

## Research Article

# MsDC-DEQ-Net: Deep Equilibrium Model (DEQ) with Multiscale Dilated Convolution for Image Compressive Sensing (CS)

**Youhao Yu and Richard M. Dansereau** 

*Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada*

Correspondence should be addressed to Richard M. Dansereau; [rdanse@sce.carleton.ca](mailto:rdanse@sce.carleton.ca)

Received 21 July 2023; Revised 12 December 2023; Accepted 5 January 2024; Published 18 January 2024

Academic Editor: Sebastian Miron

Copyright © 2024 Youhao Yu and Richard M. Dansereau. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compressive sensing (CS) is a technique that enables the recovery of sparse signals using fewer measurements than traditional sampling methods. To address the computational challenges of CS reconstruction, our objective is to develop an interpretable and concise neural network model for reconstructing natural images using CS. We achieve this by mapping one step of the iterative shrinkage thresholding algorithm (ISTA) to a deep network block, representing one iteration of ISTA. To enhance learning ability and incorporate structural diversity, we integrate aggregated residual transformations (ResNeXt) and squeeze-and-excitation mechanisms into the ISTA block. This block serves as a deep equilibrium layer connected to a semi-tensor product network for convenient sampling and providing an initial reconstruction. The resulting model, called MsDC-DEQ-Net, exhibits competitive performance compared to state-of-the-art network-based methods. It significantly reduces storage requirements compared to deep unrolling methods, using only one iteration block instead of multiple iterations. Unlike deep unrolling models, MsDC-DEQ-Net can be iteratively used, gradually improving reconstruction accuracy while considering computation tradeoffs. Additionally, the model benefits from multiscale dilated convolutions, further enhancing performance.

## 1. Introduction

Compressive sensing (CS) is a signal processing technique used to efficiently acquire signals that exhibit sparsity or compressibility in a sparse domain, and this information can then be reconstructed back to the original domain with high probability [1, 2]. In CS, a signal is measured through a small number of linear projections obtained by multiplying the signal with a sensing matrix, typically a random matrix. For images, CS allows fewer measurements to be acquired than the number of pixels in an image, making it particularly advantageous as it can significantly reduce storage requirements. Reconstruction of the image is usually obtained using algorithms that leverage the sparsity of the signal, such as sparse optimization or convex optimization techniques, but these techniques are often computationally expensive and slow to converge. CS finds applications in various fields, including magnetic resonance imaging [3], radar signal

sampling [4], cryptosystems [5], snapshot imaging [6], and video sensing [7, 8]. It proves especially useful when dealing with large amounts of data, as it can lead to significant reductions in storage and processing requirements.

A multitude of optimization-based CS reconstruction methods have been developed. One such method is basis pursuit, an algorithm that tackles the underdetermined linear system by finding the sparsest solution [9]. It assumes signal sparsity on a specific basis and solves a convex optimization problem to determine the sparsest representation. Another iterative algorithm, iterative hard thresholding, updates the signal estimate by applying thresholding at each iteration [10]. The algorithm computes a gradient descent step and enforces sparsity through thresholding, where only the  $k$  largest coefficients ( $k$  being the desired sparsity level) are retained, while the rest are set to zero. Compressive sampling matching pursuit (CoSaMP), another iterative algorithm, iteratively refines the signal estimate by selecting the support that best

aligns with the measurements [11]. At each iteration, CoSaMP identifies the  $k$  largest entries in the product of the sensing matrix  $\Phi$  and the current residual. It then solves a least-squares problem to obtain the coefficients of the selected atoms. Approximate Message Passing (AMP), an iterative algorithm utilizing a message-passing framework, estimates the signal by combining the current estimate with the noisy measurements and applying a soft thresholding operator [12]. The result undergoes linear combination and soft thresholding at each iteration. An optimization algorithm frequently employed in sparse signal recovery and regression problems is the iterative shrinkage thresholding algorithm (ISTA). It is a variant of the proximal gradient descent algorithm [13]. ISTA iteratively updates the estimate of the sparse signal by taking a gradient step and subsequently applying thresholding to promote sparsity. The thresholding operation sets small entries to zero, and the degree of sparsity is controlled by the threshold value. These methods, while effective, suffer from high computational complexity due to the necessity of multiple iterations to achieve convergence. Additionally, some parameters require careful tuning for optimal performance.

