Learning-Based Image Reconstruction for Spatial-Variant Single-Pixel Imaging

Zhenyong Shin, Chang Hong Pua, Tong-Yuen Chai, Xin Wang, and Sing Yee Chua

Original scientific article

***Abstract*—Improving single-pixel imaging efficiency can be achieved through spatially-variant resolution (SVR) sensing patterns, which adaptively adjust resolution to enhance image acquisition. This paper proposes a convolutional neural network (CNN) architecture specifically designed for SVR-based single-pixel imaging with compressed sensing (CS) as a more efficient and non-iterative approach to image reconstruction. The results show that the combination of SVR sensing patterns and the proposed CNN model outperforms the uniform resolution (UR) sensing patterns in terms of image reconstruction quality. Furthermore, the CNN-based approach achieves greater time efficiency compared to established methods such as ReconNet and TVAL3, thus reducing the overall computational load without compromising output image quality. These findings highlight the potential of the proposed learning-based SVR approach to effectively balance reconstruction accuracy and processing speed in single-pixel imaging. The study optimizes both the image acquisition and reconstruction process in single-pixel imaging, making it suitable for real-time applications that require rapid imaging capabilities while maintaining high image quality.**

***Index Terms*—Single-Pixel Imaging, Spatially-Variant Resolution, Compressed Sensing, Convolutional Neural Network.**

I. Introduction

Charge-coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) technologies have drastically improved imaging sensor technologies over the last few decades. However, the cost and efficiency of image sensors remain a challenge in invisible wavelengths and low light conditions [1]. Single-pixel imaging techniques were developed to mitigate the aforementioned challenges. It has been applied in numerous areas, such as multispectral imaging [2], video imaging [3], infrared imaging [4], etc. In recent years, the development of single-pixel imaging has moved from 2-D to 3-D imaging [5], [6]. In essence, a single-pixel imaging system captures the correlation between a target scene and a set of sensing patterns. First, the measurements that contain the encoded image data are recorded with a single-pixel detector. Then, an image of the target scene can be reconstructed with an image reconstruction algorithm such as l_1 minimization and total variation regularization [7].

Fundamentally, image sampling adheres to the Nyquist-Shannon Theorem. It states that a signal must be sampled at a sampling rate above the Nyquist rate, i.e., no less than two times the signal's highest frequency, to permit error-less signal reconstructions. For 2-D images, the number of measurements must be no less than the number of image pixels [8]. In fact, the number of measurements is directly proportional to the costs and time required for image data acquisition and processing. These factors prevent single-pixel imaging from being more widely applicable. Fortunately, single-pixel imaging works well with compressed sensing (CS) to achieve error-less image reconstruction with only partial measurements, provided that the images are sparse [9], [10]. Furthermore, many techniques such as adaptive foveated imaging [11] and block-based single-pixel imaging [12], [13] have also been proposed to reduce the number of measurements required and computational complexity.

In the single-pixel imaging with CS framework, image information is acquired in its compressed form through modulations with a sensing matrix. The design of the sensing matrix must satisfy certain properties required by CS, such as the restricted isometry property (RIP) [14]. Numerous studies have found that both the design of the sensing matrices and the ordering of their rows affect the quality of the reconstructed images [15], [16]. In general, randomly generated sensing matrices are the more favorable choices due to their higher probability of possessing the properties imposed by CS [9], [17]. Recently, deterministic sensing matrices such as Hadamard and Fourier have become more favorable due to the ease of implementation in practice [18], [19], [20].

In addition to sensing matrices, image reconstruction algorithms have also been studied extensively. The evolution of machine learning, specifically deep learning, has improved the efficiency of image reconstruction. Numerous studies have shown that convolutional neural networks (CNNs) is capable of improving image quality and reducing time complexity [21], [22], [23]. Recently, generative adversarial networks (GANs) have also been adapted to the CS framework [24], [25].

Reducing the number of measurements and time complexity for image reconstruction while retaining image quality are the primary goals in single-pixel imaging. The authors' previous work [26], [27] shows that one can program a spatially-variant resolution (SVR) sensing pattern to alter an image's resolution.

Manuscript received February 6, 2025; revised March 29, 2025. Date of publication July 10, 2025. Date of current version July 10, 2025. The associate editor prof. Maja Braović has been coordinating the review of this manuscript and approved it for publication.

Z. Shin was with Universiti Tunku Abdul Rahman (UTAR), Malaysia and is currently with the University of Malaya.

C. H. Pua and S. Y. Chua (corresponding author) are with the Lee Kong Chian Faculty of Engineering and Science (LKC FES), Universiti Tunku Abdul Rahman (UTAR), Malaysia (e-mail: sychua@utar.edu.my).

T.-Y. Chai is with the Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman (UTAR), Malaysia.

X. Wang is with the Monash University Malaysia.

Digital Object Identifier (DOI): 10.24138/jcomss-2025-0011

This allows a decent image to be reconstructed from fewer measurements compared to uniform resolution (UR) images. To further improve image reconstruction efficiency, this paper proposes an architecture of CNN inspired by ReconNet [21] to work with SVR sensing patterns. To the authors' knowledge, no previous publications have implemented SVR sensing matrices in combination with CNN. The proposed learning-based SVR approach is expected to further improve single-pixel imaging both in terms of image quality and time efficiency. This paper's results show that SVR sensing matrices outperform UR in the single-pixel image reconstruction using the proposed CNN. Furthermore, the proposed method outperforms other reconstruction algorithms in terms of time efficiency.

