Machine Learning for Image Deblurring Using MNIST and FashionMNIST Datasets

Predicting Clear Images from Blurred Inputs

EJ Mungovan, Xin Jing, Rigo Carreto, Al Ingold, Dr. Rajesh Menon, Dr. Apratim Majumder

Department of Electrical and Computer Engineering
University of Utah

April 29, 2025

Contents

1 Introduction

This project focuses on developing a machine-learning-based system to deblur images generated from the MNIST and FashionMNIST datasets. By leveraging a hybrid CycleGAN-Pix2Pix architecture, we aim to reconstruct clear images from both semi-blurred and heavily blurred inputs. The system employs a generator-discriminator framework where the generator learns to produce deblurred images, and the discriminator distinguishes between predicted and ground truth (GT) images.

The computational tasks are handled by a single Raspberry Pi 5 device, with heat management provided by a Raspberry Pi 5 cooler. The setup is compactly housed within a 12-inch box for portability, which is scalable down depending on the specific use case. This innovative solution demonstrates the potential of combining low-cost hardware with advanced machine learning techniques to address real-world challenges in image processing.

In addition to CycleGAN, we also explore the use of Pix2Pix, a supervised generative adversarial network specifically designed for paired image translation tasks. Given our access to paired blurred and ground truth images, Pix2Pix offers a suitable framework for direct supervised learning and allows for finer detail restoration.

1.1 Background

Image blurring is a common issue in photography and imaging systems, often caused by motion, defocus, or low-light conditions. Traditional methods for deblurring rely on computationally expensive algorithms or require prior knowledge of the blur kernel. In this project, we aim to bypass these limitations by using a hybrid CycleGAN-Pix2Pix model that can learn both unpaired domain mappings and paired pixel-level mappings between blurred and clear images.

Using the MNIST and FashionMNIST datasets as benchmarks, we generate two levels of blurred images (semi-blurred and heavily blurred) for each dataset. These images are paired with their GT counterparts to create a dataset of 50 blurred-GT pairs for each blur level. The hybrid model is trained on this dataset using a Raspberry Pi 5 and PC.

This approach offers a scalable and efficient solution for real-time image deblurring, with potential applications in various domains such as surveillance, autonomous vehicles, and medical imaging.

1.2 Project Objectives and System Overview

The objective of this project is to enhance the quality of images captured using a traditional Raspberry Pi lens by developing a machine learning-based deblurring system. The proposed solution leverages a hybrid CycleGAN-Pix2Pix model, which is trained on blurred and ground-truth image pairs derived from the MNIST and FashionMNIST datasets. This approach allows the system to benefit from both paired and unpaired image translation strategies, thereby addressing scenarios where perfectly aligned image pairs may not be available.

A key focus of the project is to optimize the deblurring process for real-time performance on Raspberry Pi 5 hardware, taking into account its computational constraints and leveraging its processing capabilities. The system will be rigorously evaluated on datasets

exhibiting varying levels of blur, including both semi-blurred and heavily blurred images, to ensure robust performance across different conditions.

Furthermore, the project aims to compare the effectiveness of unpaired versus paired training strategies, providing insights into the most efficient approaches for practical deployment. Ultimately, this work seeks to demonstrate the practical applications of the developed system in enhancing image quality for tasks such as surveillance, document scanning, and other real-world scenarios where improved image clarity is essential.

1.2.1 Equipment and System Architecture

Our proposed system integrates specialized hardware with advanced machine learning techniques to enable efficient image deblurring on embedded devices. The main hardware and software components are summarized below.

Component	**Description**
Raspberry Pi Lens	Captures images with controlled defocus to generate varying blur levels.
Raspberry Pi 5	Performs model training and inference for deblurring tasks.
Raspberry Pi 5 Cooler	Maintains stable thermal conditions during intensive computation.
Compact 12-inch Box	Houses all hardware components for portability and ease of deployment.

Table 1: Hardware components of the proposed system.

The software stack features a hybrid CycleGAN-Pix2Pix model, which performs image-to-image translation from blurred to clear domains using adversarial, cycle consistency, and L1 loss functions. Custom Python scripts, `fashionmnist.py` and `mnist.py`, are used to access and display the FashionMNIST and MNIST datasets, respectively, facilitating data preparation and capture for model training and evaluation.

1.2.2 System Workflow

The deblurring pipeline begins with controlled image acquisition using MNIST and FashionMNIST datasets, where ground truth images are artificially degraded through calibrated defocus operations to create two distinct blur categories: semi-blurred (moderate lens defocus) and heavily blurred (severe motion artifacts). These synthetic blurred-GT pairs establish the foundation for both training and validation phases.