In recent years, neural network-based CS reconstruction methods have gained popularity. These methods leverage the ability of neural networks to learn complex, nonlinear mappings between compressed measurements and reconstructed signals. Unlike traditional methods mentioned earlier, these noniterative network-based approaches significantly reduce computational requirements while achieving impressive reconstruction performance [14]. Most network-based methods are trained as black boxes, harnessing the powerful learning capacity of deep networks but lacking insights from a CS perspective. On the other hand, optimization methods involve iterating over parameters to minimize a loss function. Deep unrolling methods (DUM) in machine learning can be seen as incorporating insights from iterations in optimization methods [14]. In deep unrolling, a fixed number of architecturally identical blocks are utilized, where the output of each block serves as input to the next. This can be interpreted as a form of iteration, where the block is applied iteratively (typically 5–10 times) to capture longer-term dependencies [15]. While more iteration blocks often yield improved performance, training such models consumes substantial memory, potentially leading to out-of-memory issues [16]. Deep equilibrium models (DEQs) belong to a class of deep learning models that employ fixed-point iteration schemes to learn stable equilibrium points corresponding to optimal solutions for given optimization problems [17]. DEQ has been applied to CS reconstruction by formulating the problem as an optimization task solvable through fixed-point iteration schemes [15, 16, 18]. By utilizing a neural network to learn the fixed-point iteration scheme, DEQ demonstrates the ability to reconstruct images accurately and efficiently from compressive measurements.

This paper presents the design of a DEQ, named MsDC-DEQ-Net, for image CS, incorporating multiscale dilated convolutions. The model consists of two key components. First, we utilize the semi-tensor product (STP) theory to enable direct compressed measurement and initial reconstruction without the need for image block processing.

This approach avoids block artifacts in reconstruction and employs a learnable measurement matrix to capture essential signal information. Second, we construct a deep equilibrium layer based on one iteration of the ISTA algorithm, mapping  $l_1$  norm optimization for CS reconstruction into a deep network. We incorporate aggregated residual transformations (ResNeXt) [19] to enhance performance and employ the squeeze-and-excitation network (SENet) [20] to remove redundancy and enhance valuable information. Compared to DUM, our proposed MsDC-DEQ-Net possesses fewer learnable parameters and offers a tradeoff between accuracy and computation.

The main contributions of this work can be summarized as follows: (1) By leveraging STP without image block partitioning, we achieve image sampling and initial reconstruction, mitigating block artifacts and using a learnable measurement matrix that captures critical signal information. (2) We map one iteration of the ISTA algorithm into a network layer, referred to as the ISTA block, and enhance its performance by incorporating the ResNeXt and SENet structures. Additionally, we apply multiscale dilated convolutional layers to further improve performance. (3) To address the issue of large model size associated with DUM, we employ the ISTA block to construct a DEQ. This allows for multiple applications of the trained model, enabling multiple ISTA iterations and a continuous improvement in reconstruction performance.

## 2. Related Work

Network-based CS image reconstruction methods can be broadly classified into two categories: deep non-unfolding networks and deep unfolding networks [21]. Each category encompasses specific methods that vary in terms of network architectures, loss functions, regularization techniques, and other details. The following section provides a concise overview of both reconstruction methods. Subsequently, we delve into the exploration of the DEQ, focusing on its relevance to our own research.