The contributions of this paper include:

- Introducing a CNN architecture specifically designed for SVR-based single-pixel imaging, enabling efficient, non-iterative CS image reconstruction.
- Enhancing computational efficiency over ReconNet and TVAL3, making it more suitable for real-time imaging applications.
- Improving both image acquisition and reconstruction processes, achieving an optimal balance between accuracy and processing speed.

This paper is organized as follows. Section II gives the background of single-pixel imaging, CS, and CNN for image reconstruction. Section III describes the SVR scheme and the architecture of the proposed CNN. Section IV presents the results and discussion. Finally, Section V concludes the paper.

II. Background

As illustrated in Figure 1, a typical single-pixel imaging system consists of a digital micro-mirror device (DMD) with a light source, a single-pixel detector, and some lenses. The general process of capturing images with the single-pixel imaging system is as follows: First, a set of sensing patterns is uploaded to the DMD. The DMD is a device that consists of a 2-D array of micro-mirrors. Each micro-mirror can be programmed to orient towards different directions independently so that the light generated by the light source can be deflected towards various directions. By altering the orientations of the micro-mirrors, different designs of sensing patterns can be produced consecutively. Then, the single-pixel detector measures the overall intensity of the light deflected by the DMD towards it and saves the measurements. This completes the image acquisition process.

The single-pixel imaging with CS framework essentially performs both image acquisition and compression at the same time [17], [29]. Let $\mathbf{x}$ be the original image. The measurement vector of $\mathbf{x}$, denoted as $\mathbf{y}$, is acquired with a sensing matrix $\phi \in R^{M\times N} : M, N \in N, M < N$ as

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \phi\mathbf{x} \in R^M. \tag{1}$$

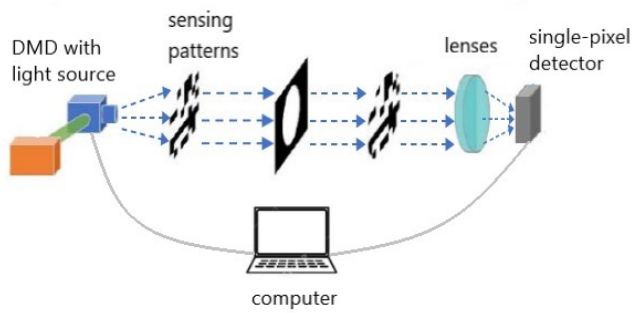


Fig. 1. Illustration of a typical single pixel imaging system [28].

where the m-th measurement is denoted by y_m. If $\mathbf{x}$ is represented by some sparsifying basis, denoted by ψ, such that $\mathbf{x} = \psi\alpha$, then

$$\mathbf{y} = \phi\mathbf{x} = \phi\psi\alpha = \Theta\alpha \tag{2}$$

where $\Theta = \phi\psi$ and α is the coefficient vector. In reality, measurements $\{y_m\}_{m=1}^{M}$ are acquired by projecting a set of sensing patterns onto the target scene and recording the residual light intensities. By reshaping a row of ϕ as a $\sqrt{N}\times\sqrt{N}$ matrix, one can obtain a sensing pattern. A visual representation of Equation 2 is shown in Figure 2.

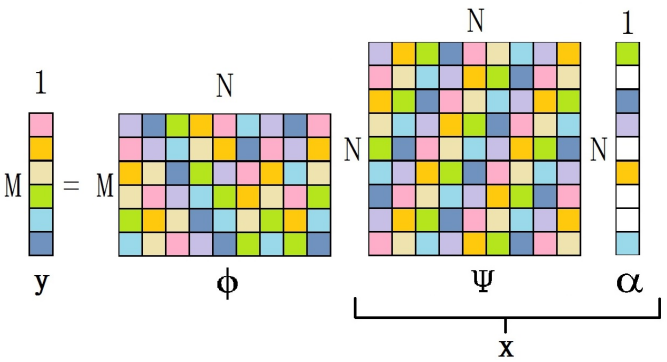


Fig. 2. Illustration of CS Equation 2.

Equation 1 is an under-determined linear system since $M < N$. Hence, for any given $\mathbf{y}$, the number of possible solutions $\mathbf{x}'$ such that $\mathbf{y} = \phi\mathbf{x}'$ is infinite, and obtaining a unique solution to Equation 1 is extremely challenging. However, since CS requires that $\mathbf{x}$ is sparse, one only needs to focus on the set $\{\mathbf{x}' : \mathbf{x}' \text{ is sparse}\}$. A unique and sparse solution $\mathbf{x}'$ to Equation 1 can be determined using l_1 minimization [8], [17]:

$$\hat{\mathbf{x}}_{l_1} = argmin||\mathbf{x}'||_1 \quad \text{subjected to} \quad \phi\mathbf{x}' = \mathbf{y}. \tag{3}$$

Various CS image reconstruction algorithms, such as the convex optimization method [14] and iterative greedy algorithms [30] have been developed and shown to successfully reconstruct error-free images [31]. Moreover, a small number of CS measurements can be discarded without affecting the image reconstruction. This permits a better reconstruction when more measurements are obtained. Finally, only the image reconstruction process of CS is computationally complex, which generally possesses more computational resources than the measurement acquisition process [32].