Following data generation, preprocessing occurs directly on the Raspberry Pi 5 hardware, where images undergo resizing to 256×256 resolution and pixel value normalization to the [-1,1] range. This on-device processing ensures compatibility with subsequent stages while leveraging the Pi's I/O capabilities.

The hybrid model training phase implements a dual-strategy approach: CycleGAN components handle unpaired image translation where perfect blurred-sharp correspondences are unavailable, while Pix2Pix layers exploit available paired data through conditional adversarial training. Three loss components – adversarial loss for realism preservation, cycle consistency for structural coherence, and L1 loss for pixel-level accuracy – work synergistically to optimize generator performance across blur intensities.

During deployment, new blurred images first pass through a lightweight classifier that estimates blur severity, then route through the appropriate generator pathway. This conditional execution balances computational load against reconstruction quality, enabling real-time deblurring on Raspberry Pi 5 hardware through optimized TensorFlow Lite model inference.

1.2.3 Key Advantages

This system leverages a low-cost hardware setup using Raspberry Pi components to achieve efficient real-time image processing. By integrating a hybrid CycleGAN-Pix2Pix architecture, the system combines the flexibility of unpaired translation with the precision of supervised learning. This approach allows for sharper reconstructions, better structural fidelity, and greater adaptability across various blurring conditions. Its scalable design makes it suitable for diverse applications.

2 Methodology

Our code and visualization is available at this GitHub repository.

2.0.1 Data Collection

To create the training dataset, the MNIST and FashionMNIST datasets were loaded as ground truth (GT) images. Two levels of blurred versions were generated to simulate different types of image degradation: semi-blurred, representing mild defocus or motion blur, and heavily blurred, simulating extreme defocus or motion blur. Using a traditional Raspberry Pi lens, 50 GT-blurred pairs were captured for each blur level, forming a comprehensive dataset for training and evaluation. This format maintains clarity while presenting the information cohesively.

2.0.2 Machine Learning Model and Implementation

The image deblurring system is built on a hybrid CycleGAN-Pix2Pix architecture, strategically combining the strengths of both unpaired and paired image-to-image translation approaches. At its core, the model employs two generators and two discriminators working in tandem. The CycleGAN backbone provides crucial cycle consistency for domain translation between blurred and clear image domains, which proves especially valuable when exact image pairings are unavailable or inconsistent in the training data. For scenarios where paired blurred and ground truth (GT) images are accessible, a Pix2Pix-style supervised L1 loss is incorporated during training to enhance detail preservation and alignment, guiding the generator to produce sharper outputs that closely match their GT counterparts.

The implementation process begins with comprehensive dataset preprocessing, where images are resized to uniform dimensions (256×256) and pixel values are normalized to the [-1,1] range to ensure consistent training dynamics. The model training phase occurs directly on the Raspberry Pi 5 hardware, leveraging data augmentation techniques such as rotation, cropping, and intensity shifts to improve generalization capabilities and prevent overfitting. Throughout training, the model benefits from a dual-strategy approach that leverages both paired and unpaired data, effectively balancing robustness

with output fidelity. Once trained, the model's performance is rigorously evaluated on previously unseen blurred-GT test pairs, using both structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) metrics to quantify reconstruction quality across varying blur intensities and image characteristics.

2.1 Performance and Evaluation

The performance of the hybrid CycleGAN-Pix2Pix model was evaluated on the MNIST dataset using SSIM and PSNR metrics. The model achieved an average SSIM of 0.7585, indicating strong structural alignment, and an average PSNR of 15.18 dB, reflecting effective noise reduction.

The hybrid CycleGAN-Pix2Pix model was evaluated on both the MNIST and FMNIST datasets. Training was conducted for 10 epochs using 41 training samples and 10 test samples for each dataset. The loss function was Binary Cross Entropy. The performance of the model was assessed on average using L1 Loss (Mean Absolute Error), L2 Loss (Mean Squared Error), SSIM (Structural Similarity Index Measure) and PSNR (Peak Signal-to-Noise Ratio) metrics.

The average evaluation results are summarized in Table 2.

Dataset	**Avg L1 Loss**	**Avg L2 Loss**	**Avg SSIM**	**Avg PSNR (dB)**
MNIST	0.1323	0.1295	0.7448	15.12
FMNIST	0.1598	0.0962	0.5794	16.32

Table 2: Quantitative evaluation results on MNIST and FMNIST datasets.

The quantitative results in Table 2 indicate that the model achieves better structural alignment on the MNIST dataset, as reflected by a higher SSIM score (0.7448 vs. 0.5794). Although the FMNIST dataset exhibits a lower SSIM, it achieves a slightly higher PSNR (16.32 dB vs. 15.12 dB) and lower L2 Loss (0.0962 vs. 0.1295), suggesting smoother reconstructions with reduced noise and improved pixel-wise accuracy. However, the higher L1 Loss on FMNIST (0.1598) implies greater average absolute differences, possibly due to its higher visual complexity. Overall, the model performs better on MNIST in terms of structural preservation, while FMNIST reconstructions are numerically cleaner but less structurally accurate.

