|  |
| --- |
| **Mask Aware Mural Inpainting** |

**Jincheng Jiang**

College of Engineering

Northeastern University

Toronto, ON

*Jiang.jinc@northeastern.edu*

**Abstract**

This project investigates mask-aware approaches for mural restoration using the DhMurals1714 dataset with randomly generated masks (ratio: 0.2-0.3) across 1564 training, 100 validation, and 50 test images resized to 256×256 pixels. I evaluated three models—Partial Convolution, Mask-Aware Dynamic Filtering (MADF), and Progressive Reasoning Network (PRN)—using SSIM, PSNR, and LPIPS metrics to assess structural similarity, signal quality, and perceptual similarity, respectively. My comparative analysis demonstrates that MADF outperforms the other approaches, offering superior feature extraction capabilities specifically tailored for mural restoration challenges.

**1 Datasets**

https://1drv.ms/u/s!AittnGm6vRKLzXorf1nkiDPRQB4D?e=Avv27i

DhMurals1714

**1.2 Why is this an interesting dataset?**

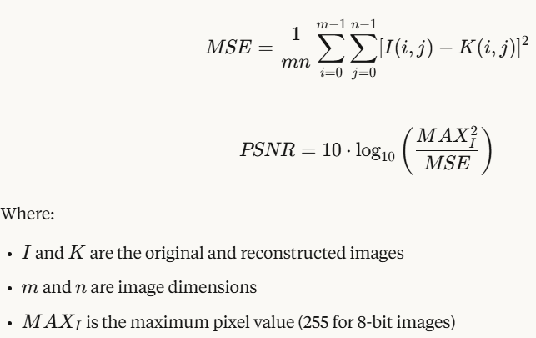
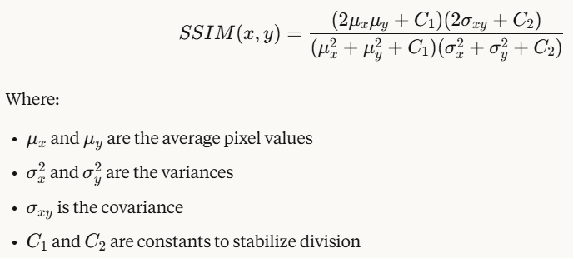
The DhMurals1714 dataset captures the complex textures, historical painting techniques, and distinctive deterioration characteristics of real murals, creating an ideal testbed for evaluating how specialized mask-aware approaches perform on historically valuable artifacts compared to generic images.

**2 Data processing**



I implemented a systematic approach to prepare the mural images and their corresponding masks. First, I center-cropped all images from the DhMurals1714 dataset to ensure a square aspect ratio before resizing them to a uniform 256×256 pixel resolution, which preserved the central details of each mural while standardizing the input dimensions for the neural networks. For mask generation, I created random stroke-like patterns that mimic realistic damage rather than using simple geometric shapes, better reflecting the irregular deterioration patterns found in actual murals. To maintain experimental consistency, I carefully controlled the mask-to-image ratio within the range of (0.2, 0.3], striking a balance between sufficient damage to test restoration capabilities while preserving enough original content for reference. This dataset consisted of 1564 training images, 100 validation images, and 50 test images, providing a robust foundation for evaluating the performance of the different mask-aware restoration approaches.

**2 Evaluation Metrics**





SSIM measures the perceived quality of images by comparing structural information

(Higher values are better)

PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise.

(Higher values are better)

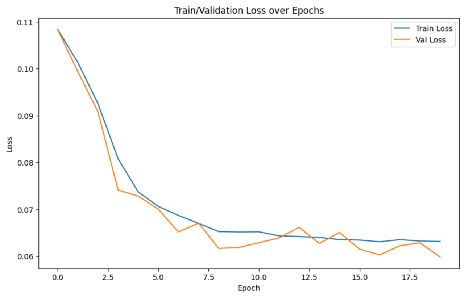
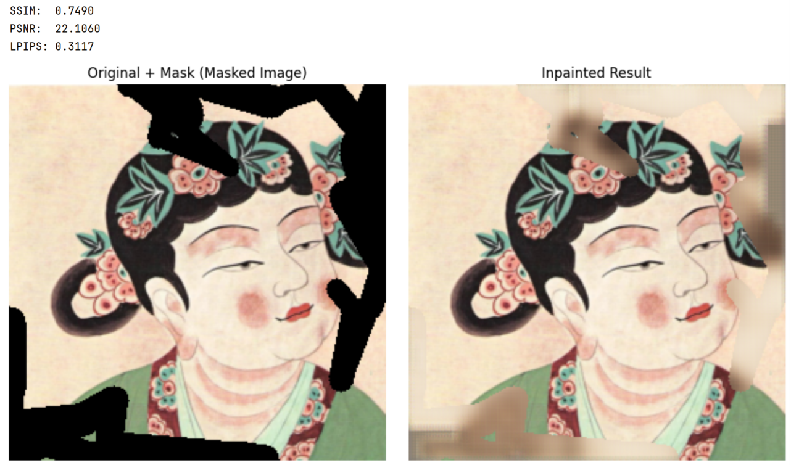
LPIPS uses deep neural networks to measure perceptual similarity

(Lower values are better)