### 2.1. Network-Based CS Image Reconstruction Methods

**2.1.1. Deep Non-Unfolding Networks.** Deep non-unfolding networks are deep neural networks trained to directly reconstruct under-sampled images from compressed measurements, without explicitly modeling the image acquisition process. These methods typically involve training a deep neural network using large-scale datasets of natural images to learn a mapping from under-sampled measurements to the fully sampled image domain.

Usually, these reconstructed images often lack fine details, particularly at low measurement rates. To address this issue, the dual-path attention network (DPA-Net) [22] employs two paths. The structure path focuses on reconstructing the dominant structural components, while the texture path recovers the remaining texture details. An attention module is utilized to transmit structure information to the texture path. To reduce the number of parameters, block-based compressed sensing (BCS) is often employed for sampling and reconstructing small image blocks. The sampling and initial reconstruction are achieved using a sampling and whole image

denoising network based on a generative adversarial network (SWDGAN) [23]. Nonoverlapping blocks are segmented from the original images, and a fully connected layer is utilized for sampling and initial reconstruction. A whole image-dense residual denoising module is then applied to improve the reconstruction quality further. The generator and discriminator are trained alternately to obtain an optimal model. Similar to DPA-Net, the parallel enhanced network (PE-Net) [24] consists of two reconstruction networks. The basic network produces the initial reconstruction, while the enhanced network progressively refines details by utilizing information from submodules of the basic network. The final reconstruction is the cumulative result of the two parallel networks.

These models all employ BCS methods, where large-scale images are processed in a block-by-block manner. In contrast, the semi-tensor product network (STP-Net) [25] treats the image as a whole without segmentation. Leveraging STP, an image can be directly sampled using a small-sized measurement matrix through matrix multiplication. The initial reconstruction can also be obtained in a similar manner. It is worth noting that these models are trained as black boxes, lacking insights from the CS domain.

**2.1.2. Deep Unfolding Networks.** Deep unfolding networks aim to unfold the iterative optimization process of traditional CS reconstruction algorithms, such as ISTA, into a single end-to-end trainable deep neural network. These networks explicitly model the physics of the image acquisition process and learn a mapping from under-sampled measurements to the original image by iteratively updating the reconstructed image estimate.

The ISTA-Net [14] approach solves CS reconstruction using the ISTA algorithm by casting it into a deep network form. This allows it to benefit from the structural insights of traditional optimization-based methods while maintaining the fast solution speed of neural networks. Nonlinear convolutional layers are employed to solve the proximal mapping associated with the sparsity-inducing regularizer. ISTA-Net's network design is well-defined, providing interpretability and allowing for structural diversity originating from the CS domain. In contrast to ISTA-Net, the optimization-inspired explicable deep network (OPINE-Net) [26] utilizes a data-driven adaptively learned matrix instead of generating the sampling matrix with a fixed random Gaussian matrix. Although the sensing matrix can take other forms, such as *noiselets* coefficients without additional memory storage or multiplications, as shown in [27–29], data-driven methods are expected to have better performance for learnable classes of data. OPINE-Net adopts the framework of ISTA-Net, with each of its nine blocks corresponding to one iteration in the traditional ISTA algorithm. Notably, all blocks share the same weights without affecting the final reconstruction performance. AMP-Net [30] unfolds the iterative denoising process of the AMP algorithm. Similar to OPINE-Net, AMP-Net consists of 9 AMP denoising iteration blocks, and the sampling matrix is trainable. Additionally, AMP-Net integrates deblocking modules to eliminate blocking artifacts. STP-ISTA-Net, introduced in [16],

combines STP-Net and ISTA-Net by connecting the output of the former as a better initial reconstruction for the latter. This model uses five iteration blocks, fewer than the aforementioned models, while still achieving competitive performance.

**2.2. DEQ.** Bai et al. [31] observed that the hidden layers of many existing sequence models converge to a fixed point. To address this, they propose the DEQ, which allows for finding the equilibrium point directly via root-finding. This method is equivalent to running an infinite-depth (weight-tied) network, and the equilibrium point can be backpropagated using implicit differentiation. The DEQ model requires only constant memory.