As opposed to CNN developed for conventional computer vision tasks, CNN developed for CS uses the CS measurements as the inputs to the network instead of images. Essentially, the CNN maps a low-dimension vector to a high-dimension vector. Such an approach was inspired by the CNN developed for image super-resolution [33]. Kulkarni et al. [21] proposed a CNN-based architecture called ReconNet, which improved the reconstruction results in terms of peak signal-to-noise ratio and time complexity as compared to iterative CS reconstruction algorithms. Shi et al. [22] proposed CSNet, which includes a sampling and reconstruction network. It was shown that the learned sampling matrices can improve the traditional image CS reconstruction.

III. Methodology

The proposed method is based on the programmable SVR scheme to alter the image resolution, which allows image reconstruction with fewer measurements. To enhance the efficiency of this reconstruction process, a CNN architecture is proposed to work in conjunction with the SVR sensing patterns.

A. Spatially-Variant Single-Pixel Imaging

SVR sensing patterns consist of various-sized pixels called cells. The authors' previous work demonstrates that SVR sensing patterns outperform UR sensing patterns. In this paper, chaotic pattern arrays (CPA) are used as the underlying sensing patterns. The superiority of CPA over other sensing patterns had been shown in the authors' prior work [27]. In accordance with the findings, $z_0 = 0.19$ is used for the generation of CPA in this paper.

Figure 3 shows the process of generating UR and SVR sensing patterns. A SVR sensing matrix ϕ_{SVR} is generated as

$$\phi^T_{SVR,m} = \tau \phi^T_m, \tag{4}$$

where τ is an $N \times N$ binary transformation matrix with entries 0 and 1, and $\phi^T_{SVR,m}$ and ϕ^T_m are the transpose of the m-th row of ϕ_{SVR} and ϕ, respectively.

As the CPA sensing matrices contain only -1 and +1, the sensing matrices are divided into two distinct sets of complementary sensing matrices $\phi_+ = \frac{1}{2}(\mathbf{1} + \phi)$ and $\phi_- = \mathbf{1} - \phi_+$, where $\mathbf{1}$ is a $M \times N$ matrix whose entries are 1. Next, the rows of ϕ_+ and ϕ_- are reshaped into two sets of complementary sensing patterns. Since there are two sets of complementary sensing matrices, there are two measurement vectors $\mathbf{y}_+ = \phi_+ \mathbf{x}$ and $\mathbf{y}_- = \phi_- \mathbf{x}$. Lastly, the CS measurements are obtained as $\mathbf{y} = \frac{1}{2}(\mathbf{y}_+ - \mathbf{y}_-)$. In such a case, the sensing matrix is replaced by $\hat{\phi} = \frac{1}{2}(\phi_+ - \phi_-)$, and Equation (1) becomes

$$\frac{1}{2}(\mathbf{y}_+ - \mathbf{y}_-) = \frac{1}{2}(\phi_+ - \phi_-)\mathbf{x}. \tag{5}$$

Figure 4 illustrates the workflows of UR and SVR image reconstruction. This paper compares the performance of UR and SVR sensing patterns and the performance of the proposed CNN with ReconNet [21] and TVAL3 [34]. BM3D [35] denoising algorithm is used to remove the edge effect caused by the image blocks.

B. The Proposed Convolutional Neural Network

Figure 5 shows the architecture of the proposed CNN. The image block size is set to 32×32 [27], [13]. The first layer of the proposed CNN is a fully-connected (FC) layer that receives a CS measurement $\mathbf{y}$ as input and returns a 1024×1 feature map as output. The second layer reshapes the feature map returned by the FC layer to a 32×32 feature map. The third and seventh layers are 2-D convolutional layers that use a 3×3 kernel and generate 64 feature maps. The fourth, fifth, eighth, and ninth layers are 2-D convolutional layers that use a 1×1 kernel and generate 32 feature maps. The sixth and the final layers are 2-D convolutional layers that use a 3×3 kernel and generate a feature map. All convolutional layers use the Rectified Linear Unit (ReLU) as an activation function except for the final layer, which uses the sigmoid function. The output of the ninth layer is the reconstructed image $\hat{\mathbf{x}}$. Zero padding is used to keep a constant feature map size across all layers.

The simulations and network training were performed on an Intel Core i5-7200U 3.1 GHz laptop with 4GB of RAM. The image dataset CIFAR-10 [36] was used in the network training, with all images converted to grayscale prior to processing. A sensing matrix ϕ is generated for each sensing ratio, $SR = M/N$. This paper selects SR values of 0.01, 0.04, and 0.10 in accordance with [21]. A set of CS measurements of the training images is obtained before training the network. This set of CS measurements is the input of the proposed CNN. The optimizer used in the network training was Adadelta with a default learning rate of 0.001. The batch size was set to 16 and the number of training epochs was 5000. Mean squared error (MSE) was used as the loss function.

IV. Results and Discussions

A. SVR versus UR Sensing Patterns using the Proposed CNN

Table I shows the mean structural similarity index measure (SSIM) and image reconstruction time with standard deviation $(n = 20)$ for SVR and UR images reconstructed using the proposed CNN. As observed, SVR images consistently show higher structural similarity compared to UR images, indicating that the CNN effectively preserves image details during the reconstruction process. This aligns with the authors' expectations, as SVR images typically exhibit greater spatial consistency due to their higher effective resolution, which the proposed CNN can leverage to enhance image quality. This improvement in SSIM could be beneficial for applications where fine image details are crucial, such as in medical imaging. On the other hand, the mean reconstruction times for SVR and UR images do not show a significant difference. This indicates that the proposed CNN processes both types of images with comparable efficiency.

