Real-time Single-pixel Imaging Based on Deep Learning

Ziran Wei[1], Jianlin Zhang[2], Wei Du[1] and Zhiruo Wang[3]
[1] Chengdu Aircraft Industrial (Group) Co. Ltd., Aviation Industry Corporation of China (AVIC), Chengdu, China
[2] Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China
[3] Xidian University, Xi'an, China
zhiruowang@stu.xidian.edu.cn

***Abstract*—The emergence of compressed sensing (CS) theory has enabled the development of single-pixel cameras that achieve high-resolution imaging using a single photodetector. However, traditional CS reconstruction algorithms require significant computational time and face an inherent trade-off between imaging resolution and frame rate, limiting current single-pixel cameras to static scene imaging. A key challenge lies in achieving real-time single-pixel imaging with both high frame rate and high resolution. This paper proposes a real-time single-pixel imaging technology based on deep learning. We design a deep convolutional neural network architecture incorporating residual networks to simulate the measurement and reconstruction process of CS-based single-pixel imaging. The network is trained on an image dataset and subsequently deployed for single-pixel imaging. The trained network can complete image reconstruction at a low sampling rate with minimal latency, enabling real-time single-pixel video capture at 128×128 resolution with 33 frames per second (fps) at a 4% sampling rate. Furthermore, we implement a four-channel parallel signal processing method to achieve real-time single-pixel imaging video at 256×256 resolution at 33 fps.**

***Index Terms*—Compressed sensing, Real-time single-pixel imaging, Deep learning**

I. Introduction

Compressed sensing (CS) theory, introduced by Donoho [1] and Candès [2] in 2006, revolutionized signal processing by enabling signal reconstruction from significantly fewer measurements than traditional Nyquist sampling requires. This breakthrough has particular significance for imaging applications, where it enables high-resolution image capture using limited measurements. The development of the single-pixel camera at Rice University [7] demonstrated the practical potential of CS theory, achieving high-resolution imaging using just one photodetector.

Despite these advances, current single-pixel imaging systems face significant challenges in real-time applications. Traditional CS reconstruction algorithms require substantial computational resources [8], [9], creating an inherent trade-off between imaging resolution and frame rate. This limitation has largely confined single-pixel cameras to static scene imaging, making real-time video capture particularly challenging.

Our paper presents a novel solution to this challenge through the integration of deep learning with single-pixel imaging technology. We introduce a custom-designed deep convolutional neural network architecture that enables real-time single-pixel video capture while maintaining high resolution and frame rates. Furthermore, we implement an innovative four-channel parallel processing method to enhance system performance.

This paper is organized as follows: Section II provides the theoretical background and recent developments in the field. Section III details our proposed network architecture and four-channel processing methodology. Section IV presents experimental results and analysis, and Section V concludes with our findings and future directions.

II. Literature Review

The convergence of deep learning and computational imaging has catalyzed significant advancements in imaging technology over the past decade. This synthesis has proven particularly transformative in addressing fundamental challenges in image acquisition and processing. Recent developments in neural network architectures have enabled breakthrough achievements in various imaging applications, from lensless computational imaging systems [10] to sophisticated non-line-of-sight reconstruction techniques [11]. Notably, researchers have demonstrated remarkable progress in ghost imaging [12], where deep learning approaches have substantially improved both image quality and reconstruction speed.

In the medical imaging domain, the integration of deep learning has revolutionized traditional imaging paradigms. Researchers have achieved significant improvements in both image quality and processing efficiency [5], [13], particularly in applications requiring real-time processing of complex imaging data. These advances have demonstrated the potential of neural networks to overcome traditional limitations in computational imaging, providing new pathways for addressing long-standing challenges in the field.

The application of deep learning to compressed sensing has emerged as a particularly promising direction for advancing imaging technology. Recent studies have demonstrated impressive results across diverse domains, including radar systems [3], medical diagnostics [5], and modern communication technologies [6]. These implementations have shown that deep learning can significantly reduce the computational overhead traditionally associated with compressed sensing reconstruction, while maintaining or even improving reconstruction quality.

Traditional compressed sensing approaches, while theoretically sound, have faced significant practical limitations in real-

ISSN:3006-2985(Print)
ISSN:3006-9297(Online)

time applications. These challenges are particularly evident in single-pixel imaging, where the trade-off between resolution and frame rate has been a persistent constraint. The work of Chen et al. [18] and Wei et al. [19] has highlighted these limitations, demonstrating that conventional reconstruction methods struggle to meet the demands of real-time, high-resolution imaging applications. Block-based compressed sensing approaches [20] have offered partial solutions to these challenges, providing improved computational efficiency through structured sampling and reconstruction techniques.

