Single pixel imaging with optical phase arrays: a machine learning assisted approach

Basak Ersoz[a], Paolo Bardella[a], Lorenzo Columbo[a], Marco Novarese[a], Maurizio Dabbicco[b], and Massimo Brambilla[c]

[a]Politecnico di Torino, Torino, Italy
[b]University of Bari Aldo Moro, Bari, Italy
[c]Politecnico di Bari, Bari, Italy

ABSTRACT

Speckle patterns used for imaging are generated with a single pixel detector employing optical phased arrays (OPAs). The OPA modulates light, which then travels through a diffraction grating, creating speckle patterns via various light paths. Once the target is illuminated, these patterns are captured by a bucket detector in single-pixel imaging, and the image is reconstructed using Compressed Sensing (CS) techniques.[1] In this work, we propose a Machine Learning (ML)[2] based approach to enhance the quality of CS reconstructions by comparing two popular architectures for image denoising: a basic Convolutional Neural Network (CNN) and a U-Net. A dataset of paired noisy CS target image and clean images was generated and used to train and evaluate both models. Quantitative results indicate that the U-Net significantly outperforms the CNN in terms of Mean Squared Error, Structural Similarity Index, and Peak Signal-to-Noise Ratio, demonstrating its effectiveness in maintaining structural information while reducing noise. This study opens the door the potential of ML in boosting reconstruction fidelity for single pixel imaging systems based on OPAs.

Keywords: optical phased arrays, speckle pattern imaging, single pixel algorithm, machine learning, convolutional neural network, u-net architecture

1. INTRODUCTION

Speckle patterns are created and exploited for imaging applications using a single pixel detector working together with OPAs.[3] The methodology is based on the fact that light modulated by the OPA and then propagating through a multimode fiber or scattered by a diffraction grating creates complex speckle patterns as a result of several light paths.[4] After illuminating the target, the intensity distribution of these speckle patterns can be detected by a bucket detector in single pixel imaging, and the image is then reconstructed using CS computational techniques.[5] The CS reconstructed target images frequently have high noise levels that may mask crucial target features. Conventional noise reduction techniques,[6] usually have trouble striking a compromise between spatial feature retention and noise reduction. ML models have recently become widely used for imagine reconstruction tasks because they provide notable gains in efficiency and quality[2].[7]

Combining CS and OPA illumination with a single pixel detector minimizes the need for complex hardware while still capturing the necessary data for image reconstruction.[8] Additionally, we show that ML integration enhances the quality of CS reconstructed images, making the final image much more accurate, thus improving upon the limitations of traditional CS techniques. This approach is particularly advantageous for MIR where detector technology is less advanced and the detectors are more bulky and more expensive with respect to those available in the visible and near infrared region of the electromagnetic spectrum.

In this paper, the effectiveness of the CNN and U-Net architecture[9] for denoising is compared. Because of its encoder-decoder structure with skip connections, the U-Net performs in preserving fine grained details, whereas the CNN offers simplicity and faster training. This study highlights the advantages and disadvantages of both topologies by evaluating them on a shared dataset.

Further author information: (Send correspondence to Paolo Bardella)
P. Bardella: E-mail: paolo.bardella@polito.it

2. METHODS

2.1 Dataset Preparation

In our work, a dataset consisting of couples of clean target images and their corresponding noisy CS reconstructions are simulated using a MATLAB algorithm. Both the original clean target images and noisy CS target images were resized to 128×128 pixels using image processing functions to match the input size required by the models. The images were converted to grayscale and normalized by scaling pixel values to a range between 0 and 1. This preprocessing ensured consistency in the dataset, making it suitable for training and evaluation of the CNN and U-Net models.

2.2 CNN Architecture

The CNN model was created using the basic encoder-decoder configuration. The decoder uses a single convolutional layer with a sigmoid activation function to reconstruct the image, while the encoder uses a series of convolutional layers with decreasing filter sizes ($32 \rightarrow 16 \rightarrow 8 \rightarrow 4$). With a learning rate of 0.001, the model is improved using the Adam optimizer to minimize the pixel-wise Mean Squared Error (MSE). With a batch size of 16, the model was trained for five epochs.

