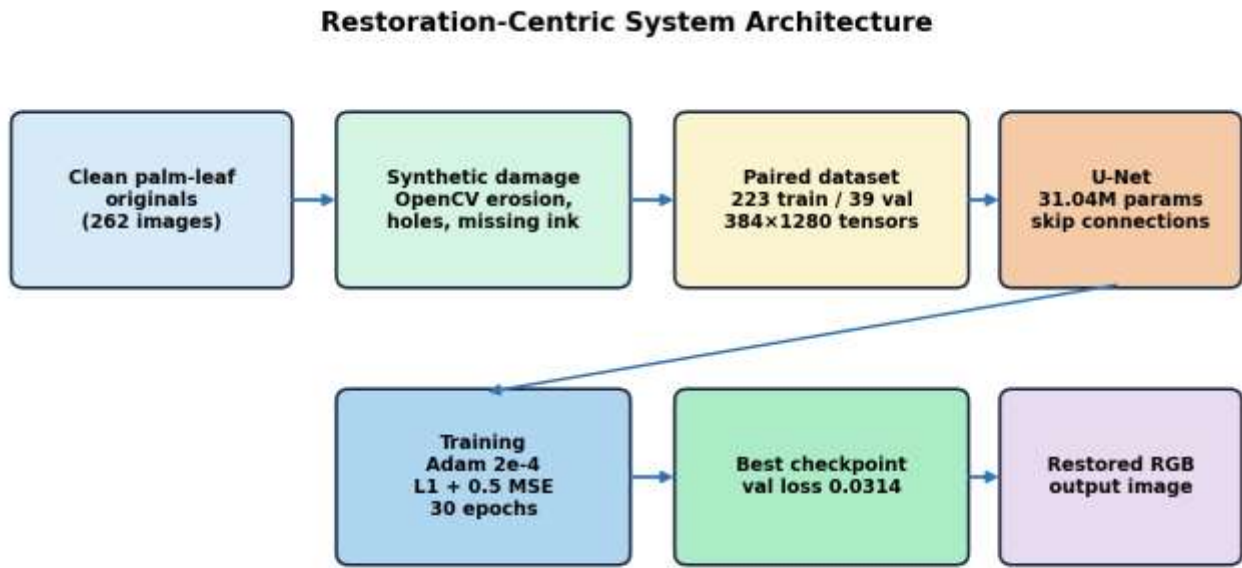


4. Proposed Restoration Pipeline

4.1 System Overview

Figure 2 summarises the implemented pipeline. Clean originals are damaged, packed as pairs, and consumed by a U-Net. Training logs epoch loss, wall-clock time, and peak graphics memory (VRAM). The checkpoint with the lowest validation loss is kept and used for RGB inference. Transcription and translation are reserved as downstream modules. They consume restored pixels, through OCR or handwritten text recognition (HTR), or Unicode text, through neural machine translation (NMT). They are not stages inside the restoration loss.

Figure 2: Restoration-Centric System Architecture.