Recent innovations in deep residual learning architectures [21], [22] have opened new possibilities for addressing these limitations. Residual networks have demonstrated exceptional capability in learning complex image features and reconstruction patterns, making them particularly well-suited for single-pixel imaging applications. These architectures offer improved training stability and enhanced performance compared to traditional neural network designs, providing a robust foundation for real-time image reconstruction systems.

The synthesis of these various technological streams – deep learning, compressed sensing, and parallel processing – presents exciting opportunities for advancing single-pixel imaging technology. Our work builds upon these foundations while introducing novel architectural elements and processing strategies. By combining advanced neural network designs with innovative parallel processing techniques, we address the fundamental challenges of achieving both high frame rates and high resolution in real-time single-pixel imaging applications. This approach not only overcomes traditional performance limitations but also establishes a new paradigm for real-time computational imaging systems.

III. Formulation

The signal measurement and reconstruction process based on Compressed Sensing can be mathematically expressed as:

$$y = \Phi x \tag{1}$$

where $x \in \mathbb{R}^n$ represents the original signal, $\Phi \in \mathbb{R}^{m \times n}$ denotes the measurement matrix, and $y \in \mathbb{R}^m$ is the measurement vector, with $m \ll n$. This system is underdetermined [14]–[16]. Signal reconstruction involves recovering x from y through CS-based reconstruction algorithms, which requires specific theoretical conditions to be satisfied [17], [18].

We have implemented a single-pixel imaging system based on CS measurement and reconstruction theory, as illustrated in Fig. 1. The system employs a Digital Mirror Device (DMD, ViAlUX V-7001) to express the sampling sequences of the measurement matrix. Operating in 1-bit mode (0-1 binary), the DMD achieves a maximum switching rate of 22,727 Hz. To optimize DMD encoding efficiency, we utilize a 0-1 binary measurement matrix. Signal detection is performed using a Silicon PhotoMultiplier (SPM, Hamamatsu C13369), with analog-to-digital conversion operating at a 50MHz sampling frequency. Scene reconstruction is subsequently performed computationally from the digitized data. To enhance structural balance and minimize artifacts in the imaging results, we implement bilateral projection [19], as depicted in Fig. 2. This approach leads to an updated expression of equation (1):

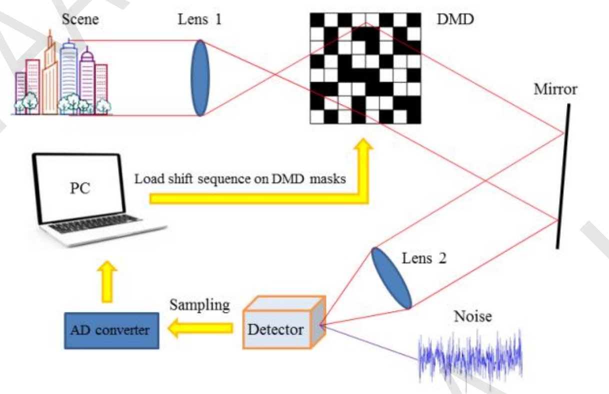

Fig. 1. Schematic diagram of the single-pixel imaging.

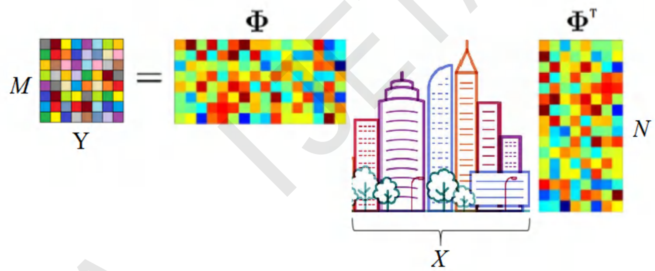

Fig. 2. Schematic diagram of the bilateral projection.

$$\Phi = \begin{bmatrix} \Phi_B & 0 & \cdots & 0 \\ 0 & \Phi_B & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Phi_B \end{bmatrix}$$

Fig. 3. The block diagonal measurement matrix.

$$y = \begin{bmatrix} \Phi \\ -\Phi \end{bmatrix} x \tag{2}$$

While the bilateral projection method potentially increases computational complexity in image reconstruction, we mitigate this effect by implementing a block diagonal measurement matrix [20], as shown in Fig. 3. This optimization helps maintain computational efficiency while preserving reconstruction quality.