While DUMs like ISTA-Net, OPINE-Net, and AMP-Net achieve good performance by simulating a fixed number of iterations of an optimization method in their architectures, the number of iterations must be limited due to the difficulty of training large-sized networks. Additionally, significant errors arise when expecting more optimization iterations through multiple applications of the trained model [15]. In contrast, the DEQ model can be executed for more optimization iterations, leading to consistent improvements in reconstruction quality while requiring only constant memory in both training and testing [15]. There exists a tradeoff between reconstruction quality and computation.

The DEQ model has been applied to inverse problems in imaging [15, 16] and video snapshot compressive imaging (SCI) reconstruction [18]. Although the DEQ model simplifies the structure compared to DUM, a bottleneck arises with single-scale convolutions, limiting the ability to extract and propagate useful information. Dilated convolutions are widely used in various domains, such as image denoising [32], feature detection [33], image super-resolution [34], and CS reconstruction [35]. Dilated convolutions expand the receptive field without increasing the number of parameters, thereby maintaining the same amount of computation. This inspires us to extract features of different scales using a model that incorporates multiple convolution channels in parallel, with each channel having different dilation factors.

In comparison to deep non-unfolding networks, the architecture of MsDC-DEQ-Net offers good interpretability as it borrows insights from traditional optimization methods. MsDC-DEQ-Net also allows for structural diversity in its model design, providing ample room for optimizing network structures. Compared to deep unfolding networks, MsDC-DEQ-Net only requires one iteration block, significantly reducing memory requirements and addressing computation issues in large-scale models [15]. Extensive experiments demonstrate that MsDC-DEQ-Net achieves competitive performance compared to existing network-based CS image reconstruction methods.

### 3. Proposed MsDC-DEQ-Net for Image CS

In this section, we will first introduce the relevant concepts that we have utilized and then provide a detailed explanation of the design of the proposed MsDC-DEQ-Net.

3.1. *ISTA Optimization for CS.* Suppose the original signal  $\mathbf{x} \in \mathbb{R}^N$  is CS measured by a linear random projection  $\Phi \in \mathbb{R}^{M \times N}$  giving measurements  $\mathbf{y} \in \mathbb{R}^M$  as follows:

$$\mathbf{y} = \Phi \mathbf{x}, \quad (1)$$

were  $M \ll N$  and the CS ratio is given as  $M/N$ . The purpose of CS reconstruction is to infer  $\mathbf{x}$  from  $\mathbf{y}$ . Traditionally, we obtain the reconstruction by solving the optimization problem as follows:

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\Psi \mathbf{x}\|_1 \right\}, \quad (2)$$

where  $\Psi$  is some sparse transform and the result of  $\Psi \mathbf{x}$  is the coefficients of  $\mathbf{x}$  in a sparse domain. The sparsity of  $\Psi \mathbf{x}$  is encouraged by the  $l_1$  norm with regularization parameter  $\lambda$ .

The problem in Equation (1) can be solved using various optimization algorithms, such as the ISTA [13], alternative direction method of multipliers (ADMM) [36], and AMP [12]. In this paper, we adopt the ISTA algorithm for simplicity. ISTA is a widely used first-order proximal method and is particularly suitable for solving linear inverse problems [14]. Each iteration of the ISTA algorithm consists of the two steps as follows:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho \Phi^T (\Phi \mathbf{x}^{(k-1)} - \mathbf{y}), \quad (3)$$

$$\mathbf{x}^{(k)} = \Psi^{-1} \left\{ (|\Psi \mathbf{r}^{(k)}| - \lambda)_+ \text{sgn}(\Psi \mathbf{r}^{(k)}) \right\}, \quad (4)$$