Figure 6 shows the ground truths and images reconstructed using the proposed CNN with $SR = 0.10$ for SVR and UR sensing patterns. The results show that SVR outperforms UR sensing patterns. Visually, the image details of images reconstructed with SVR sensing patterns appear clearer. This observation, together with the quantitative results shown in Table I, shows that SVR sensing patterns performed better than UR sensing patterns using the proposed CNN.

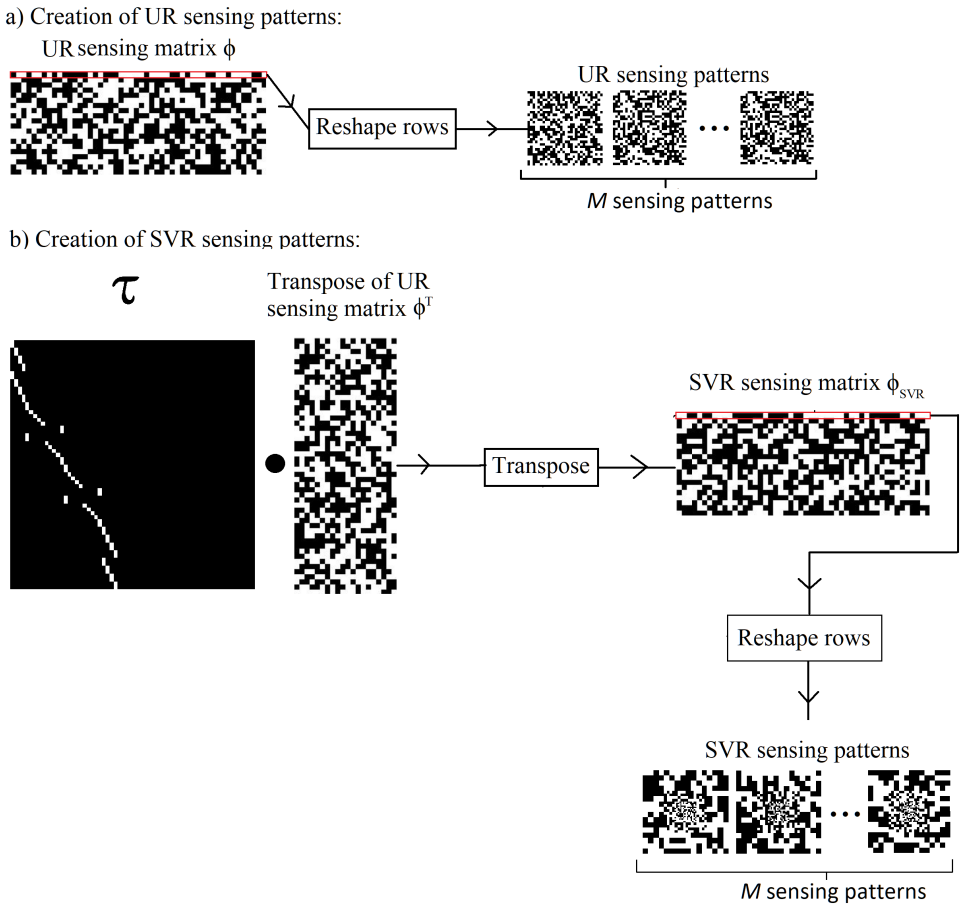


Fig. 3. The process of creating a set of (a) UR and (b) SVR sensing patterns [27]. The black and white pixels are entries 0 and 1, respectively.

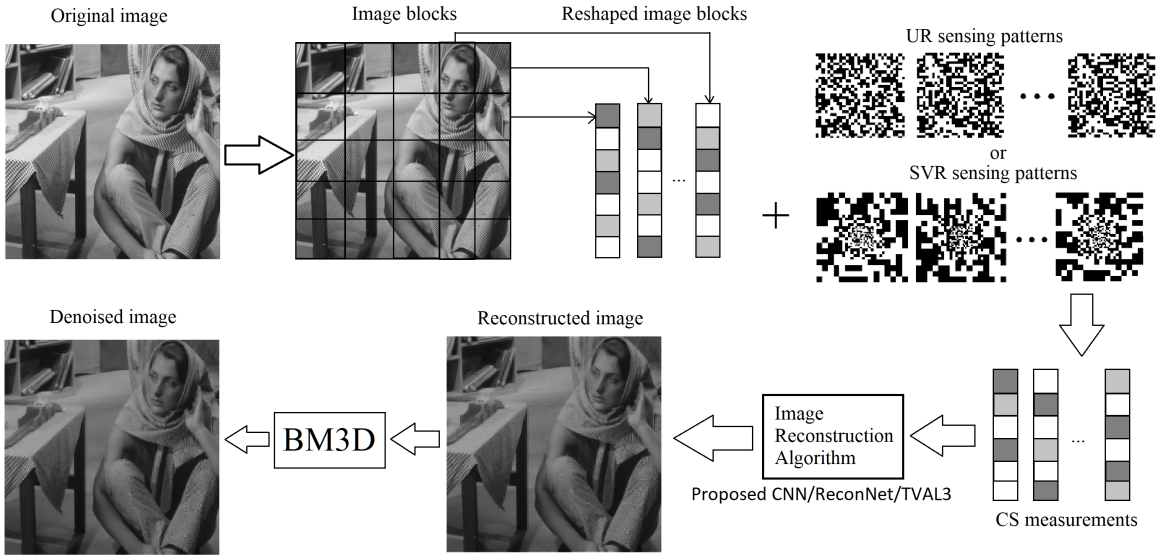


Fig. 4. Workflows of UR and SVR image reconstruction.