2.1.1 MNIST Results

Figures 1 and 2 illustrate example deblurring outputs and the test loss curve for the MNIST dataset.

Figure 1 presents qualitative deblurring results on the MNIST dataset. Each row displays a blurred input digit (left), the model's deblurred prediction (center), and the corresponding ground truth (right). The results demonstrate the model's ability to recover digit shapes with high perceptual similarity, even under strong blur. For example, the digits "8" and "5" are reconstructed with correct topology and detail. The displayed SSIM values (ranging from 0.7372 to 0.7976) and PSNR values (14.58 to 16.59 dB) confirm the structural and signal-level accuracy of the predictions. Low L1 and L2 losses further indicate strong pixel-wise reconstruction quality.

Figure 2 shows the test L1 and L2 loss trends over 10 training epochs. Both loss curves show a steady downward trend, with minor fluctuations. L1 loss decreases more consistently, while L2 loss initially rises slightly before converging. By epoch 9, both losses reach their lowest values (L1 $\approx$ 0.132, L2 $\approx$ 0.130), indicating that the model effectively learns to minimize reconstruction error. The convergence behavior suggests that even with limited training samples, the model generalizes well and avoids overfitting.

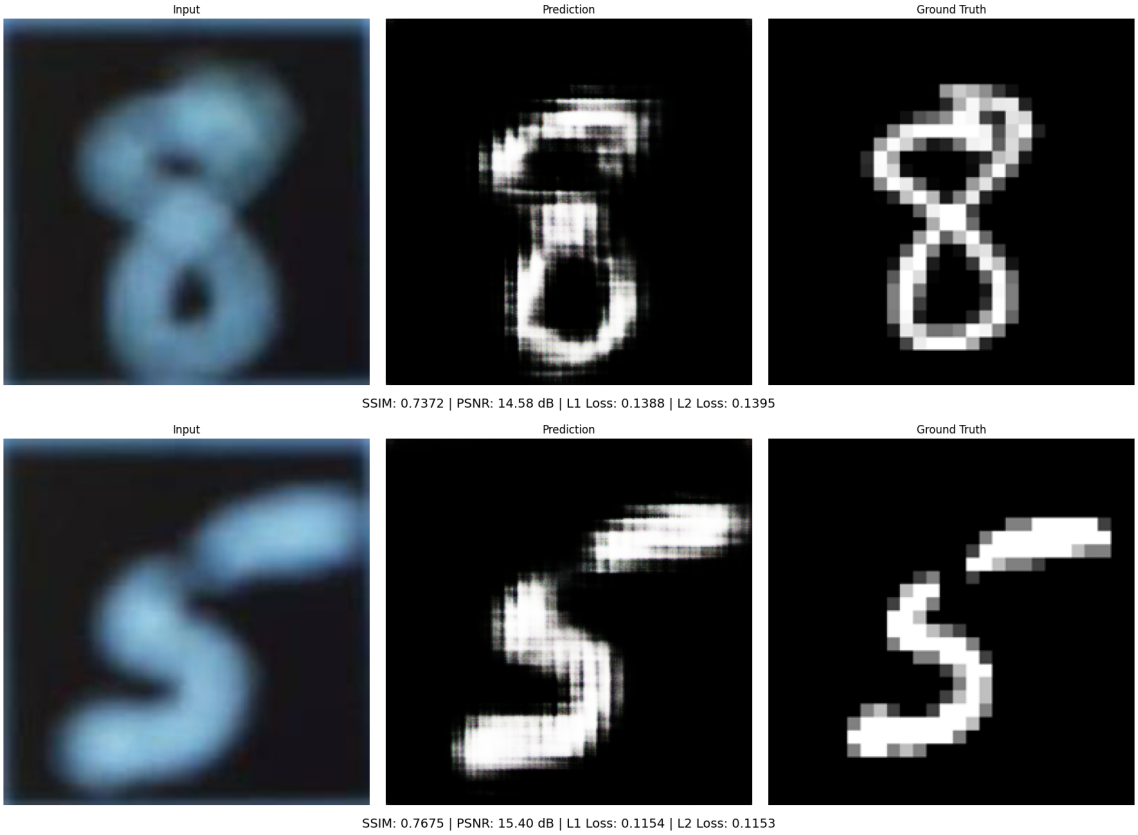

Figure 1: Example deblurring results on MNIST. **Left:** Input (blurred digit), **Center:** Model prediction (deblurred output), **Right:** Ground truth (original sharp digit).