**3 Model structure and performance**

**3.1 Partial Convolutions**

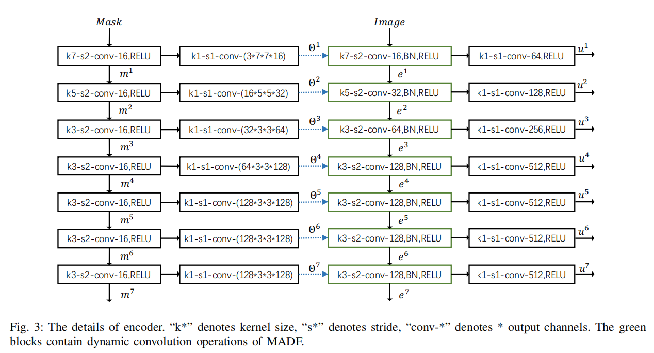
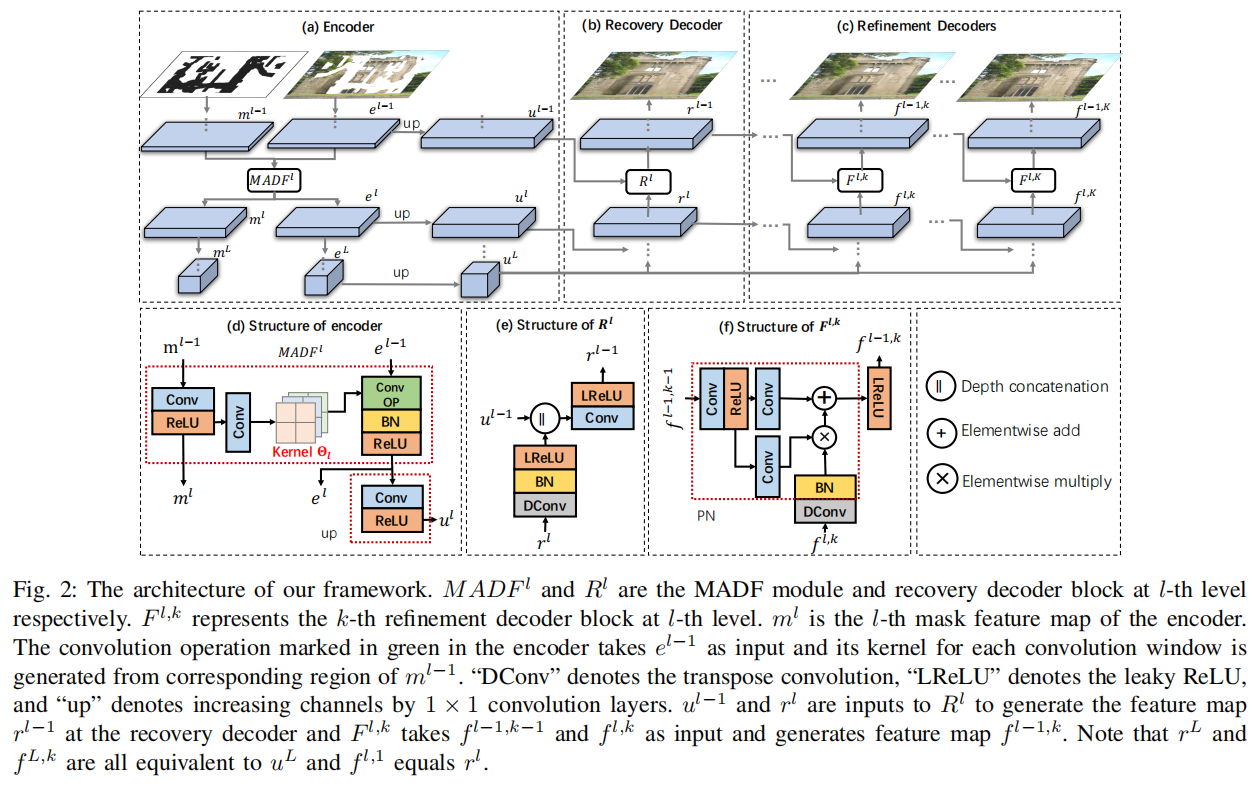
For the model structure, I implemented a U-Net architecture featuring an encoder with partial convolutions and a decoder with transposed convolutions. The encoder progressively processes the input through increasing channel dimensions (3→64→128→256), while the decoder reconstructs the image through decreasing channel dimensions (256→128→64→3). The core innovation lies in the use of partial convolutions, which fundamentally differ from standard convolutions by dynamically adapting to missing data—while standard convolutions treat all pixels equally regardless of validity, partial convolutions only consider valid pixels when computing features and automatically update the mask during forward propagation, making them particularly effective for mural inpainting where damage patterns are irregular.