IV. The Proposed Methods

A. Network Architecture

We present a novel deep convolutional neural network architecture for real-time single-pixel imaging. Our design incorporates the residual network approach [21], which facilitates the training of deeper network architectures through improved

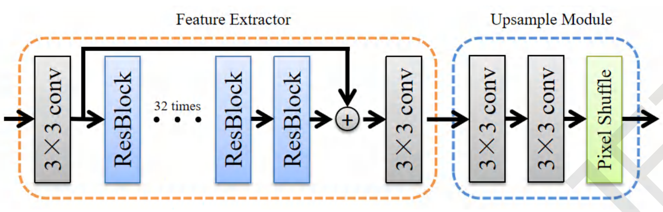

Fig. 4. The flowchart of our network

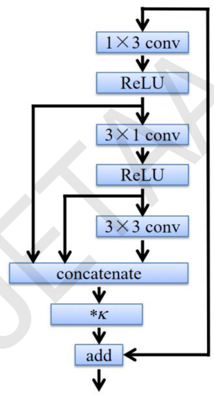

Fig. 5. The improved residual block network

optimization capabilities. The residual network's distinctive feature lies in its skip connections, which effectively address the gradient vanishing problem that typically emerges in deep neural networks with increasing layers.

Our network architecture employs a residual network-style backbone for feature extraction, complemented by pixel shuffle [22] for feature upsampling. Fig. 4 illustrates the flowchart of our network architecture. The Feature Extractor component incorporates an improved residual block, shown in Fig. 5, which comprises three distinct convolution operations and includes a learnable parameter α.

Compared to the original residual network structure [21], we have deliberately removed the batch normalization layers from the residual block. This modification stems from our observation that single-pixel image reconstruction primarily focuses on recovering image details and textures, without requiring significant correlations between input images. Additionally, batch normalization layers consume substantial GPU memory. By eliminating these layers, we achieve significantly reduced memory requirements, enabling the construction of larger and more sophisticated network models within limited computational resources. This enhancement theoretically improves performance for single-pixel imaging reconstruction.

Our network implements a large number of filters, generating numerous feature maps. However, we recognized that increasing feature parameters beyond certain thresholds could potentially destabilize the network training process. To address this challenge, we introduced residual scaling with a factor α to maintain training stability. The scaling layer is positioned after the final convolution layer, as depicted in Fig. 5. We initialize α at 0.1 and treat it as one of many learnable parameters that adaptively adjust during network training.

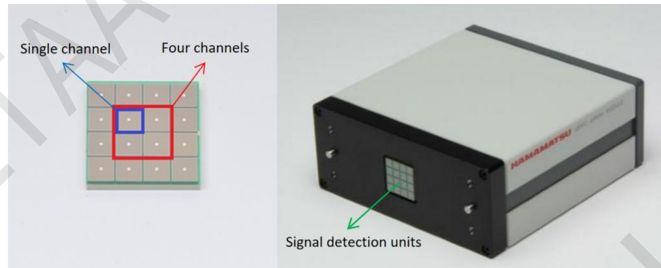

Fig. 6. The schematic diagram of single-channel and four channels of SPM.

B. Training the Network

Our single-pixel imaging system operates at a resolution of 256×256. We train the network at a 4% sampling rate, resulting in 2704 (52×52) samples per imaging scene. To facilitate network training across large datasets, we employ a fixed measurement matrix following the structure shown in Fig. 3, with a constant distribution of 0 and 1 elements.

The network implementation utilizes four GTX 1080Ti GPUs and is developed using PyTorch [23]. We selected the DIV2K dataset [24] for training, which contains 900 2K-resolution images particularly well-suited for image super-resolution reconstruction in our single-pixel camera application. The network undergoes training for 960 epochs on DIV2K. We implement data augmentation through random crop, mirror, and flip operations. The crop size is set to 256×256, establishing the label image size for our network, with a corresponding input size of 2704 (52×52) at the 4% sampling rate. Prior to network input, we subtract 128 from all pixel values.

Experimental evidence has demonstrated that the L1 loss function outperforms L2 loss in terms of PSNR and SSIM metrics for network training in image reconstruction quality [25]. Therefore, we implement L1 loss as our objective function and employ stochastic gradient descent optimization with momentum=0.9 and weight decay=1e-4. The learning rate initiates at 1e-4 and halves every 160 epochs.