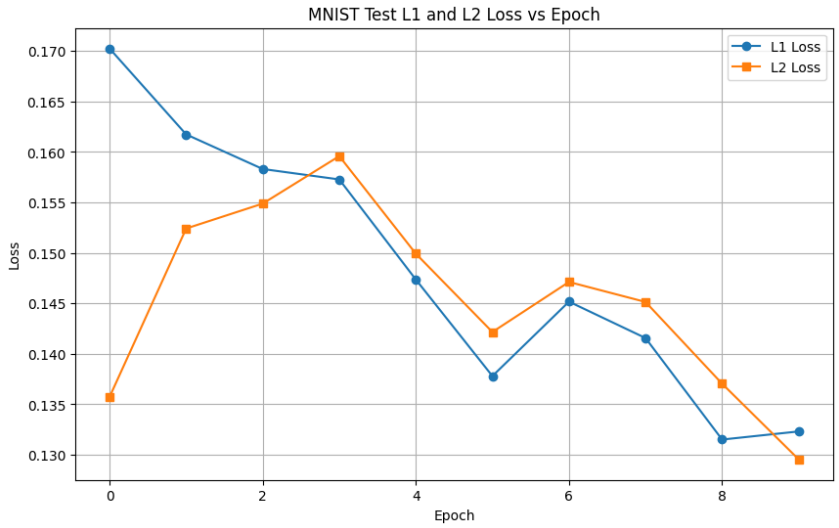

Figure 2: Test L1 loss and L2 loss versus training epochs on MNIST dataset. The model shows steady convergence within 10 epochs.

2.1.2 FMNIST Results

Figures 3 and 4 show the corresponding outputs and test loss curve for the FMNIST dataset.

Figure 3 presents deblurring results on the FMNIST dataset. The model effectively restores the general structure of fashion items from severely blurred inputs, although finer textures and details are not always preserved. For instance, the outlines of dresses and shirts are recovered, but high-frequency features (e.g., logos or sleeve edges) remain distorted. SSIM values range from 0.5552 to 0.6711, and PSNR values from 15.65 to 18.23 dB, reflecting reasonable visual quality. However, compared to MNIST, predictions show higher L1 losses (up to 0.1945), likely due to FMNIST's greater intra-class variation and structural complexity.

Figure 4 illustrates the L1 and L2 test losses over 10 epochs. The L1 loss shows a steep initial drop but remains noisier across epochs than MNIST, suggesting the model struggles more with consistent convergence on FMNIST. Nonetheless, the L2 loss remains relatively stable and low, indicating that the model still learns effective pixel-level mappings. Overall, the model generalizes well but exhibits more training variability on FMNIST due to its higher visual complexity.