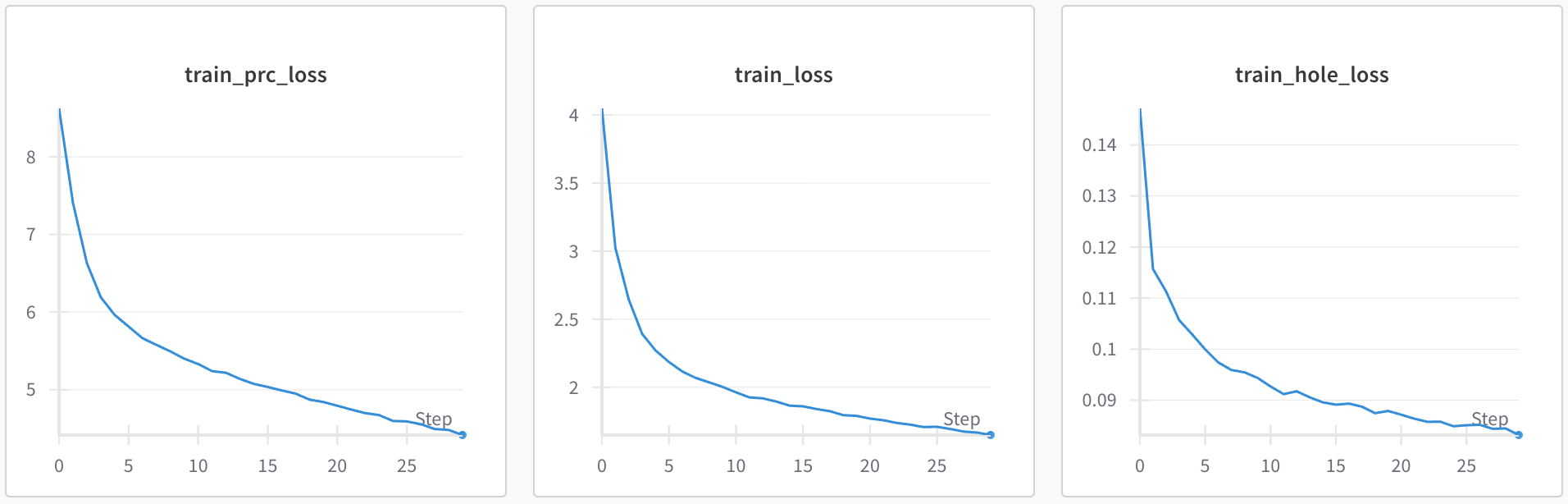
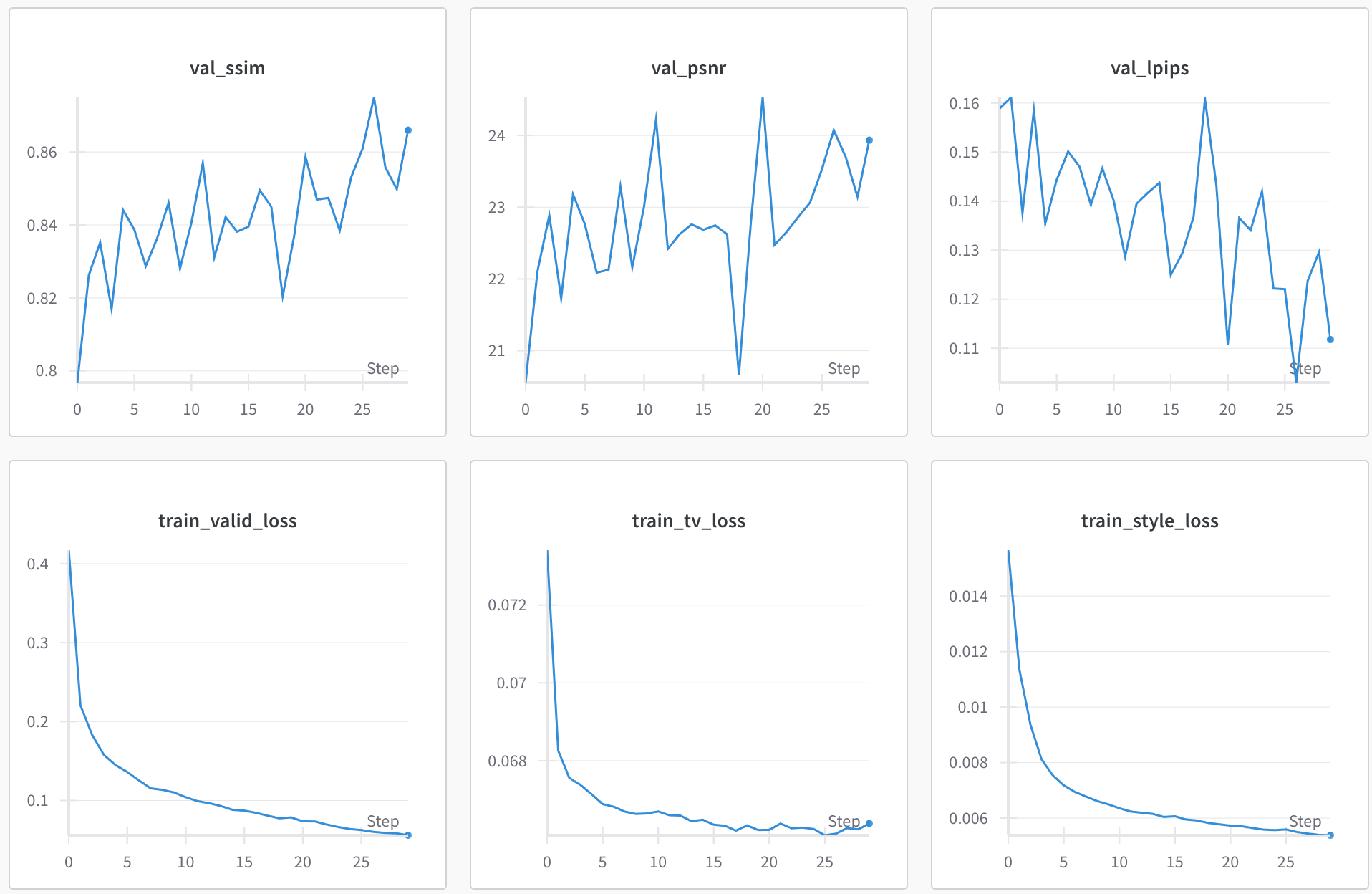
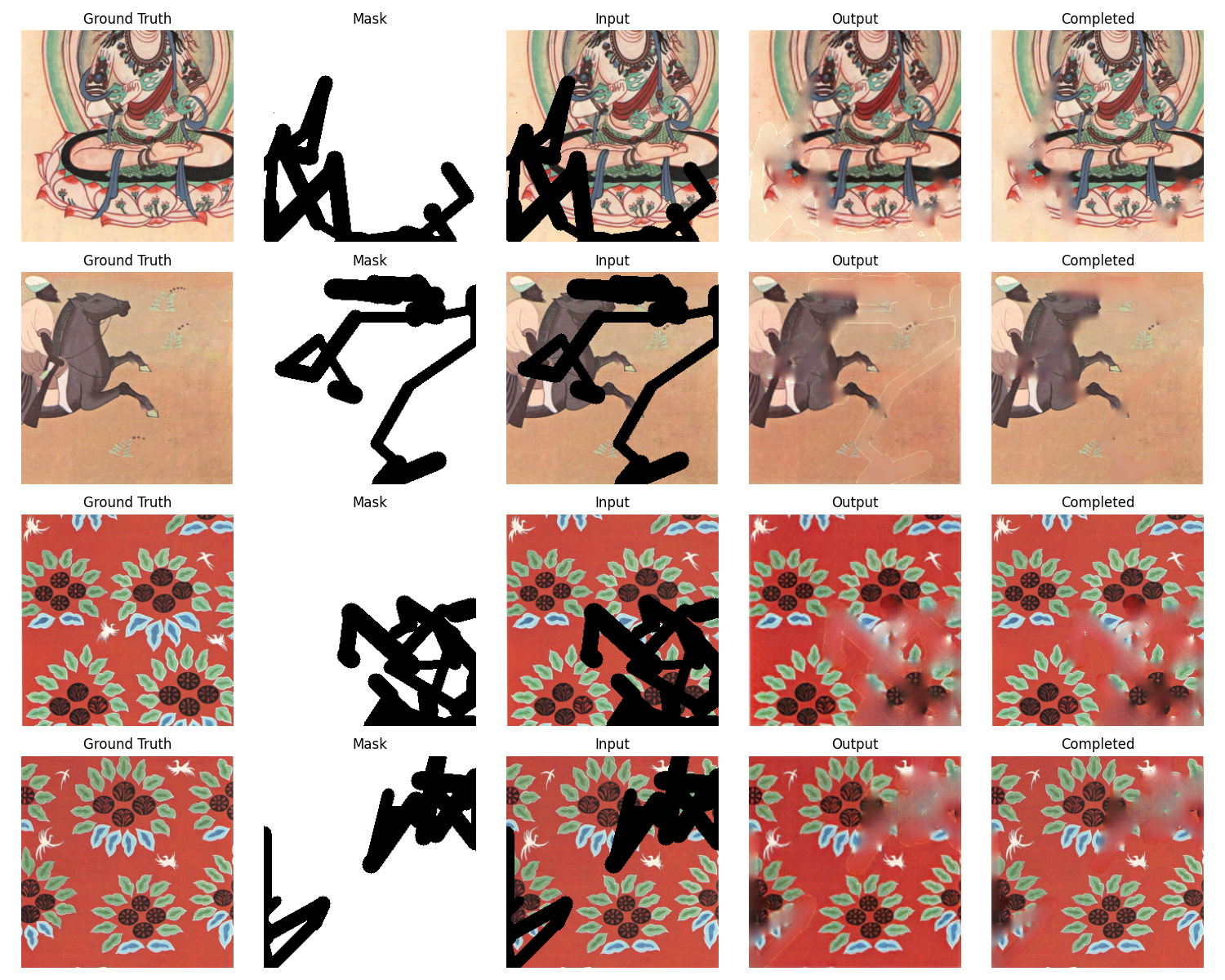