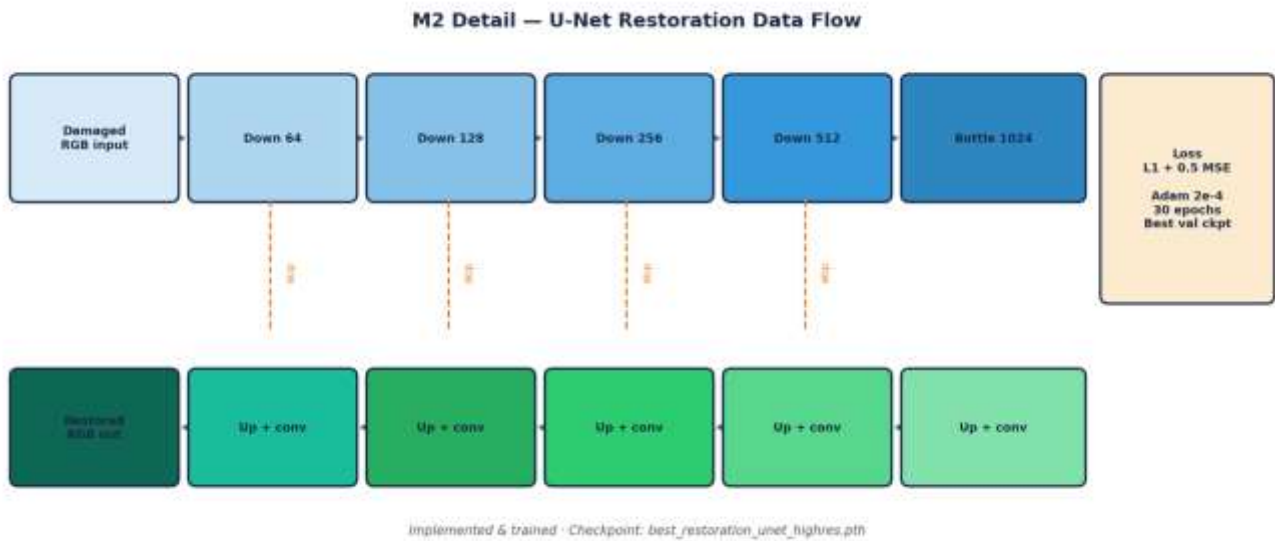
Future modules (not implemented in this report): transcription (OCR/HTR) and translation (NMT).

4.2 U-Net Architecture

The restorer is a classical U-Net adapted to RGB inpainting, as drawn in Figure 3 [3]. The encoder applies four down-sampling stages after an initial double convolution, with channel counts of 64, 128, 256, 512, and 1024 at the bottleneck. Each double-convolution block is two convolutions of size 3 by 3, with batch normalisation and a rectified linear unit (ReLU). The decoder uses transposed convolutions with a kernel size of 2 and a stride of 2. It concatenates the corresponding encoder feature maps through skip connections, and then applies a matching double convolution. A convolution of size 1 by 1, followed by a sigmoid, produces a 3-channel RGB image with values in the range 0 to 1. Skip connections are the reason thin ink strokes survive. They copy high-resolution spatial detail around the bottleneck rather than forcing the decoder to hallucinate it.

The network has 31.04 million parameters. Residual or attention U-Nets could be substituted later. They were not required to obtain the first working checkpoint.

Figure 3: U-Net Encoder-Decoder with Skip Connections Used for Palm-Leaf Restoration.



4.3 Loss and Optimisation

Let x be the damaged tensor, y the clean target, and $\hat{y}$ the restoration produced by the network. The training objective is given in Equation (1).

$$L = \|\hat{y} - y\|_1 + 0.5 \times \|\hat{y} - y\|_2^2 \quad (1)$$

In Equation (1), $\hat{y}$ is the restored image and y is the clean target. The first term is the L1 distance. The second term is 0.5 times the squared L2 distance, which is the MSE term. The L1 term penalises absolute deviation and preserves median intensity. The MSE term more strongly penalises large residual holes. An adversarial term is omitted in order to reduce the risk of invented glyphs. The optimiser is Adam, with a learning rate of 0.0002 [19]. No cosine schedule and no additional regulariser are used, so the run can be reproduced from Table 1 alone.

5. Experimental Setup

Training was performed on a single AMD Instinct MI210 GPU with 64 GB of high-bandwidth memory, under Radeon Open Compute (ROCm) 6.1 and PyTorch 2.6 [20, 21]. The host is a server in the AMD EPYC 9634 class. For reproducibility on this accelerator, cuDNN-style autotuning was disabled, and a single data-loading worker was used without pinned memory. That choice avoided MIOpen crashes observed in earlier trials.

Table 1: Training Hyper-Parameters

Hyper-Parameter	Value
Input Resolution	384×1280 RGB
Batch Size	16
Epochs	30
Training Pairs	223

Validation Pairs	39
Optimiser	Adam
Learning Rate	0.0002
Loss	$L1 + 0.5 \times \text{MSE}$
Parameters	31.04 million
Framework	PyTorch 2.6 with ROCm 6.1

Images are resized with bilinear interpolation and converted to tensors in the range 0 to 1. No random geometric augmentation is applied, because the binding holes and the line geometry of a palm leaf are identity cues that should not be warped away in this first protocol. The best validation checkpoint is written to disk as `best_restoration_unet_highres.pth`, a file of about 118 MB. Per-epoch metrics are serialised to a JavaScript Object Notation (JSON) log, so that every number in Section 6 can be traced to a log rather than to a screenshot.