C. Four Channel Signal Processing of SPI

Our single-pixel imaging system incorporates an SPM featuring a 4×4 photodetector array, as illustrated in Fig. 6. To enhance real-time video imaging frame rates, we utilize the central 2×2 four detection units (Channels) for 256×256 resolution real-time imaging, with each detection unit responsible for 128×128 single-pixel imaging. We implement four-channel parallel analog-to-digital conversion to process the detection signals simultaneously, as shown in Fig. 7, achieving a frame rate four times higher than single-channel processing.

V. Experiments Analysis and Discussion

In this section, we evaluate our proposed methods through implementation in our single-pixel camera system for real-time video imaging experiments. Our deep learning approach demonstrates a significant advancement in processing efficiency. While the initial network training requires considerable computational time, the subsequent image reconstruction using the trained network is remarkably fast, requiring only

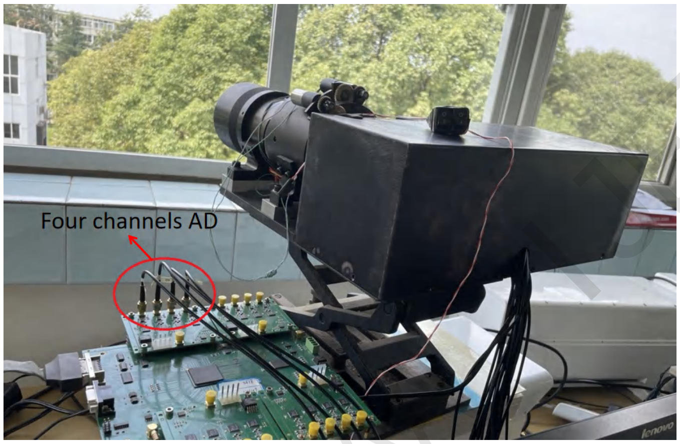

Fig. 7. Four-channels parallel analog-to-digital converter.

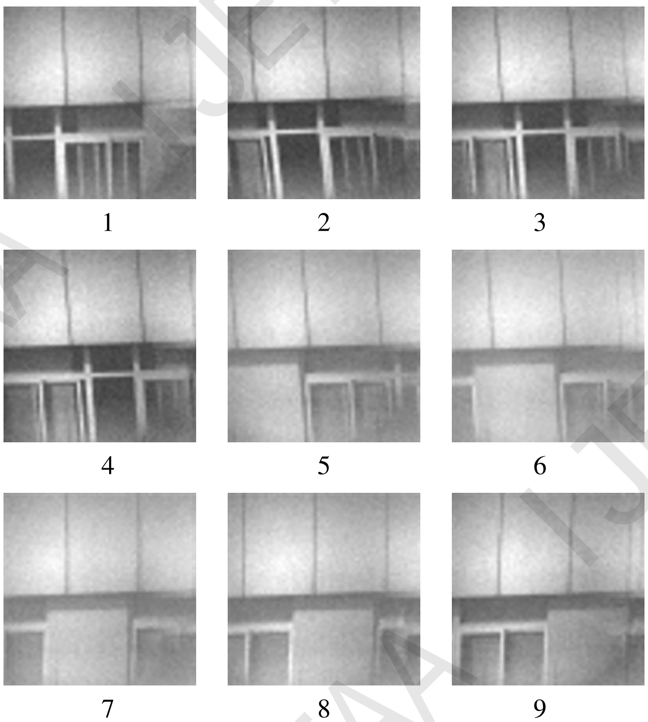

Fig. 8. Single-channel, real-time 256×256 single-pixel imaging video (1 to 9 frames).

approximately 19 ms for 256×256 resolution images and 9 ms for 128×128 resolution images—latencies that are practically negligible in real-time applications.

The frame rate of our real-time single-pixel imaging system is primarily constrained by the DMD modulation time. Operating at a 4% sampling rate with the DMD's maximum switching rate of 22,727 Hz, our system achieves real-time single-pixel video imaging at 33 fps for 128×128 resolution and 8 fps for 256×256 resolution using a single channel. The experimental results are presented in Fig. 8.