My training process showed promising convergence as evidenced by the smoothly decreasing loss curves for both training and validation sets, with validation loss stabilizing around 0.06 after approximately 15 epochs, indicating good generalization without overfitting. When evaluated on the test set, the model achieved moderate performance with an SSIM of 0.7490, PSNR of 22.106, and LPIPS of 0.3117. While these metrics indicate partial success in restoration, they also reveal limitations in fully reconstructing the original mural details. As shown in the example female figure with floral decorations, the model managed to regenerate the basic structure and coloration but struggled with finer details and texture coherence, particularly in areas with complex patterns. This suggests that while partial convolutions offer improvements over standard approaches for mural restoration, there remains considerable room for enhancement in handling the unique characteristics of ancient mural artwork.

**3.2 MADF**

I implemented a Mask-Aware Dynamic Filtering (MADF) framework featuring a sophisticated architecture with three interconnected components: an encoder, a recovery decoder, and refinement decoders. Unlike partial convolutions that merely ignore damaged pixels with a fixed mask updating mechanism, MADF dynamically generates customized convolution kernels specifically tailored to each window's unique mask pattern. This fundamental difference allows MADF to adaptively process different damage patterns with specialized filters rather than applying the same filter everywhere, resulting in more context-appropriate feature extraction for mural restoration challenges.