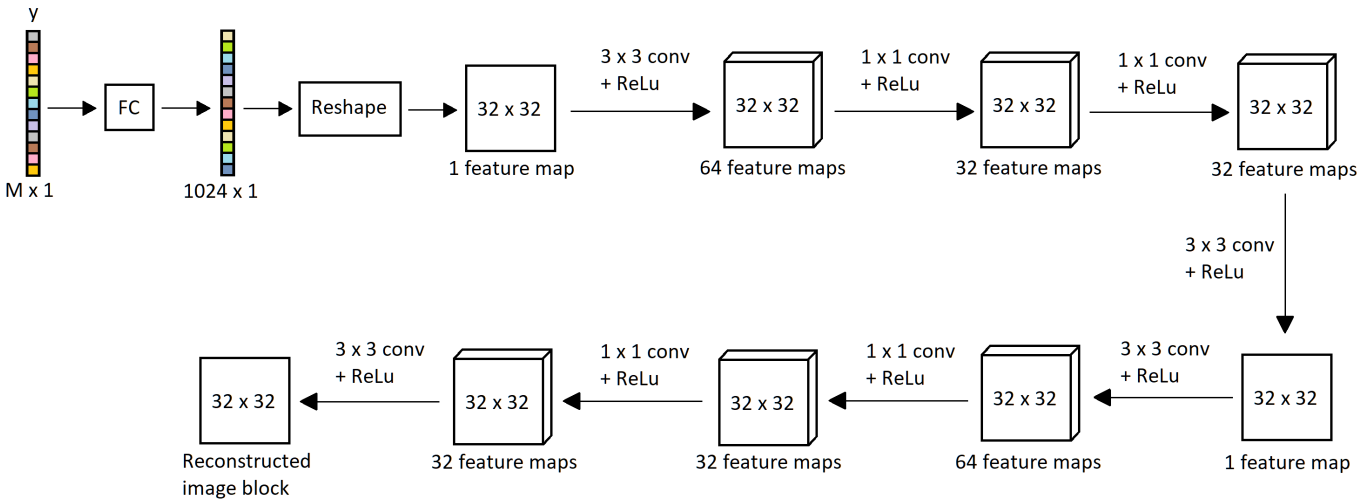


Fig. 5. Architecture of the proposed CNN.

TABLE I
THE MEAN SSIM AND IMAGE RECONSTRUCTION TIME (s) WITH STANDARD DEVIATION $(n = 20)$ FOR IMAGES RECONSTRUCTED USING THE PROPOSED CNN, FOR SVR AND UR FRAMEWORKS.

	SR=0.10							
	SVR				UR			
	SSIM		Time		SSIM		Time	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
airplane	0.786	1.14E-16	2.7883	0.0229	0.774	3.42E-16	2.8640	0.0691
baboon	0.454	2.85E-16	2.7258	0.0280	0.416	1.14E-16	2.8565	0.0564
Barbara	0.659	0.00E+00	2.7242	0.3782	0.624	3.42E-16	2.7766	0.0599
boat	0.652	2.28E-16	2.8242	0.1009	0.614	1.14E-16	2.8043	0.0407
cat	0.675	0.00E+00	2.9250	0.2011	0.643	1.14E-16	3.3482	0.3592
fruits	0.744	2.28E-16	2.6539	0.1491	0.695	0.00E+00	2.7347	0.0555
goldhill	0.651	0.00E+00	2.6484	0.0927	0.623	0.00E+00	2.8103	0.0626
cameraman	0.819	0.00E+00	2.9773	0.1119	0.776	2.28E-16	2.7008	0.0634
peppers	0.800	2.28E-16	2.5500	0.1146	0.779	2.28E-16	3.1526	0.3682
Zelda	0.786	1.14E-16	2.5359	0.0841	0.764	1.14E-16	3.4967	0.5790
	SR=0.04							
	SVR				UR			
	SSIM		Time		SSIM		Time	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
airplane	0.727	1.14E-16	2.7336	0.0206	0.659	1.14E-16	2.7951	0.0610
baboon	0.364	1.14E-16	2.7031	0.1374	0.324	5.70E-17	2.7556	0.0551
Barbara	0.594	1.14E-16	3.2742	0.8833	0.542	1.14E-16	2.7083	0.0603
boat	0.559	0.00E+00	2.8508	0.3604	0.506	1.14E-16	2.7472	0.0645
cat	0.620	1.14E-16	3.0305	0.3245	0.570	2.28E-16	3.2516	0.4793
fruits	0.609	2.28E-16	2.5938	0.1559	0.552	1.14E-16	2.7466	0.1132
goldhill	0.572	1.14E-16	2.6344	0.1508	0.503	0.00E+00	2.9137	0.1700
cameraman	0.701	2.28E-16	2.6984	0.0362	0.660	1.14E-16	2.6469	0.0440
peppers	0.732	2.28E-16	2.5211	0.1389	0.684	0.00E+00	3.4001	0.5367
Zelda	0.665	0.00E+00	2.5156	0.0888	0.645	2.28E-16	3.4245	0.6704
	SR=0.01							
	SVR				UR			
	SSIM		Time		SSIM		Time	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
airplane	0.542	2.28E-16	2.5898	0.0593	0.502	1.14E-16	2.7139	0.0718
baboon	0.252	0.00E+00	2.5195	0.0232	0.236	0.00E+00	2.6262	0.0503
Barbara	0.424	1.14E-16	2.9539	0.6824	0.414	1.14E-16	2.6386	0.0499
boat	0.357	1.14E-16	2.6172	0.1873	0.360	1.14E-16	2.6635	0.0515
cat	0.449	1.71E-16	2.7516	0.2617	0.422	0.00E+00	2.7850	0.1691
fruits	0.366	1.71E-16	2.4289	0.1719	0.360	1.14E-16	2.6927	0.1394
goldhill	0.334	1.14E-16	2.3656	0.0861	0.325	5.70E-17	3.0895	0.6060
cameraman	0.461	1.14E-16	2.5414	0.0554	0.449	2.85E-16	2.6078	0.1095
peppers	0.539	0.00E+00	2.4156	0.1529	0.516	2.28E-16	3.7746	0.6382
Zelda	0.384	1.71E-16	2.3336	0.3656	0.364	1.14E-16	3.4768	0.6348