Further enhancement is achieved through our four-channel parallel processing method, which enables 256×256 resolution imaging at 33 fps. The corresponding video frames are shown in Fig. 9. Analysis of the experimental results reveals several key findings. The single-channel video capture of distant buildings and trees, obtained by rotating the single-pixel camera, demonstrates clear visualization of detailed features such as building windows. However, we observed two technical challenges in the single-channel implementation: dynamic noise induced by camera movement and moderately leaning vertical contour lines in each frame. These artifacts arise from frame rate limitations and video data stream reading delays during real-time dynamic imaging.

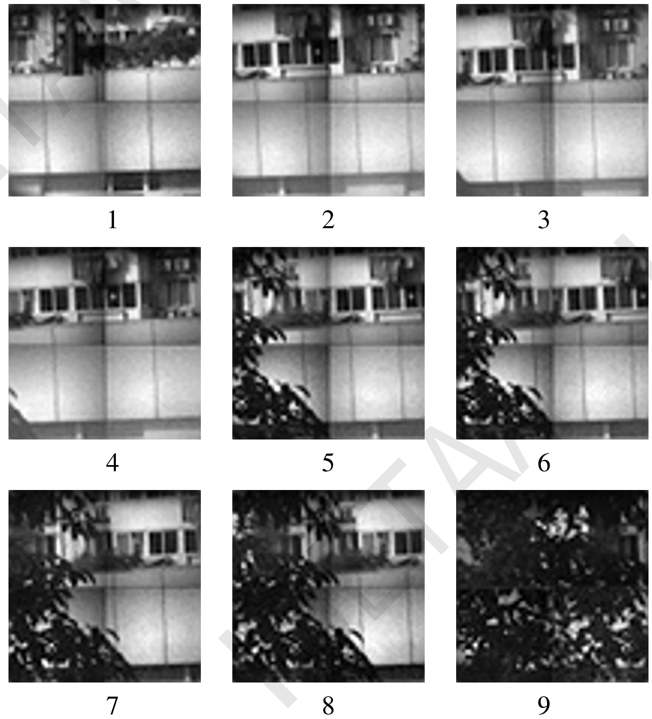

Fig. 9. Four channels, real-time 256×256 single-pixel imaging video (1 to 9 frames)

The four-channel implementation addresses these limitations effectively. Compared to single-channel video imaging, the increased frame rate and enhanced signal processing speed per channel yield several improvements:

- Reduced leaning of vertical contour lines in video frames
- Minimized video data stream reading delays
- Expanded imaging field of view
- Enhanced balance in reconstructed frame images
- Improved video stream smoothness

Additionally, distributing imaging signals of the same resolution across four independent channels for processing results in lower visual noise compared to single-channel imaging. However, one trade-off emerges: because images are reconstructed separately and in parallel across the four imaging channels, the final results exhibit some cross mark artifacts.

These experimental results validate the effectiveness of our proposed deep learning approach for real-time single-pixel imaging while highlighting areas for potential future optimization. The demonstrated improvements in frame rate and resolution through four-channel parallel processing represent a significant advancement in single-pixel imaging technology.

VI. Conclusion

This paper presents a significant advancement in real-time single-pixel video imaging through the application of deep learning techniques. Our research demonstrates that deep learning-based approaches can effectively overcome traditional limitations in single-pixel imaging, particularly regarding the trade-off between resolution and frame rate. The implementation of four-channel parallel signal processing represents a novel solution for achieving higher resolution and frame rates in single-pixel imaging systems.

Our experimental results validate that the proposed methodology successfully achieves real-time single-pixel video imaging with 256×256 resolution at 33 frames per second, a notable improvement over existing systems. The parallel processing architecture not only enhances the frame rate but also improves overall image quality by reducing artifacts and noise typically associated with single-channel implementations.

The integration of deep learning with single-pixel imaging opens new possibilities for applications requiring high-speed, high-resolution imaging in challenging environments. Our approach demonstrates that neural networks can effectively learn and optimize the complex relationships between compressed measurements and reconstructed images, providing a more efficient alternative to traditional compressed sensing reconstruction algorithms.

This research establishes a foundation for future developments in single-pixel imaging technology. The parallel processing method described here may inspire further exploration of multi-channel architectures and alternative neural network designs for enhanced imaging performance. Future research directions could focus on optimizing network architectures for specific applications, reducing computational requirements, and developing more sophisticated parallel processing schemes to further improve image quality and frame rates.

The success of this approach suggests that similar principles could be applied to other computational imaging systems where real-time performance is crucial. As deep learning techniques continue to evolve, we anticipate further improvements in both the efficiency and capability of single-pixel imaging systems.