6. Results and Discussion

6.1 Quantitative Training

Table 2 reports the logged summary. Validation loss falls from 0.3691 in epoch 1 to 0.0314 in epoch 30. Training loss tracks the same trajectory, from 0.2855 to 0.0311. That agreement indicates that the model is not merely memorising the 223 training leaves at the expense of the held-out 39. Two transient validation spikes, at epochs 18 and 22, recover within one epoch. They are retained in the log rather than smoothed away. Peak allocated VRAM is 42.54 GB throughout, which fits on a 64 GB MI210 and explains why an input of 384 by 1280 pixels at a batch size of 16 is feasible. The mean epoch time after warm-up is about 55 seconds. The total wall-clock time is 1682.46 seconds, which is 28.04 minutes.

Table 2: Logged Restoration Training Summary on the MI210

Metric	Value
Best Validation Loss	0.0314
Final Training Loss	0.0311
Total Time	28.04 minutes
Mean Epoch Time	56.08 seconds
Peak Allocated VRAM	42.54 GB
Steady Throughput	About 4 images per second

Figure 4: Training and Validation Loss over 30 Epochs.

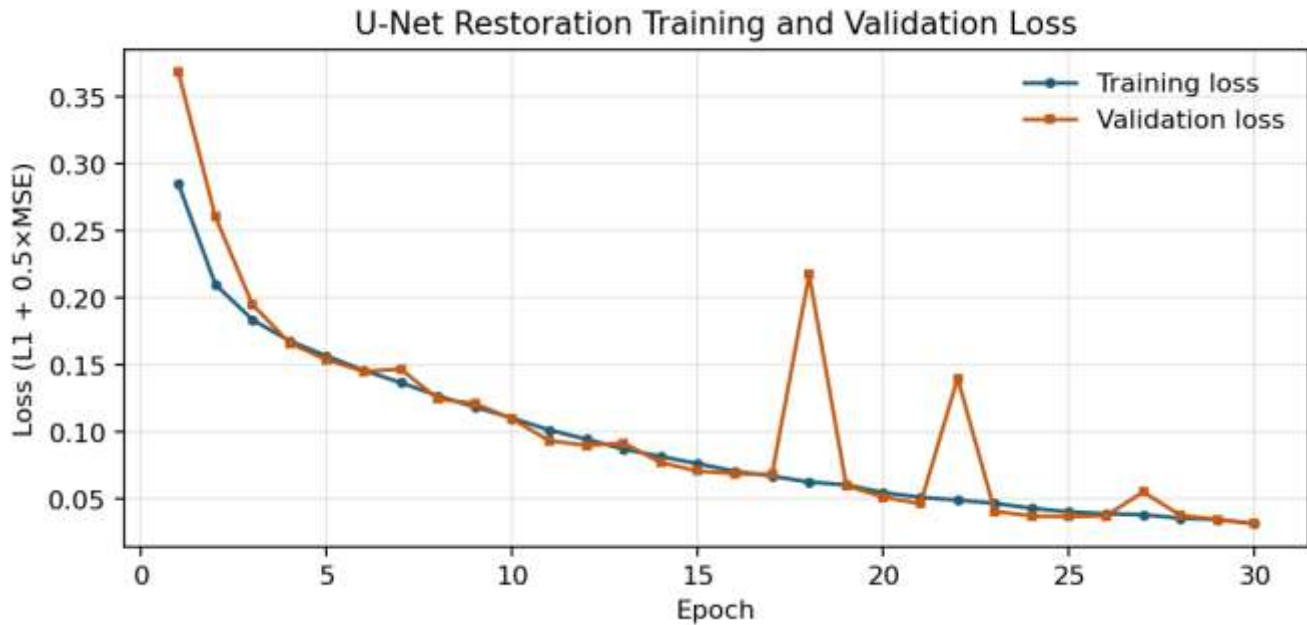


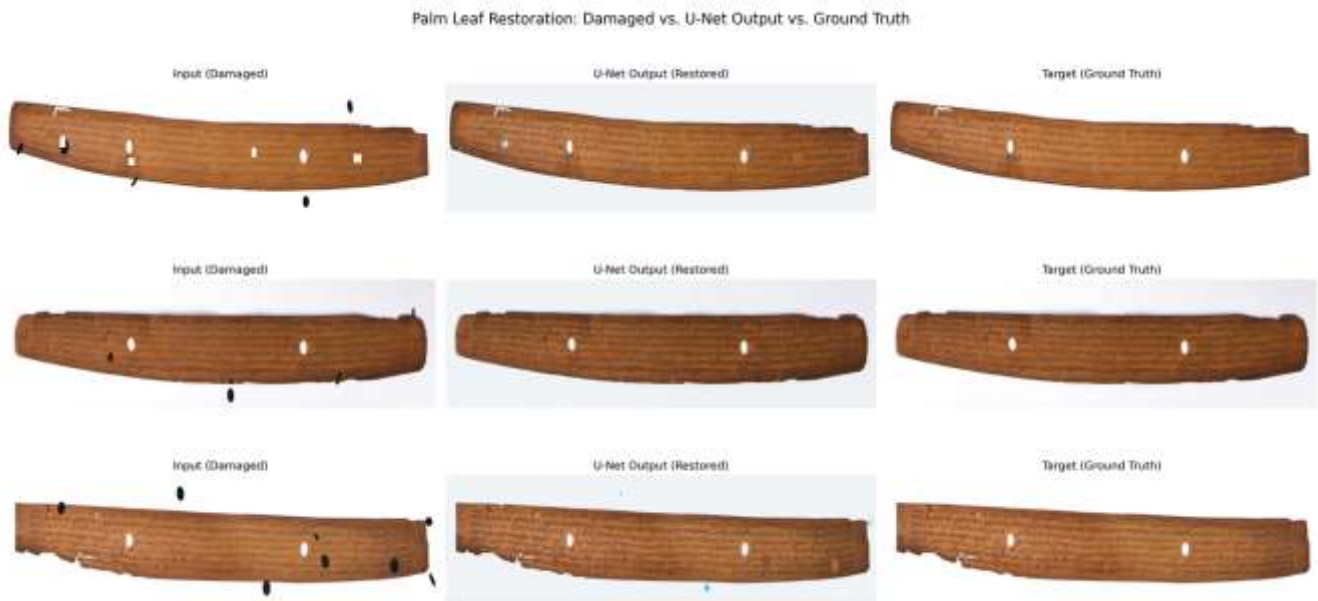
Figure 4 shows the loss curves. The absence of a large, persistent gap between training loss and validation loss is consistent with a pixel-reconstruction task on a homogeneous domain, namely Tamil palm-leaf scans of three works, rather than with a highly diverse natural-image problem. The limitation is the converse. Leaf-level diversity is modest, even though the pixel-level sample is large.

6.2 Qualitative Restoration

Figure 5 compares the damaged input, the U-Net output, and the clean ground truth on three representative leaves. Simulated white patches and dark occlusions are filled with plausible fibre texture and with continuations of the strokes. Binding holes are preserved rather than painted over, which is the correct archival behaviour: a hole is part of the object, not noise. Restored script is slightly softer than the target, which is a known effect of averaging under an L1 and MSE loss, but stroke continuity is recovered in regions where the input was blank. Torn geometry at a leaf tip is not hallucinated into a complete rectangle. The model inpaints texture inside the existing silhouette.

These panels support a limited claim. The network inverts the documented synthetic operators. They do not prove that a museum leaf with centuries of real damage will be restored to an original agreed by an expert, because no such paired ground truth was available. Structural similarity (SSIM), peak signal-to-noise ratio (PSNR) on the synthetic split, and the OCR gain measured as CER on restored crops against unrestored crops, are the named next metrics. They are not invented here.