where  $\mathbf{r}$  is the immediate reconstruction, and  $\rho$  is the step size [13, 14]. The subscript  $+$  in Equation (4) takes the positive part, setting any negative part to zero. In Equation (3) is the gradient descent step where the estimate of the reconstructed signal  $\mathbf{x}$  is updated by taking a step in the direction of the negative gradient of the objective function using the data fidelity term  $\frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2$  from Equation (2). This step is essentially a gradient descent update, where the step size is chosen based on the Lipschitz constant of the gradient of the objective function [13]. Equation (4) is the soft-threshold step, where the soft-thresholding operator is applied to the current estimate of the signal in its sparse domain. To be specific, the operator first sets the elements whose absolute values are below a certain threshold to zero and then shrinks the nonzero coefficients toward zero by the threshold. These two steps are repeated iteratively until convergence is achieved or a maximum number of iterations is reached. The ISTA algorithm is commonly used for solving sparse linear regression and compressed sensing problems [13].

3.2. *STP.* According to STP theory, a small matrix  $\Phi$  can be multiplied with a tall vector  $\mathbf{x}$  as follows:

$$\mathbf{y} = \Phi(t) \bowtie \mathbf{x}, \quad (5)$$

where  $\Phi(t) \in \mathbb{R}^{\frac{M}{t} \times \frac{N}{t}}$ ,  $\bowtie$  is the left product operator for STP, and  $t$  is a shrinkage factor chosen as a common divisor of  $M$  and  $N$  [37–39]. The operation in Equation (5) is equivalent to the following:

$$\mathbf{y} = \{\Phi(t) \otimes I_t\} \mathbf{x}, \quad (6)$$

where  $\otimes$  is the Kronecker product of matrices and  $I_t \in \mathbb{R}^{t \times t}$  is an identity matrix. When  $t = \sqrt{N}$ , Equation (6) can be written in matrix form [25] as follows:

$$\mathbf{Y} = \mathbf{X} \cdot \Phi(t)^T, \quad (7)$$

where  $\mathbf{X} \in \mathbb{R}^{\sqrt{N} \times \sqrt{N}}$ ,  $\mathbf{Y} \in \mathbb{R}^{\sqrt{N} \times \frac{M}{\sqrt{N}}}$ , and  $\Phi(t) \in \mathbb{R}^{\frac{M}{\sqrt{N}} \times \sqrt{N}}$ . Here, we reshape the vectors  $\mathbf{x}$  and  $\mathbf{y}$  to be matrices maintaining column-wise order. The square matrix  $\mathbf{X}$  is easy to associate with a square image.

If  $\Phi(t)$  satisfies the restricted isometry property (RIP) [40], it can be used as a measurement matrix since mutual coherence of  $\Phi(t) \bowtie \Psi$  still satisfies the RIP [41]. The application of STP brings significant convenience to CS. In [41], the authors perform column-wise measurements of an image, reducing the dimensions of the measurement matrix to  $1/t^2$  compared to conventional CS. This reduction greatly reduces the memory footprint.

If we consider an image with a size of  $256 \times 256$  and directly apply CS measurements at a ratio of 10%, the size of the measurement matrix  $\Phi(1)$  would be  $6,554 \times 65,536$ . However, to conserve memory, BCS is commonly employed. In approaches such as [14, 26, 30, 42–44], the image is divided into smaller blocks, typically with a size of  $33 \times 33$ . These blocks are then vectorized and measured individually. In the case of a 10% CS ratio, this means a measurement matrix with a size of  $109 \times 1,089$  would be required, but at the disadvantage of creating blocking artifacts when these blocks are reassembled. As shown in [27], block-based sensing is less efficient compared with global sensing in recovery performance because a measurement in the block-based approach has information only about the block, while a measurement in the global approach has information about the whole image.

Using the same  $256 \times 256$  image, if the image is vectorized directly without breaking it into smaller blocks and STP is applied, when  $t$  is chosen as 256, the measurement matrix  $\Phi(256)$  only has size  $26 \times 256$ . So, the measurement matrix size is smaller than the block-based method that used blocks of size  $33 \times 33$  instead of the full image of size  $256 \times 256$ .