VII. Acknowledgment

The authors would like to express their sincere gratitude for the financial support provided by several organizations that made this research possible. This work was supported in part by the National High Technology Research and Development Program of China (863 Program) under Grant G158207, the West Light Foundation for Innovative Talents under Grant YA18K001, and the Key Research and Development Plan of Shaanxi Province under Grant 2024GX-YBXM-232. The support from these organizations has been instrumental in facilitating our research and advancing the field of single-pixel imaging technology.

References

[1] D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[2] E. Candès, "Compressive sampling," in Proc. Int. Congress Math., Madrid, Spain, 2006, vol. 3, pp. 1433–1452.
[3] H. Li, J. Xu, H. Song, and Y. Wang, "Deep unfolding for sparse aperture ISAR imaging via compressed sensing: CIST+," in IET Int. Radar Conf., Chongqing, China, 2023, pp. 3483–3488.
[4] D. Milioris, "Classification in twitter via compressive sensing," in Proc. IEEE Conf. Comput. Commun. Workshops, Hong Kong, China, 2015, pp. 95–96.
[5] Z. Xue et al., "DARCS: Memory-efficient deep compressed sensing reconstruction for acceleration of 3D whole-heart coronary MR angiography," IEEE Trans. Med. Imaging, Jan. 2024.
[6] B. Xia and Y. Tang, "Compressive sensing based channel estimation method for MIMO-OFDM power line communication," in Int. Conf. Energy, Power and Grid, Guangzhou, China, 2024, pp. 1658–1661.
[7] M. Duarte et al., "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag., vol. 25, no. 2, pp. 83–91, 2008.
[8] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," IEEE Trans. Image Process., vol. 23, no. 8, pp. 3618–3632, 2014.
[9] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," IEEE Trans. Inf. Theory, vol. 62, no. 9, pp. 5117–5144, 2016.
[10] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica, vol. 4, pp. 1117–1125, 2017.
[11] X. He, P. Wang, H. Liu, and X. Yu, "Passive non-line-of-sight imaging reconstruction based on dual input U-Net," in Int. Conf. Neural Netw., Inf. Commun. Eng., Guangzhou, China, 2023, pp. 337–341.
[12] Y. Hualong and G. Daidou, "Research on double encryption of ghost imaging by SegNet deep neural network," IEEE Photon. Technol. Lett., vol. 36, no. 10, pp. 669–672, May 2024.
[13] N. Mishra and S. Mishra, "Impact of deep learning applications in medical hyperspectral imaging," in Healthcare Big Data Analytics: Computational Optimization and Cohesive Approaches. De Gruyter, 2024, pp. 179–206.
[14] F. Hashimoto et al., "Compressed-sensing magnetic resonance image reconstruction using an iterative convolutional neural network approach," Appl. Sci., vol. 10, no. 6, pp. 1902, 2020.
[15] K. Haris et al., "Free-breathing fetal cardiac MRI with doppler ultrasound gating, compressed sensing, and motion compensation," J. Magn. Reson. Imaging, vol. 51, no. 1, pp. 260–272, 2020.
[16] G. R. Matcuk, J. S. Gross, and J. Fritz, "Compressed sensing MRI," Adv. Clin. Radiol., vol. 2, pp. 257–271, 2020.
[17] V. B. Manur and L. Ali, "Compressed sensing channel estimation for STBC-SM based hybrid MIMO-OFDM system for visible light communication," Int. J. Commun. Syst., vol. 33, no. 11, pp. e4403, 2020.
[18] Y. Chen et al., "Hyperspectral image compressive sensing reconstruction using subspace-based nonlocal tensor ring decomposition," IEEE Trans. Image Process., vol. 29, pp. 6813–6828, 2020.
[19] Z. R. Wei, J. L. Zhang, Z. Y. Xu et al., "Improving the signal-to-noise ratio of super resolution imaging based on single pixel camera," IEEE Photon. J., pp. 1–16, 2019.
[20] J. E. Fowler, S. Mun, and E. W. Tramel, "Block-based compressed sensing of images and video," Found. Trends Signal Process., vol. 4, no. 4, pp. 297–416, 2010.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.
[22] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1874–1883.
[23] A. Paszke et al., "Automatic differentiation in PyTorch," in NIPS Workshop Autodiff, 2017.
[24] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1110–1121.
[25] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," arXiv preprint arXiv:1511.08861, 2015.