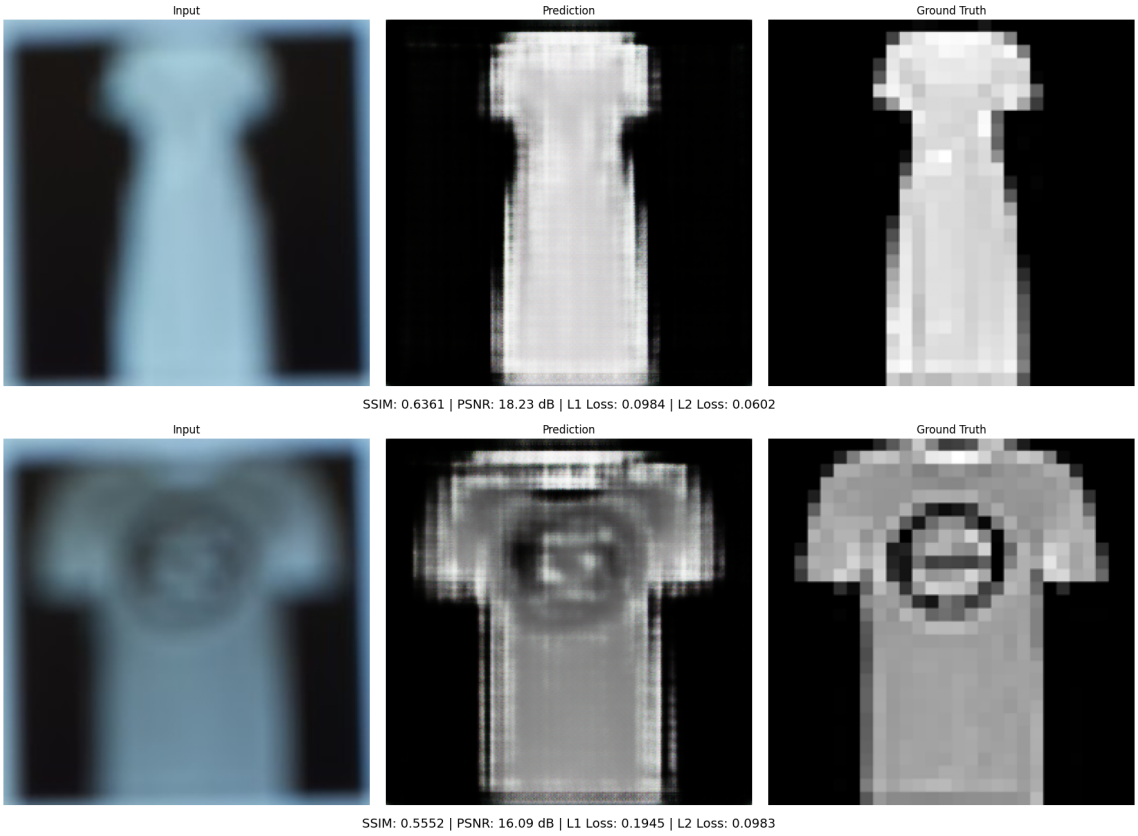

Figure 3: Example deblurring results on FMNIST. **Left:** Input (blurred item), **Center:** Model prediction (deblurred output), **Right:** Ground truth (original sharp item).

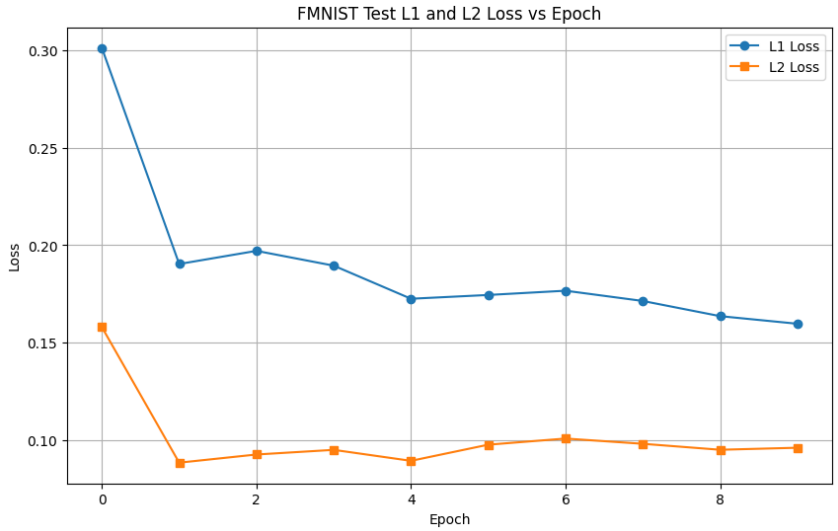

Figure 4: Test L1 loss versus training epochs on FMNIST dataset. While noisier, the model converges well within 10 epochs.

2.2 Robustness Evaluation

To assess the robustness of the trained hybrid CycleGAN-Pix2Pix model, we introduced Gaussian blur to the FMNIST test images with increasing levels of severity (standard de-

viation $\sigma \in \{0, 10, 20, 26, 30, 40, 50\}$). The goal was to evaluate the model's generalization ability to increasingly degraded inputs.

2.2.1 Visual Robustness Across Blur Levels

Figure 5 shows example outputs for FMNIST test samples across different levels of Gaussian blur. As the blur level increases, the model demonstrates decreasing recovery performance, but still manages to extract meaningful structures even at $\sigma = 50$.