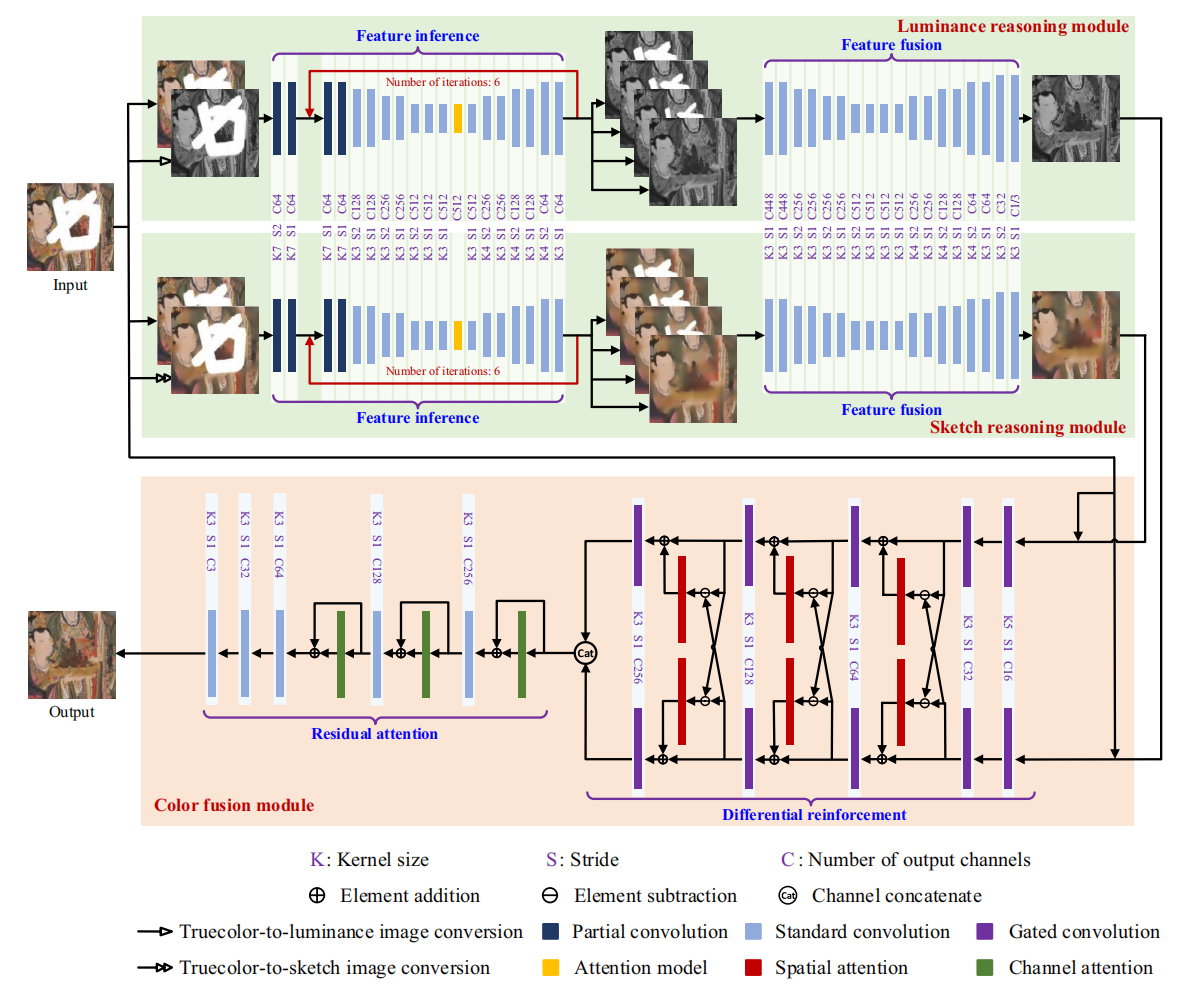


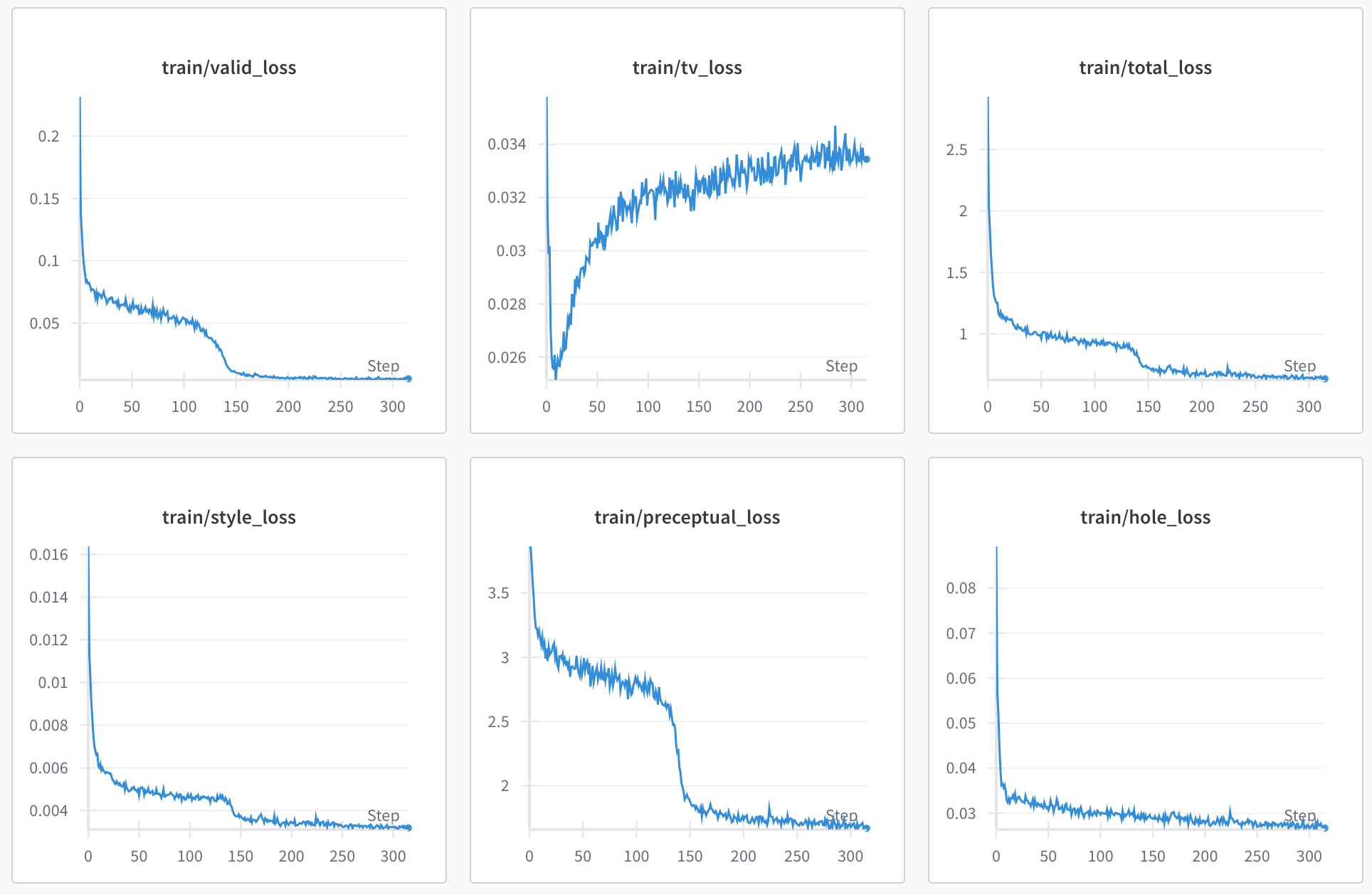
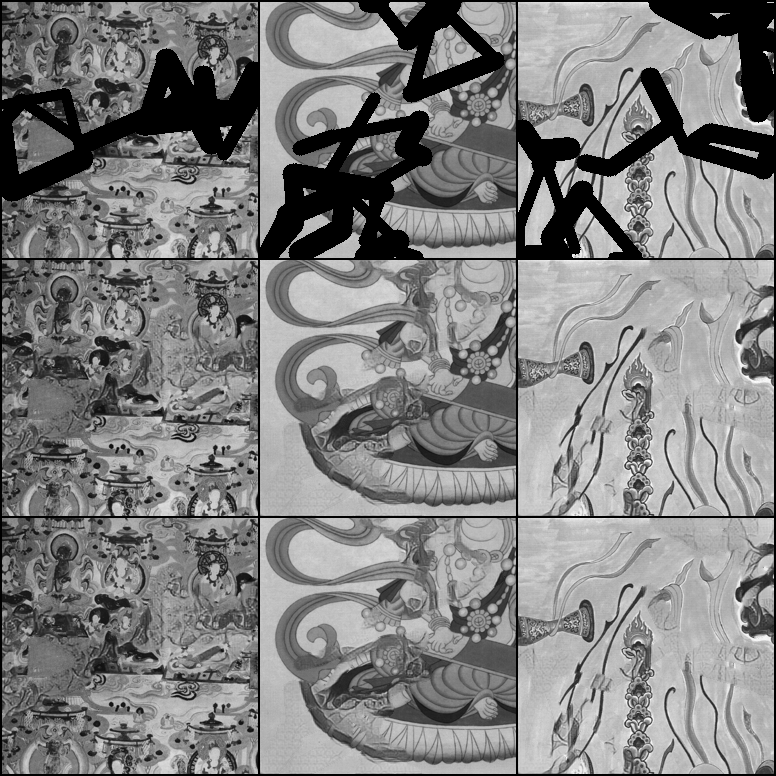
Despite limited training of only 30 epochs without full convergence due to time constraints, the MADF model demonstrated remarkable performance improvements over the partial convolution approach. Quantitatively, MADF achieved significantly better metrics with an SSIM of 0.875 (versus 0.749), PSNR of 24.075 (versus 22.106), and LPIPS of 0.103 (versus 0.3117). The training curves show clear improvement trends across all loss components. Qualitatively, as evident in the visual comparison examples, MADF produced more coherent restorations with better preservation of artistic styles, textures, and structural details across diverse mural examples including figurative paintings, animal motifs, and floral patterns, suggesting that dynamic kernel generation is particularly effective for handling the complex damage patterns in historical murals.