Fig. 6. The ground truths and images reconstructed using the proposed CNN with $SR = 0.10$ for SVR and UR sensing patterns.

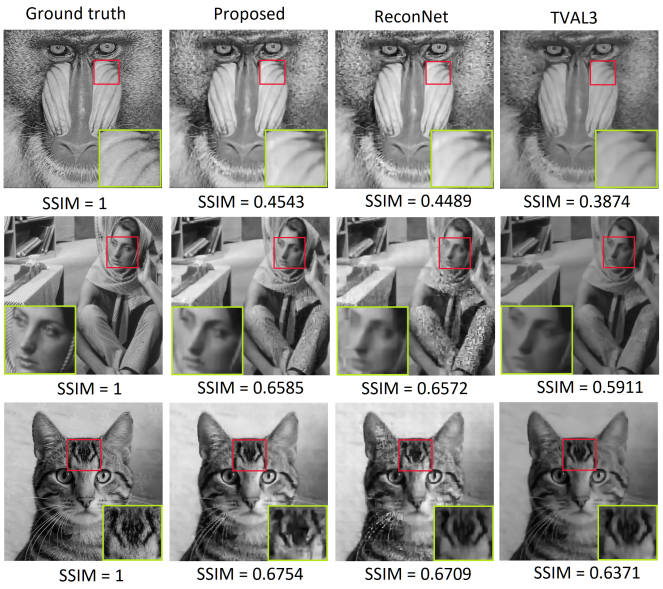


Fig. 7. The ground truth and the images reconstructed using various methods with $SR = 0.10$.

B. Comparison of Image Reconstruction Algorithms

Table II shows the mean SSIM and image reconstruction time with standard deviation ($n = 20$) for SVR images reconstructed using different image reconstruction methods. The results show that for $SR = 0.10$ and 0.04, the proposed CNN shows similar performance to TVAL3 in terms of mean SSIM but significantly lower reconstruction time. Additionally, the proposed CNN generally shows better SSIM and time performance as compared to ReconNet. The results demonstrate that the proposed CNN can achieve image quality comparable to the well-established TVAL3 method, while also offering an advantage in computational efficiency. The performance improvement over TVAL3 and ReconNet highlights the potential of the proposed CNN as an effective solution for scenarios where both high-quality image reconstructions and timely results are essential.

Figure 7 shows the ground truth and images reconstructed with $SR = 0.10$ using the proposed CNN, ReconNet, and TVAL3. The results show that the proposed CNN outperforms ReconNet and TVAL3 in terms of SSIM. Visually, the image details of the images reconstructed with the proposed CNN are clearer than those reconstructed with other methods.

The results in Table II and Figure 7 show that the proposed CNN outperforms ReconNet in terms of mean SSIM and time efficiency. It is worth mentioning that the image reconstruction time required for the proposed CNN is drastically lower than TVAL3, although TVAL3 may yield better image quality in some cases. Hence, the proposed CNN is more favorable when considering the efficiency and image quality trade-off, as it demonstrates significant time reduction and comparable image quality to TVAL3.

V. Conclusion

This paper proposes a learning-based SVR approach aimed at effectively optimizing both image quality and processing speed in single-pixel imaging. In general, the primary goal in single-pixel imaging is to enhance the image acquisition and reconstruction efficiency while preserving high image quality. SVR sensing patterns were proven efficient, especially for lower SR conditions, that fewer measurements are required as compared to UR images yet maintain a comparable image quality. CNN had been shown to improve the time efficiency of the single-pixel imaging with CS.

The image reconstruction time of the proposed CNN using both SVR and UR sensing patterns does not show an obvious difference. In addition, the proposed SVR with CNN demonstrates improved image quality and time efficiency in most cases compared to ReconNet. Notably, the proposed CNN achieves similar performance to TVAL3 in terms of image quality while offering much greater efficiency in image reconstruction.

Overall, the results indicate that the proposed CNN, combined with SVR sensing patterns, enhances both the image quality and time efficiency in single-pixel imaging with CS framework. However, its performance may still be constrained under diverse imaging conditions. Future work could examine the robustness of the approach across a wider range of scenarios. Further studies on network architecture and training efficiency may help reduce computational cost and enhance reconstruction quality. Additionally, investigations into the effects of different training strategies and hyperparameter selections on image reconstruction could further refine the performance of the proposed CNN.

TABLE II
THE MEAN SSIM AND IMAGE RECONSTRUCTION TIME (s) WITH STANDARD DEVIATION ($n = 20$) FOR SVR IMAGES RECONSTRUCTED USING DIFFERENT IMAGE RECONSTRUCTION METHODS.