There can be fluctuations in the resulting mutual coherence of different generated measurement matrices, especially when the value of  $t$  is large, resulting in a decreased probability of satisfying the RIP [41]. This compels us to adopt a larger-sized measurement matrix from a smaller value of  $t$ . In accordance with [25], an image can be measured using two steps with two larger measurement matrices as follows:

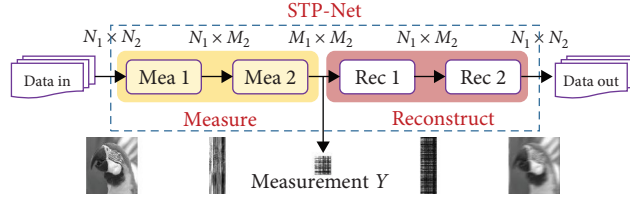


FIGURE 1: STP-Net [25].

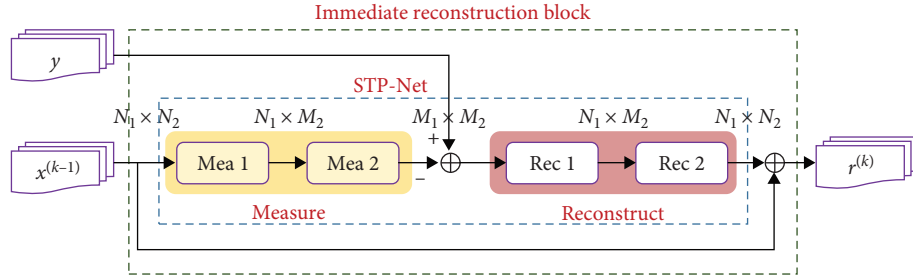


FIGURE 2: Immediate reconstruction block (IRB).

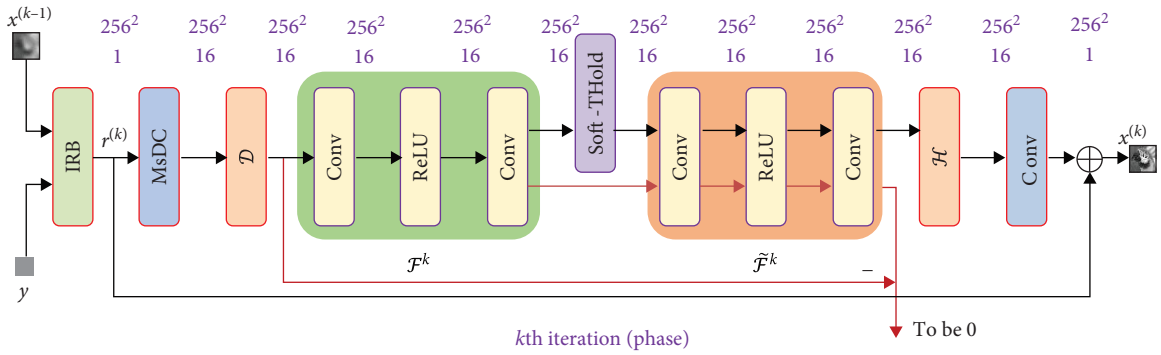


FIGURE 3: ISTA block.

$$Y = \Phi_1(t) \cdot X \cdot \Phi_2(t)^T, \quad (8)$$

where the combined steps result in fewer resulting measurements in  $Y$ . For instance, to achieve a CS ratio equal to 10%, the size of the two measurement matrices  $\Phi_1(256)$  and  $\Phi_2(256)$  are set to  $81 \times 256$ , since  $(81/256)^2 \approx 10\%$ . The memory footprint they occupy is still smaller than a  $33 \times 33$  block-based method. Here,  $\Phi_1(t)$  and  $\Phi_2(t)$  can even be selected to be the same. Based on Equation (8), we build STP-Net as shown in Figure 1 [25], which can measure an image directly without segmenting it into blocks and provides an initial reconstruction. Mea1 and Mea2 correspond to  $\Phi_1(t)$  and  $\Phi_2(t)$ , showing the two measurement steps. Rec1 and Rec2 are the inverse operations of Mea1 and Mea2 for a total of four matrices for the measurement and initial reconstruction phases.