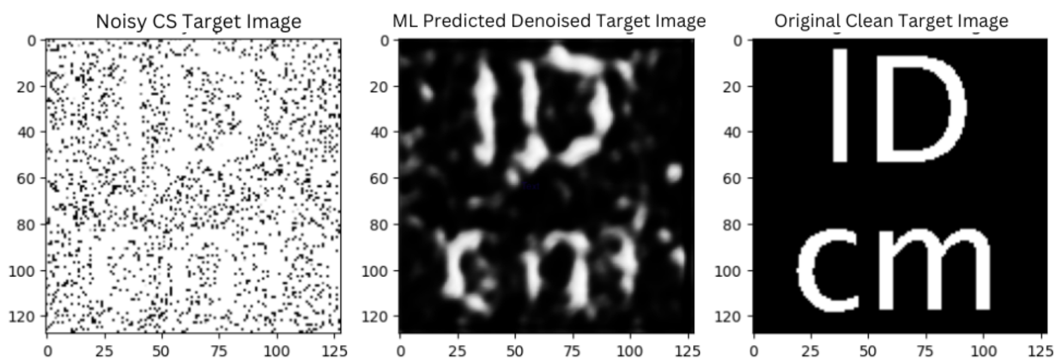

Figure 1. Illustrated results of the CNN model showing noisy inputs, predicted clean images, and the original clean ground truth.

Fig. 1 highlights the differences between the architectures. The CNN provides a smooth denoising effect but CNN is unable to recreate complex patterns.

2.3 U-Net Architecture

The U-Net model[9] uses bypass connections between corresponding layers and a symmetric encoder-decoder structure. Convolutional layers ($64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1024$ filters) are set up with max-pooling layers to extract hierarchical features. This structure is mirrored by the decoder, which uses transpose convolutions for upsampling and concatenating encoder features to maintain spatial information. A grayscale image with pixel values limited to [0, 1] via a sigmoid activation function is the model's output. The Adam optimizer was used to train it for 100 epochs with a batch size of 256 and a learning rate of 0.0001.

In Fig. 2, the U-Net output maintains both structural integrity and visual clarity by matching closely the ground truth.

2.4 The Evaluation Metrics

The following measures were used to assess the model's performance:

a. Mean Squared Error (MSE): Quantifies the average squared difference between predicted and ground truth pixel values. **b. Structural Similarity Index (SSIM):** Measures the structural similarity between images, capturing perceptual quality. **c. Peak Signal-to-Noise Ratio (PSNR):** Assesses image quality based on the ratio of signal power to noise power.

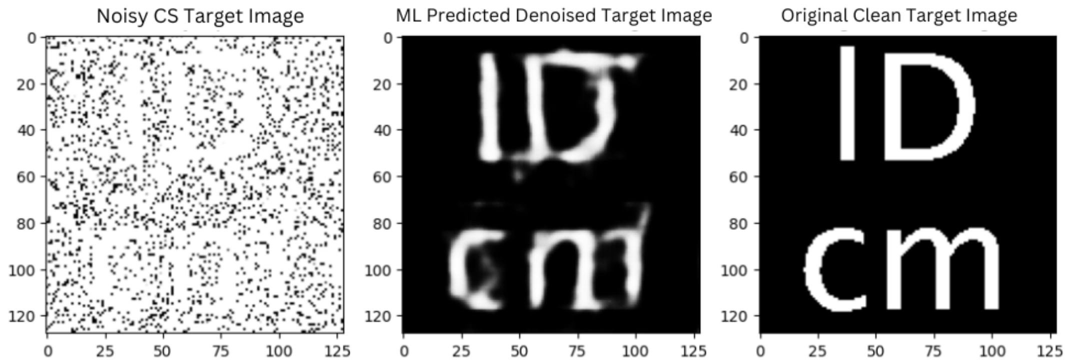

Figure 2. Illustrated results of the U-Net model showing noisy inputs, predicted clean images, and the original clean ground truth. The U-Net demonstrates a smooth denoising effect but struggles to reconstruct fine details.

3. RESULTS

The training set size of 750, and the validation set that includes 250 noisy-clean image pairs was used to evaluate the performances of CNN and U-Net models. Quantitative results, including MSE, SSIM, and PSNR, are summarized in Table 1. and illustrated results are shown in Fig. 1 and 2, showing noisy CS target images inputs, ML predicted denoised target image outputs, and original clean target images.