	SR=0.10											
	Proposed				ReconNet				TVAL3			
	SSIM		Time		SSIM		Time		SSIM		Time	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
airplane	0.786	1.14E-16	2.788	0.023	0.737	0.00E+00	2.923	0.020	0.794	1.52E-09	20.673	1.876
baboon	0.454	2.85E-16	2.726	0.028	0.401	1.71E-16	2.922	0.028	0.387	1.45E-08	23.775	1.208
Barbara	0.659	0.00E+00	2.724	0.378	0.618	1.14E-16	2.860	0.027	0.591	2.13E-07	25.442	1.582
boat	0.652	2.28E-16	2.824	0.101	0.580	1.14E-16	2.926	0.047	0.653	4.24E-08	22.850	1.356
cat	0.675	0.00E+00	2.925	0.201	0.630	2.28E-16	2.965	0.077	0.637	2.45E-08	22.374	1.301
fruits	0.744	2.28E-16	2.654	0.149	0.684	2.28E-16	3.127	0.267	0.714	2.42E-06	25.980	1.472
goldhill	0.651	0.00E+00	2.648	0.093	0.590	0.00E+00	3.400	0.428	0.658	8.98E-09	26.573	11.339
cameraman	0.819	0.00E+00	2.977	0.112	0.801	3.42E-16	2.769	0.022	0.843	2.20E-08	19.127	0.621
peppers	0.800	2.28E-16	2.550	0.115	0.769	0.00E+00	3.681	0.596	0.786	2.65E-08	25.391	10.622
Zelda	0.786	1.14E-16	2.536	0.084	0.730	2.28E-16	2.817	0.029	0.792	5.15E-08	23.515	4.921
	SR=0.04											
	Proposed				ReconNet				TVAL3			
	SSIM		Time		SSIM		Time		SSIM		Time	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
airplane	0.727	1.14E-16	2.734	0.021	0.575	2.28E-16	2.888	0.038	0.725	1.45E-05	22.819	1.208
baboon	0.364	1.14E-16	2.703	0.137	0.307	0.00E+00	2.805	0.028	0.340	2.31E-06	23.116	1.486
Barbara	0.594	1.14E-16	3.274	0.883	0.527	2.28E-16	2.783	0.034	0.585	2.19E-06	24.032	1.796
boat	0.559	0.00E+00	2.851	0.360	0.453	0.00E+00	2.942	0.122	0.588	4.51E-06	22.667	1.458
cat	0.620	1.14E-16	3.030	0.325	0.539	2.28E-16	2.821	0.056	0.595	4.28E-06	23.442	1.120
fruits	0.609	2.28E-16	2.594	0.156	0.533	1.14E-16	2.912	0.138	0.526	4.47E-06	25.751	1.493
goldhill	0.572	1.14E-16	2.634	0.151	0.460	0.00E+00	3.949	1.018	0.589	2.51E-06	22.932	4.301
cameraman	0.701	2.28E-16	2.698	0.036	0.704	0.00E+00	2.781	0.138	0.750	2.63E-08	19.185	0.564
peppers	0.732	2.28E-16	2.521	0.139	0.653	2.28E-16	3.278	0.544	0.734	4.30E-07	23.799	4.804
Zelda	0.665	0.00E+00	2.516	0.089	0.559	0.00E+00	2.778	0.043	0.754	1.16E-05	27.591	12.741

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MoHE) through Fundamental Research Grant Scheme (FRGS/1/2022/TK08/UTAR/02/8). The authors gratefully acknowledge Universiti Tunku Abdul Rahman which supported this work under UTARRF (IPSR/RMC/UTARRF/2020-C2/C07).

REFERENCES

[1] T. Lu, Z. Qiu, Z. Zhang, and J. Zhong, "Comprehensive comparison of single-pixel imaging methods," *Optics and Lasers in Engineering*, vol. 134, p. 106301, 2020.

[2] F. Rousset, N. Ducros, F. Peyrin, G. Valentini, C. D'andrea, and A. Farina, "Time-resolved multispectral imaging based on an adaptive single-pixel camera," *Optics express*, vol. 26, no. 8, pp. 10 550–10 558, 2018.

[3] Y. Xu, L. Lu, V. Saragadam, and K. F. Kelly, "A compressive hyperspectral video imaging system using a single-pixel detector," *Nature Communications*, vol. 15, no. 1, p. 1456, 2024.

[4] Y. Wang, K. Huang, J. Fang, M. Yan, E. Wu, and H. Zeng, "Mid-infrared single-pixel imaging at the single-photon level," *Nature Communications*, vol. 14, no. 1, p. 1073, 2023.

[5] M.-J. Sun and J.-M. Zhang, "Single-pixel imaging and its application in three-dimensional reconstruction: a brief review," *Sensors*, vol. 19, no. 3, p. 732, 2019.

[6] G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," *Optics Express*, vol. 28, no. 19, pp. 28 190–28 208, 2020.

[7] M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nature photonics*, vol. 13, no. 1, pp. 13–20, 2019.

[8] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.

[9] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[10] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[11] D. B. Phillips, M.-J. Sun, J. M. Taylor, M. P. Edgar, S. M. Barnett, G. M. Gibson, and M. J. Padgett, "Adaptive foveated single-pixel imaging with dynamic supersampling," *Science advances*, vol. 3, no. 4, p. e1601782, 2017.

[12] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *2009 16th IEEE international conference on image processing (ICIP)*. IEEE, 2009, pp. 3021–3024.

[13] Z. Shin, T.-Y. Chai, C. H. Pua, X. Wang, and S. Y. Chua, "Efficient spatially-variant single-pixel imaging using block-based compressed sensing," *Journal of Signal Processing Systems*, pp. 1–15, 2021.