**3.3. Structure of ISTA Iteration Block.** To implement Equation (3), we build the immediate reconstruction block, as shown in Figure 2, by means of STP-Net.

The residuals of an image are known to exhibit higher compressibility [14], and incorporating residual learning can facilitate the training of deeper networks [45]. Let's assume that  $\mathbf{x}^{(k)}$  comprises three components: the immediate reconstruction result  $\mathbf{r}^{(k)}$ , the high-frequency components  $\mathbf{h}^{(k)}$  missing in  $\mathbf{r}^{(k)}$  found using an appropriate operator, and any remaining noise  $\mathbf{n}^{(k)}$  embedded in  $\mathbf{x}^{(k)}$  [14]. We can express  $\mathbf{x}^{(k)}$  as follows:

$$\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathbf{h}^{(k)} + \mathbf{n}^{(k)}. \quad (9)$$

Then, Equation (4) can be reformulated as follows:

$$\mathbf{x}^{(k)} = \mathbf{r}^{(k)} + \mathcal{H} \left\{ \Psi^{-1} \left( \left( |\Psi(\mathcal{D}(\mathbf{r}^{(k)}))| - \lambda \right)_+ \text{sgn}(\Psi(\mathcal{D}(\mathbf{r}^{(k)}))) \right) \right\} + \mathbf{n}^{(k)}, \quad (10)$$

where  $\mathcal{D}$  represents a denoising operation and  $\mathcal{H}$  represents a high-pass filter [14].

Since network-based methods allow for structural diversity originating from the CS domain [14], we construct the ISTA block, as shown in Figure 3, which implements Equations (3)

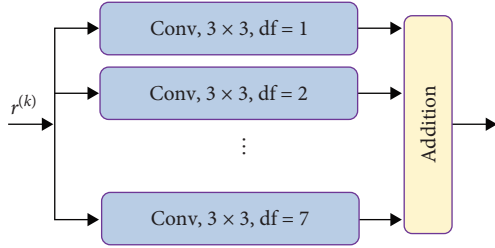


FIGURE 4: Multiscale dilated convolution layer (df: dilation factor).

and (10). All “Conv” layers represent convolutional layers with a filter size of  $3 \times 3$ . The top line in Figure 3 represents the output feature sizes, while the second line represents the number of features. The blocks  $\mathcal{F}^k$  and  $\mathcal{F}^k$  correspond to  $\Psi$  and  $\Psi^{-1}$ , respectively, implementing the sparse transform and its inverse operation [14]. The second output of Figure 3 is enforced to be zero, indicating that a signal passes through  $\mathcal{F}^k$  and  $\mathcal{F}^k$  without any changes.

MsDC (multiscale dilated convolution) is illustrated in Figure 4, which consists of multiple branches with different dilation factors ranging from 1 to 7. This enables the extraction of both structural and detailed information from natural images, as different dilation factors capture different levels of information [32].

The denoise layer  $\mathcal{D}$  and high-pass filter layer  $\mathcal{H}$ , depicted in Figures 5(a) and 5(b), respectively, are inspired by the aggregated residual transformations for deep neural networks (ResNeXt) [19] and the SENet [20]. ResNeXt increases the number of branches in a residual block while the SE network adaptively recalibrates channel-wise feature responses. Since MsDC extracts multifeatures from  $\mathbf{r}^{(k)}$ , the structure of the SE network is expected to enhance important features, while the structure of ResNeXt is expected to improve the performance of the denoise layer and high-pass filter.