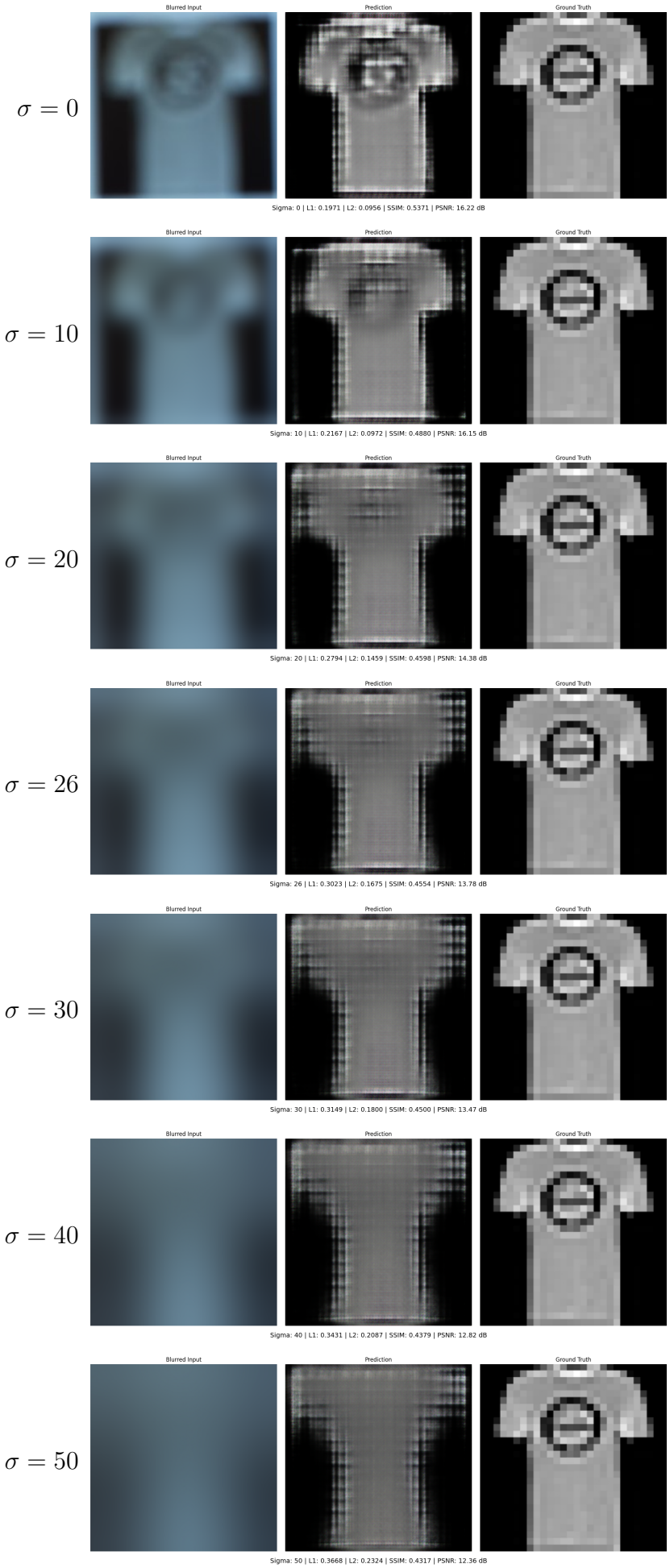

Figure 5: FMNIST deblurring predictions under increasing Gaussian blur levels. The model maintains partial structure even under extreme degradation.

As blur severity increases from $\sigma = 0$ to $\sigma = 50$, the model progressively struggles to recover fine details, particularly around edges and textures. Nonetheless, it maintains a recognizable global structure up to moderate blur levels.

At $\sigma = 10$, predictions retain sharp contours and preserve key shape features. By $\sigma = 20$, the outputs begin to lose edge clarity, though the overall silhouette remains intact. When $\sigma = 26$, the model reaches a "critical threshold": it can still infer the rough outline of the item, but high-frequency information such as internal patterns becomes indiscernible. Beyond this point ($\sigma \geq 30$), predictions show significant degradation, with outputs becoming increasingly uniform and losing semantic detail.

Despite the challenging conditions, the model shows a degree of robustness, with PSNR values remaining above 12 dB and SSIM values above 0.47 even under extreme blur. This highlights the model's ability to generalize structural priors from training data, though it becomes unreliable past a certain noise level.

2.2.2 Quantitative Robustness: Metrics vs Blur Level

We report PSNR, SSIM, L1, and L2 losses as functions of the blur level. Figures 6 and 7 illustrate the impact of increasing Gaussian blur (σ) on model performance using a representative FMNIST sample and the full test set, respectively.