[14] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[15] T. N. Canh and B. Jeon, "Restricted structural random matrix for compressive sensing," *Signal Processing: Image Communication*, vol. 90, p. 116017, 2021.

[16] H. Gan, S. Xiao, Y. Zhao, and X. Xue, "Construction of efficient and structural chaotic sensing matrix for compressive sensing," *Signal Processing: Image Communication*, vol. 68, pp. 129–137, 2018.

[17] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[18] Z. Ye, H. Wang, J. Xiong, and K. Wang, "Simultaneous full-color single-pixel imaging and visible watermarking using hadamard-bayer illumination patterns," *Optics and Lasers in Engineering*, vol. 127, p. 105955, 2020.

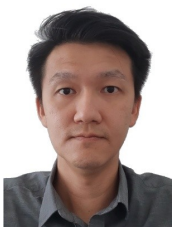
[19] H. Gan, S. Xiao, T. Zhang, Z. Zhang, J. Li, and Y. Gao, "Chaotic pattern array for single-pixel imaging," *Electronics*, vol. 8, no. 5, p. 536, 2019.

[20] W. L. Tey, M.-L. Tham, Y.-N. Phua, and S. Y. Chua, "Self-adaptive sampling for fourier single-pixel imaging using probability estimation," *Journal of Electronic Imaging*, vol. 33, no. 2, pp. 023 044–023 044, 2024.

[21] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed

measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 449–458.
[22] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Image compressed sensing using convolutional neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 375–388, 2019.
[23] Y. Tian, Y. Fu, and J. Zhang, "Joint supervised and unsupervised deep learning method for single-pixel imaging," *Optics & Laser Technology*, vol. 162, p. 109278, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030399223001718
[24] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo *et al.*, "Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.
[25] B. H. Woo, M.-L. Tham, and S. Y. Chua, "Adaptive coarse-to-fine single pixel imaging with generative adversarial network based reconstruction," *IEEE Access*, vol. 11, pp. 31 024–31 035, 2023.
[26] Z. Y. Shin, H. S. Lin, T.-Y. Chai, X. Wang, and S. Y. Chua, "Programmable single-pixel imaging," in *2019 13th International Conference on Sensing Technology (ICST)*. IEEE, 2019, pp. 1–6.
[27] Z. Shin, H. S. Lin, T.-Y. Chai, X. Wang, and S. Y. Chua, "Programmable spatially variant single-pixel imaging based on compressive sensing," *Journal of Electronic Imaging*, vol. 30, no. 2, pp. 1 – 15, 2021.
[28] A. Mathai, X. Wang, and S. Y. Chua, "Transparent object detection using single-pixel imaging and compressive sensing," in *2019 13th International Conference on Sensing Technology (ICST)*. IEEE, 2019, pp. 1–6.
[29] F. Magalhães, F. M. Araújo, M. V. Correia, M. Abolbashari, and F. Farahi, "Active illumination single-pixel camera based on compressive sensing," *Applied optics*, vol. 50, no. 4, pp. 405–414, 2011.
[30] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
[31] S. Gunasheela and H. Prasantha, "Compressed sensing for image compression: Survey of algorithms," in *Emerging Research in Computing, Information, Communication and Applications*, 2019, pp. 507–517.
[32] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE signal processing magazine*, vol. 24, no. 4, pp. 118–121, 2007.
[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.
[34] C. Li, W. Yin, and Y. Zhang, "User's guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.
[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
[36] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," *Unpublished manuscript*, vol. 40, no. 7, pp. 1–9, 2010.

Zhenyong Shin has been a Ph.D. student at the University of Malaya, Institute of Mathematical Sciences, Malaysia, since 2023. His research focuses on additive combinatorics. Zhen Yong received his Bachelor of Science (Honours) Physics degree and M. EngSc. degree in Optical Engineering from Universiti Tunku Abdul Rahman, Lee Kong Chian Faculty of Engineering and Science, Malaysia.



Chang Hong Pua received his Ph.D. in Physics from University of Malaya in 2012. He is currently an Associate Professor of Universiti Tunku Abdul Rahman. In year 2024, he was elected as the council member of Institute Physics Malaysia. His research interest is in Photonics and Sensors. His works cover lasers, fiber lasers, optical waveguide fabrication, optical sensors, acoustic sensors etc.

Tong-Yuen Chai received his B.Eng. (Hons) and M.Sc. degrees in Mechatronics Engineering from Universiti Malaysia Perlis, Malaysia. He received his Ph.D. in Engineering from Universiti Tunku Abdul Rahman (UTAR). He is currently an Associate Professor in UTAR. His research focuses on Large Vision Model, deep learning and AI-driven sustainability technologies. He leads vision AI solutions for manufacturing and has also consulted for industry partners on AI applications.

Xin Wang obtained her PhD from Nanyang Technological University (NTU). Currently, she is an Associate Professor at Monash University Malaysia campus. Dr. Wang's research interests include machine vision, image processing, laser sensing, optical metrology, deep learning, non-destructive testing and structural health monitoring. She has published 2 book chapters and over 70 peer-reviewed journal papers. Dr. Wang is a Chartered Engineer with IMechE (UK) and a Senior Member of IEEE.

Sing Yee Chua received her BEng in Electrical and Electronics Engineering from the University of Technology Malaysia and her PhD in Engineering from Monash University. She is currently an Assistant Professor at Universiti Tunku Abdul Rahman. Her research interests include computer vision, signal and image processing, and optical engineering.