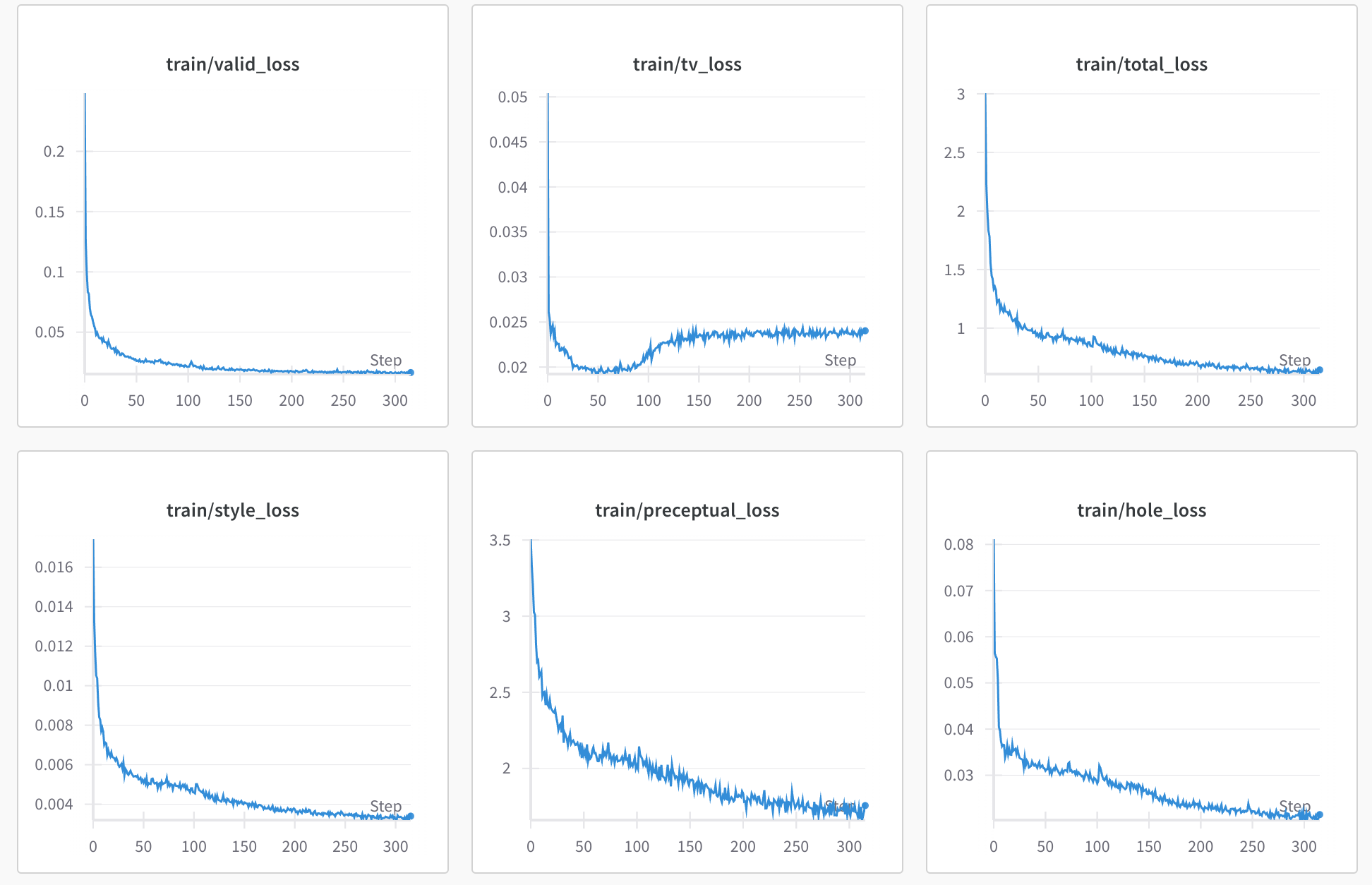
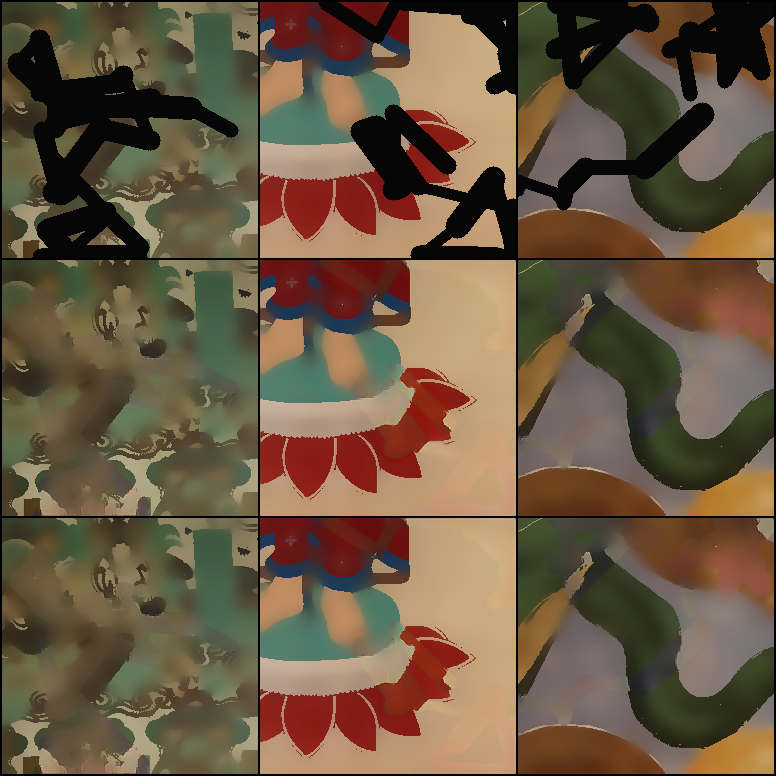
In designing the loss function for mural restoration, I implemented a weighted composite approach to balance different restoration priorities. The total loss function combines multiple components: L\_total = L1 + 0.05L\_perc + 120L\_style + 0.1L\_tv, where L1 = L\_valid + 6L\_hole. This formulation assigns significantly higher weight (6×) to pixels in damaged regions compared to valid regions, forcing the model to prioritize accurate hole filling. The perceptual loss (L\_perc) using VGG features helps ensure semantic consistency, while the style loss (L\_style) receives a substantially higher weight (120) because the raw values of style loss are typically much smaller than other loss components, requiring scaling to make its contribution meaningful to the overall optimization. The total variation loss (L\_tv) with a modest weight (0.1) promotes local smoothness without over-blurring. These carefully calibrated weights create a balanced optimization landscape where each component contributes proportionally to guide the restoration process, ensuring both structural accuracy and artistic coherence in the restored murals.