Table 1. Comparison of Performance Metrics for CNN and U-Net Architectures

Metric	CNN	U-Net
MSE	0.055	0.025
SSIM	0.32	0.75
PSNR (dB)	12.6	15.9

As seen by its lower SSIM and PSNR values, the CNN struggled to preserve small details despite demonstrating faster training and inference durations. But because the U-Net was able to preserve spatial features through skip connections, it performed better on all criteria.

4. CONCLUSION

This study highlights the potential of deep learning in photonics by applying CNN and U-Net architectures for denoising complex pattern images. The CNN model, with its simple encoder-decoder design, is computationally efficient but struggles with fine detail reconstruction. In contrast, the U-Net achieves very good results, effectively preserving structural integrity and delivering higher accuracy in terms of SSIM and PSNR. Additionally, each model can be enhanced by using their optimized version.

The findings show that U-Net could promise better performance in photonics applications, such as optical imaging, which demand high quality image reconstruction.[10] This method provides scalable and effective solutions for future photonic systems, demonstrating the growing significance of machine learning in improving imaging techniques.

ACKNOWLEDGMENTS

This study was carried out within the MId infRAred laBel free Interferometric detectorLess Imaging photonic circuitS - MIRABILIS project (CUPM. D53D23002780006) – funded by European Union – Next Generation EU within the PRIN 2022 program (D.D. 104 - 02/02/2022 Ministero dell'Università e della Ricerca). This manuscript reflects only the authors' views and opinions and the Ministry cannot be considered responsible for them.

REFERENCES

[1] Fukui, T., Kohno, Y., Tang, R., Nakano, Y., and Tanemura, T., "Single-pixel imaging using multimode fiber and silicon photonic phased array," *Journal of Lightwave Technology* **39**(3), 839–844 (2021).

[2] Savio, D. and Bardella, P., "Deep-learning assisted control of optical phased array: A case study," *IEEE Access* **10**, 15421–15426 (2022).

[3] Guerber, S., Fowler, D., Charlet, I., Grosse, P., Abdoul-Carime, K., Faugier-Tovar, J., and Szelag, B., "Development, calibration and characterization of silicon photonics based optical phased arrays," in [*Smart Photonic and Optoelectronic Integrated Circuits XXIII*], He, S. and Vivien, L., eds., **11690**, 1169006, International Society for Optics and Photonics, SPIE (2021).

[4] Acoleyen, K. V., Bogaerts, W., Jágerská, J., Thomas, N. L., Houdré, R., and Baets, R., "Off-chip beam steering with a one-dimensional optical phased array on silicon-on-insulator," *Opt. Lett.* **34**, 1477–1479 (May 2009).

[5] Fukui, T., Kohno, Y., Tang, R., Nakano, Y., and Tanemura, T., "Single-pixel imaging using multimode fiber and silicon photonic phased array," *Journal of Lightwave Technology* **39**(3), 839–844 (2021).

[6] Choi, H. and Jeong, J., "Speckle noise reduction in ultrasound images using srad and guided filter," in [*2018 International Workshop on Advanced Image Technology (IWAIT)*], 1–4 (2018).

[7] Kim, K., Wu, D., Gong, K., Dutta, J., Kim, J. H., Son, Y. D., Kim, H. K., El Fakhri, G., and Li, Q., "Penalized pet reconstruction using deep learning prior and local linear fitting," *IEEE Transactions on Medical Imaging* **37**(6), 1478–1487 (2018).

[8] Dabbicco, M., Columbo, L., Ersoz, B., Brambilla, M., and Bardella, P., "Mid infrared label free interferometric detectorless imaging photonic circuit," in [*2024 IEEE 14th International Conference Nanomaterials: Applications Properties (NAP)*], 1–5 (2024).

[9] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*], 234–241, Springer (2015).

[10] Li, Y., Yu, Z., Chen, Y., He, T., Zhang, J., Zhao, R., and Xu, K., "Image reconstruction using pre-trained autoencoder on multimode fiber imaging system," *IEEE Photonics Technology Letters* **32**(13), 779–782 (2020).