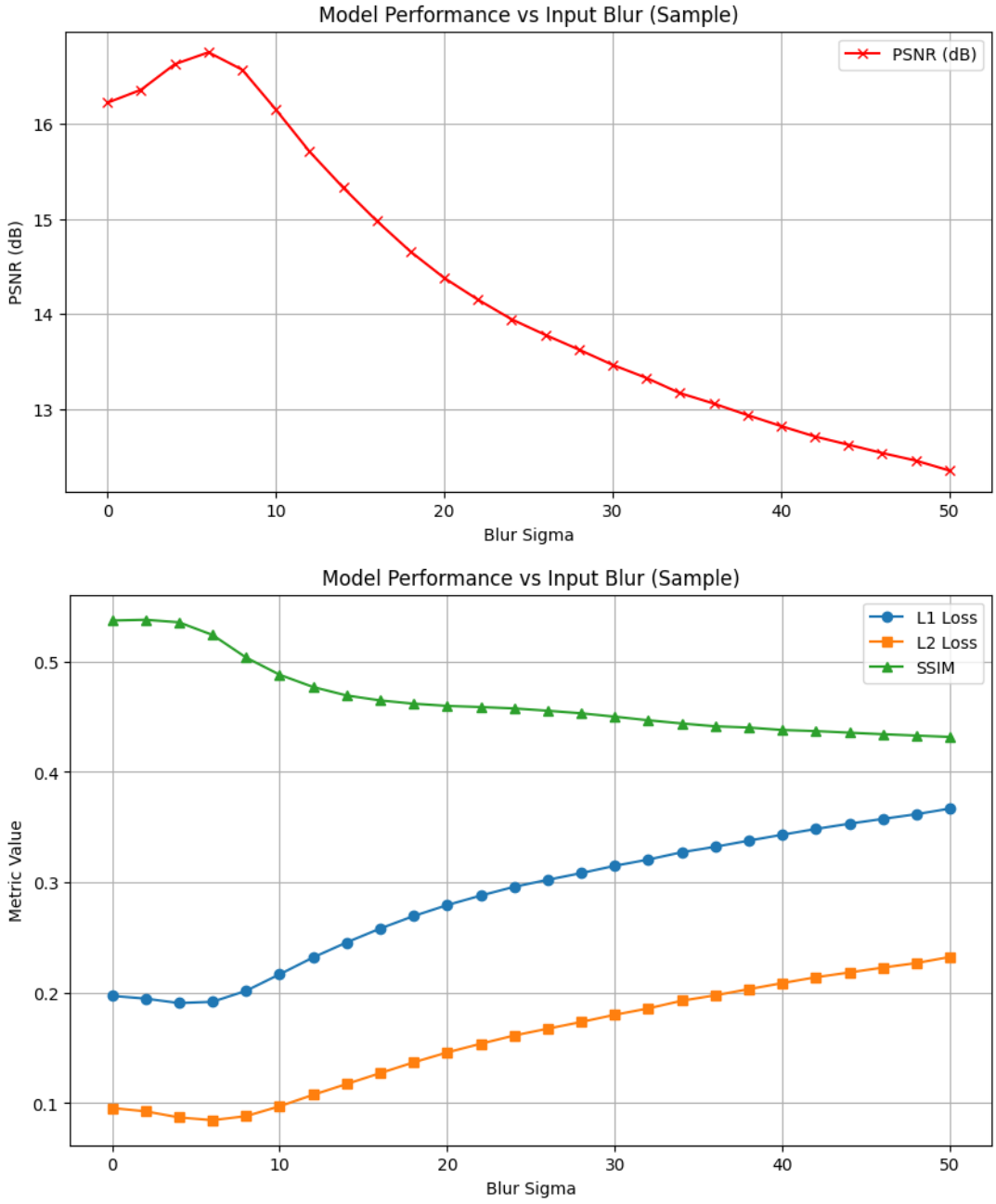

Figure 6: Robustness metrics for a representative FMNIST sample. **Top:** PSNR. **Bottom:** SSIM, L1 Loss, and L2 Loss. These plots highlight model behavior on individual input across increasing blur.