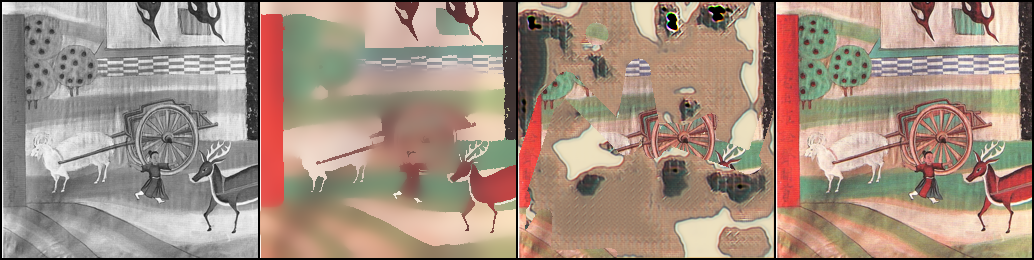
**3.3 PRN**

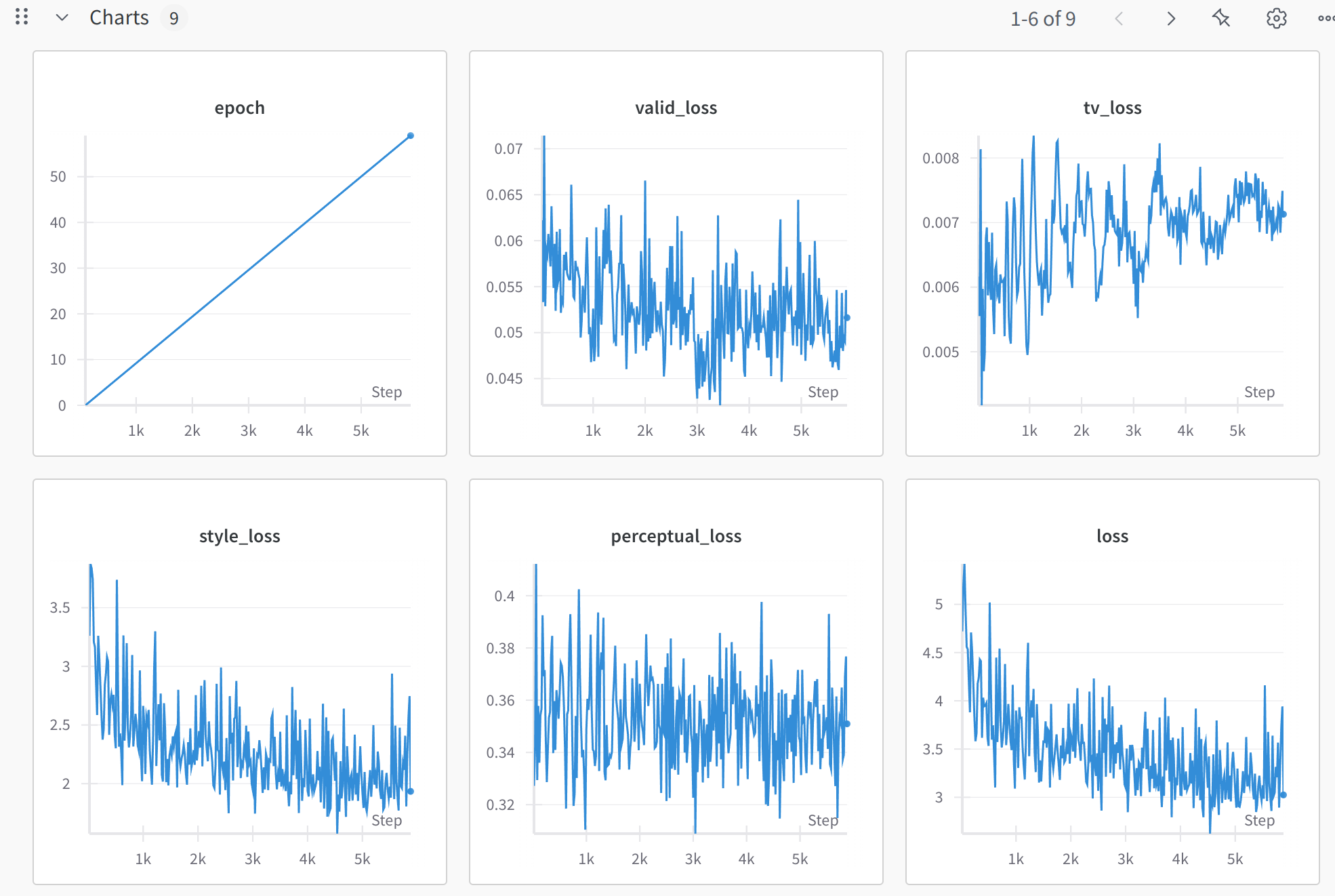
I implemented the Progressive Reasoning Network (PRN), which uniquely incorporates the traditional mural creation process—from sketch to coloring—into its architecture for restoration tasks. Unlike previous approaches, PRN decomposes the complex restoration challenge into three specialized modules: a luminance reasoning module that infers light intensity patterns, a sketch reasoning module that reconstructs structural outlines, and a color fusion module that integrates these intermediate representations to produce the final restoration. This progressive approach mirrors how artists actually create murals, with each module featuring its own cyclic double-codec framework for feature inference and fusion, complemented by specialized attention mechanisms to capture both spatial and channel relationships critical for artistic coherence.











Despite using the same comprehensive loss function as MADF (L\_total = L1 + 0.05L\_perc + 120L\_style + 0.1L\_tv, where L1 = L\_valid + 6L\_hole), the PRN model achieved less satisfactory results with metrics of PSNR: 15.358, SSIM: 0.456, and LPIPS: 0.513. This underperformance can be attributed to two primary factors: first, the modular training approach where each component was trained separately introduced cumulative errors during inference as imperfections propagated between stages; second, the model's sophisticated architecture requires substantial computational resources and longer training periods than were available for this experiment. The training curves show initial improvement followed by instability, suggesting that while the progressive reasoning concept has theoretical merit for mural restoration, practical implementation faces significant challenges in achieving stable optimization across interconnected components.

**4 Examine Ethical Implications**

First, while digital restoration helps preserve cultural heritage without physically altering original artifacts, questions arise about authenticity and interpretation—my algorithms make subjective decisions about what constitutes "correct" restoration, potentially imposing modern aesthetic judgments on historical works. Second, there are concerns about cultural appropriation and ownership, as advanced AI technologies might enable wider access to cultural artifacts without proper attribution or permission from their cultural stewards. Additionally, the carbon footprint of training resource-intensive models like PRN raises environmental ethical questions. Going forward, responsible development in this field requires transparent documentation of restoration decisions, collaborative relationships with cultural heritage experts and stakeholders, and ongoing dialogue about the appropriate boundaries between preservation, restoration, and reinterpretation of irreplaceable cultural artifacts.

**6 Conclusion**

I investigated mask-aware deep learning approaches for ancient mural restoration, exploring three progressively sophisticated models: Partial Convolution, Mask-Aware Dynamic Filtering (MADF), and Progressive Reasoning Network (PRN). Using the DhMurals1714 dataset with randomly generated masks, I developed a comprehensive evaluation framework measuring SSIM, PSNR, and LPIPS metrics. My experiments demonstrated that MADF achieved the best performance (SSIM: 0.875, PSNR: 24.075, LPIPS: 0.103) by dynamically generating convolution kernels tailored to specific damage patterns, significantly outperforming the baseline Partial Convolution approach. While the art-process-inspired PRN showed theoretical promise by decomposing restoration into luminance, sketch, and color stages, its practical implementation faced challenges with error propagation across modules.

**7 Reference**

https://www.nature.com/articles/s41598-024-72368-1

<https://github.com/qinnzou/mural-image-inpainting>

<https://arxiv.org/html/2211.06649v2>

<https://github.com/albestobe/PRN>

https://arxiv.org/abs/2104.13743

<https://paperswithcode.com/task/image-inpainting>

<https://github.com/MADF-inpainting/Pytorch-MADF>

<https://arxiv.org/abs/1804.07723>

https://1drv.ms/u/s!AittnGm6vRKLzXorf1nkiDPRQB4D?e=Avv27i