Figure 5: Qualitative Restoration of Three Tamil Palm-Leaf Samples.



6.3 Choice of Objective and Canvas Size

An unconstrained adversarial restorer could look sharper in a demonstration. For heritage images, sharpness that is not present in the target is a liability. The L1 and MSE objective is a deliberate ethical choice as well as a technical one. The residual can be inspected, and the damaged original is always kept beside the restoration. The canvas of 384 by 1280 pixels is the matching geometric choice. Compressing a leaf whose aspect ratio is about 6 to 1 into a square of 256 by 256 pixels would destroy the spacing between characters and the horizontal ruling implied by the fibre. The institutional GPU is used so that the leaf can keep its shape.

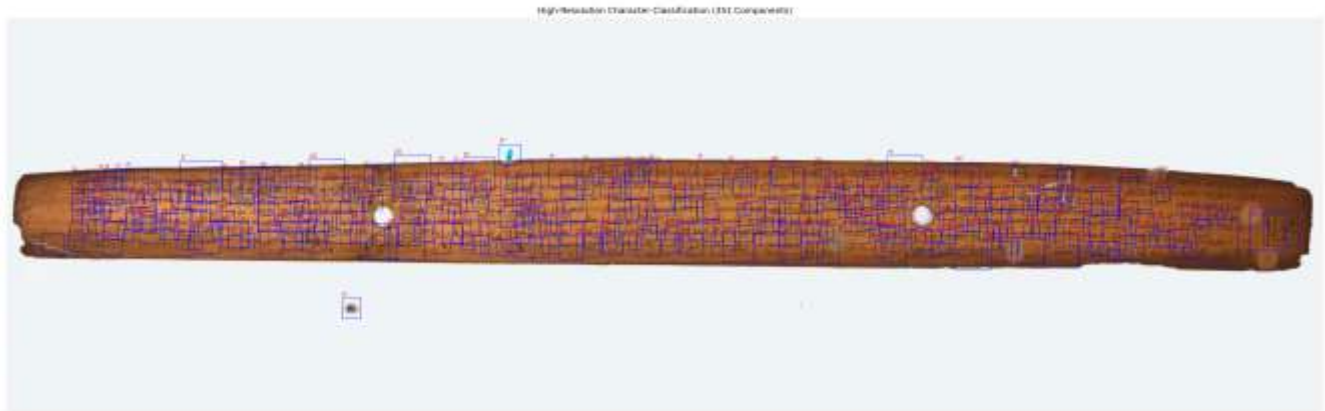
7. Toward Transcription and Translation

Restoration is justified for reading only if it eventually reduces character error or expert effort. That experiment is not closed in this paper, but the control stack is in place.

A Tesseract OCR engine with Tamil language data was evaluated on Mozhi printed image-and-text pairs and produced a mean CER of approximately 0.34 [4, 15]. This result is a usable printed baseline. It must not be read as a palm-leaf HTR score. The Pracalit collection in Transkribus contains 441 manuscript pages with XML ground truth, on which plain OCR is expected to fail [16]. That expectation motivates a specialised HTR upgrade, either ConvNeXt character classifiers [14] or PaliGemma-style vision-language prompting [22]. Figure 6 shows a high-resolution character-component overlay on a restored strip, with 351 boxed components. The figure demonstrates localisation on a palm leaf. It does not demonstrate that Unicode transcripts are already correct.

Translation is a planned proof of concept. IndicTrans2 is specified to run offline for Tamil-to-English [5], scored with the bilingual evaluation understudy (BLEU) and the character n-gram F-score (chrF++) on the FLORES benchmark [17]. The acceptance criterion is ten restored-page transcripts plus a manual spot-check. No translation checkpoint is claimed. The module is blocked on stable transcription text, which is the scientifically honest dependency.

Figure 6: Character-Component Localisation on a Restored Palm-Leaf Strip.