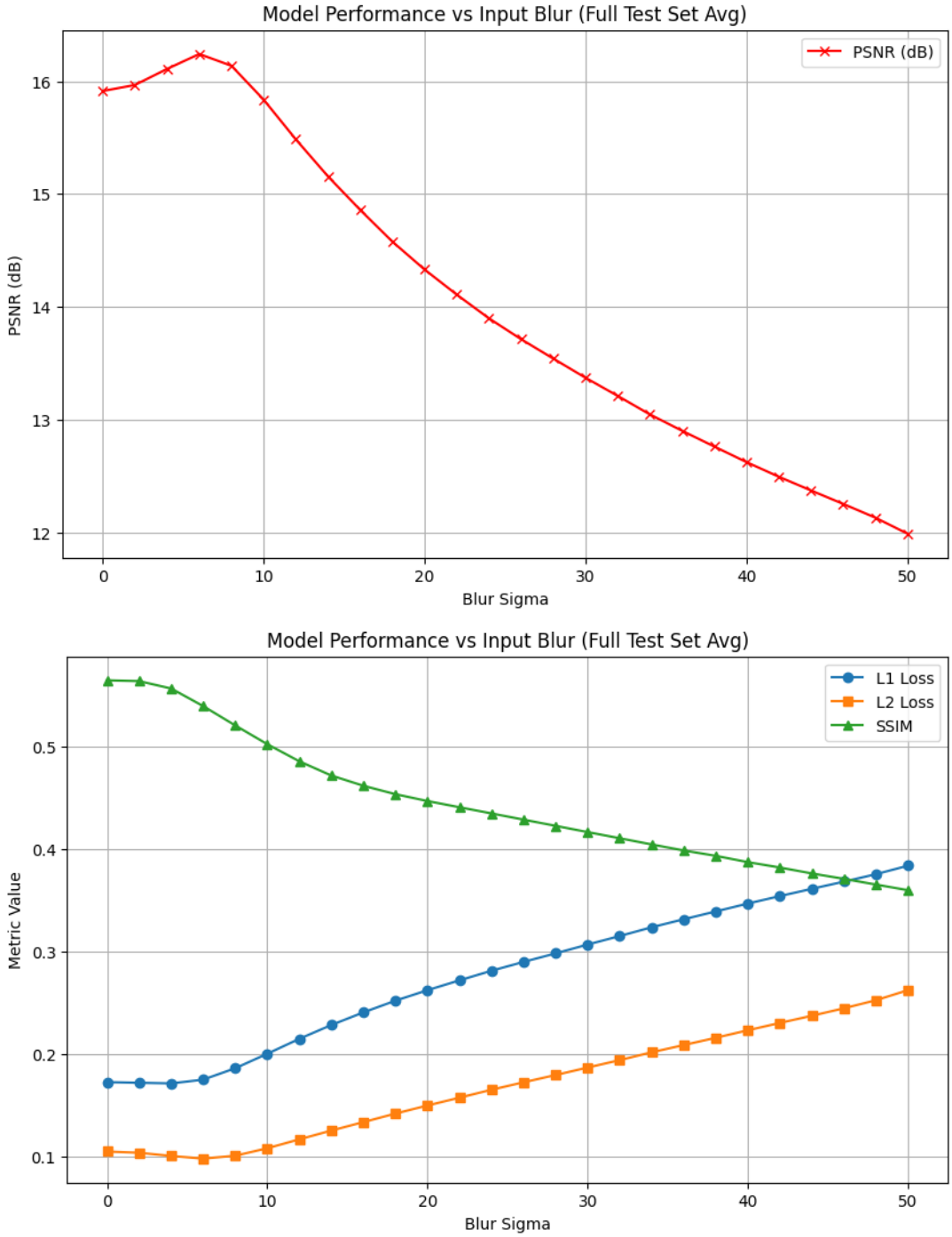

Figure 7: Average robustness metrics across the FMNIST test set as blur level increases. **Top:** PSNR. **Bottom:** SSIM, L1 Loss, and L2 Loss. As expected, PSNR and SSIM degrade with higher blur, while L1 and L2 losses increase.

In Figure 6, PSNR initially increases slightly for small blur levels ($\sigma \leq 5$), possibly due to minor noise removal. However, as σ increases beyond 10, PSNR steadily declines, and both L1 and L2 losses rise, reflecting degradation in reconstruction accuracy. SSIM shows a consistent downward trend, indicating loss of structural detail. This confirms that the model is sensitive to increasing blur and struggles to recover spatial features beyond moderate noise levels.

Figure 7 generalizes this behavior across the test set. The trends are consistent: PSNR

and SSIM degrade as blur increases, while L1 and L2 losses rise monotonically. Notably, SSIM drops more sharply between $\sigma = 0$ and $\sigma = 20$, suggesting early blur has the most significant impact on structural perception. Beyond this point, the decline becomes more gradual. The model exhibits graceful degradation, maintaining a predictable drop in performance without catastrophic failure, highlighting its robustness to moderate blur levels but limited capacity for extreme recovery.

3 Future Work

Building upon the current results, several directions have been identified to further extend and improve this project:

3.1 Integration of High-Resolution Datasets

Another key step is the integration of a **higher-resolution image dataset** to validate the scalability of our approach. Unlike the low-resolution digits and clothing items in MNIST and FashionMNIST, high-res datasets will present more complex textures, patterns, and features. We aim to benchmark the model's performance on this new dataset and investigate architectural adjustments or training optimizations (e.g., progressive resizing or multi-scale loss functions) that may be necessary for handling larger input dimensions.

3.2 Hardware Optimization and Deployment

In parallel with dataset expansion, we intend to explore optimization strategies for faster inference on the Raspberry Pi 5, including model quantization, pruning, and possibly the use of an external AI accelerator (such as Google's Coral TPU). These enhancements will support real-time deployment scenarios, especially for applications involving high-resolution input streams.

These future directions aim to enhance both the technical depth and the practical relevance of the project, ensuring that the system remains robust, portable, and adaptable across various image deblurring challenges.

References

[1] Hojjatkashani, H. (2017). *MNIST Dataset*. Kaggle. Retrieved April 26, 2025, from https://www.kaggle.com/datasets/hojjatk/mnist-dataset

[2] Xiao, H., Rasul, K., & Vollgraf, R. (2017). *Fashion-MNIST Dataset*. Kaggle. Retrieved April 26, 2025, from https://www.kaggle.com/datasets/zalando-research/fashionmnist

[3] Jain, N. (2020). *Image-Deblurring*. GitHub. Retrieved April 26, 2025, from https://github.com/navyajain16/Image-Deblurring

[4] xinhjing. (2025, April 29). *CompPhoto-Final-Project-pix2pix-cycleGAN* [Source code]. GitHub. Retrieved April 29, 2025, from https://github.com/xinhjing/CompPhoto-Final-Project-pix2pix-cycleGAN