8. Limitations

Several limitations follow from the protocol. First, supervision is synthetic. Real unpaired museum damage is a domain shift that has not been measured. Second, 39 validation leaves are a small sample at the level of the leaf, even though each leaf contains hundreds of thousands of pixels. Third, the reported loss is the training objective, not SSIM, PSNR, or CER. Fourth, the Tesseract CER is measured on printed Mozhi text, not on palm-leaf handwriting. Fifth, an English gloss, if shown in a demonstration, must be labelled as illustrative until IndicTrans2 has been trained and reviewed. These constraints are listed so that a reviewer can map every claim to its evidence.

9. Conclusion

This paper described a high-resolution U-Net pipeline for restoring degraded Tamil palm-leaf manuscript images. A paired dataset of 262 leaves was built with documented OpenCV damage operators. A U-Net of 31.04 million parameters, trained for 30 epochs at 384 by 1280 pixels on an AMD Instinct MI210, reached a validation loss of 0.0314 in 28.04 minutes. The qualitative outputs invert simulated holes, erosion, and missing blocks while preserving leaf geometry and binding holes. Transcription is a partial proof of concept, covering a Tesseract Tamil baseline on Mozhi and a character-component visualisation. Translation is a designed IndicTrans2 plan, not a trained system. Completing those two stages, with the same logging discipline that was required of restoration, is the work of the next phase. The present contribution is a reproducible restoration front-end for Indic heritage images. It is not a claim that the manuscripts have already been read.

Acknowledgement

The authors thank the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, and the administrators of the institutional AMD Instinct MI210 server, for compute access. Dr. Golda Dilip, Professor and Head of the Department, is thanked for institutional support.

References

1. M.A. Souibgui, Y. Kessentini, “DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
2. Z. Yang, et al., “Damaged Historical Document Appearance Prediction”, Proceedings of the AAAI Conference on Artificial Intelligence, 2025.
3. O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, 234–241.
4. Indic OCR Community, “Mozhi Indic OCR Dataset”, Hugging Face, darknight054/indic-mozhi-ocr.
5. AI4Bharat, “IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for Indian Languages”, 2023.
6. United Nations, “Sustainable Development Goal 11, Target 11.4”, Cultural Heritage Protection.
7. Document Restoration Literature, “HDR-Style Paired Degrade-to-Clean Training (diffHDR Family)”.
8. K. Nikolaidou, M. Seuret, H. Liwicki, M. Liwicki, “A Survey of Historical Document Image Datasets”, International Journal on Document Analysis and Recognition, 2022.
9. I. Jailingeswari, S. Gopinathan, “Tamil Handwritten Palm Leaf Manuscript Dataset (TH-PLMD)”, 2024.
10. N. Thuon, J. Du, P. Theang, R. Thuon, “PALM-LAY: A Multi-Script Cross-Regional Dataset for Layout Analysis of Palm Leaf Manuscripts”, Proceedings of the International Conference on Document Analysis and Recognition, 2025.
11. D. Valy, M. Verleysen, S. Chhun, J.C. Burie, “A New Khmer Palm Leaf Manuscript Dataset for Document Analysis and Recognition: Sleukrith Set”, 2017.
12. M.W.A. Kesiman, et al., “AMADI LontarSet: The First Handwritten Balinese Palm Leaf Manuscripts Dataset”, 2016.
13. L. Li, “Handwriting Recognition in Historical Documents with Multimodal LLM”, 2024.
14. Z. Liu, et al., “A ConvNet for the 2020s”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
15. Tesseract OCR Project, “Tesseract Open-Source OCR Engine with Tamil Language Data”.
16. Transkribus, “Pracalit Manuscript Collection with Page-Level XML Ground Truth”, Zenodo.
17. N. Goyal, et al., “The FLORES-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation”, Transactions of the Association for Computational Linguistics, 2022.
18. G. Bradski, “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools, 2000.
19. D.P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization”, Proceedings of the International Conference on Learning Representations, 2015.
20. AMD, “ROCm and Instinct MI210 Documentation”.
21. A. Paszke, et al., “PyTorch: An Imperative Style, High-Performance Deep Learning Library”, Proceedings of the Conference on Neural Information Processing Systems, 2019.
22. Google, “PaliGemma Vision-Language Models”.