**3.4. Proposed MsDC-DEQ-Net.** From Figure 3, an ISTA iteration can be described as follows:

$$\mathbf{x}^{(k)} = f(\mathbf{x}^{(k-1)}, \mathbf{y}), \quad (11)$$

where  $f(\cdot)$  represents the operations of all the layers in Figure 3. Equation (11) merges (3) and (10) into one equation. According to the ISTA algorithm, after a certain number of iterations  $\mathbf{x}^{(k-1)}$  will approach  $\mathbf{x}^{(k)}$ , converging to an equilibrium point. This observation aligns with the concept of an equilibrium model, which has a fixed point  $\mathbf{x}^*$  represented as follows:

$$\mathbf{x}^* = f(\mathbf{x}^*, \mathbf{y}), \quad (12)$$

where  $\mathbf{y}$  serves as the input injection, playing a crucial role in ensuring that the equilibrium point aligns with the original signal [17]. Therefore, we can consider the ISTA block as an equilibrium layer within the model.

The proposed MsDC-DEQ-Net for image CS is illustrated in Figure 6. Both the STP-Net and the equilibrium layer are jointly trained. The STP-Net provides the measurement  $\mathbf{y}$  as the input injection for the equilibrium layer, along with the initial reconstruction  $\mathbf{z}_0$ , which serves as a suitable starting point for the iterative solution within the equilibrium layer. Figure 6 shows three outputs: (1) Output1 aims to approach the original images. (2) Output2 ensures the reversibility of the sparse transform, meaning that the signal remains unchanged as it passes through  $\mathcal{F}^k$  and  $\mathcal{F}^k$ . Therefore, Output2 should equal zero, as depicted in Figure 3. (3) Output3 focuses on the initial reconstruction, striving to approach the equilibrium point efficiently. The parameters of the MsDC-DEQ-Net are optimized by minimizing the half mean square error between the outputs and the expected signals.

## 4. Experimental Results

The ILSVRC2014 ImageNet dataset consists of 1.2 million images with 1,000 object categories and is commonly used in computer vision competitions [46]. For our experiments, we randomly selected 20,000 natural images from this dataset, with 14,000 images used for training, 3,000 for validation, and 3,000 for testing. Each image was cropped to a central  $256 \times 256$  region and converted to an 8-bit grayscale. During training, we employed the Adam solver with a learning rate of  $1e-5$  and a minibatch size of 16. To evaluate the performance of our model, we used two widely used benchmark datasets: Set11 [42] and BSD68 [47]. Set11 contains 11 grayscale images, while BSD68 contains 68 grayscale images. The reconstruction results were reported for compression ratios of 1%, 4%, 10%, and 25%. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used as evaluation criteria.

All experiments were conducted on MATLAB R2019a, running on a computer with an Intel i7-8700K CPU operating at 3.7 GHz, a GeForce GTX 1,080 GPU, and 16 GB RAM. To find the fixed point of the DEQ model, we employed the Anderson acceleration method to improve convergence and prevent divergence. This method determines the promising direction for iterations by updating the input of the deep equilibrium layer with a linear combination of previous outputs. During test time, we set the number of iterations to 50 for the Anderson acceleration method.

**4.1. Performance Comparison.** We conducted a comparison between our proposed MsDC-DEQ-Net and nine recent state-of-the-art image CS methods, namely STP-Net [25], DPA-Net [22], SWDGAN [23], PE-Net [24], ISTA-Net<sup>+</sup> [26], OPINE-Net [26], AMP-Net [30], STP-ISTA-Net [16], and STP-DEQ-Net [16]. The first four are deep non-unfolding networks, while the next four are deep unfolding networks. We evaluated the average PSNR and SSIM reconstruction performance on the Set11 dataset across four CS ratios, as summarized in Table 1. The results for the other methods were obtained from their respective papers.

We anticipate that the proposed technique will perform well compared to the listed competing techniques. For the

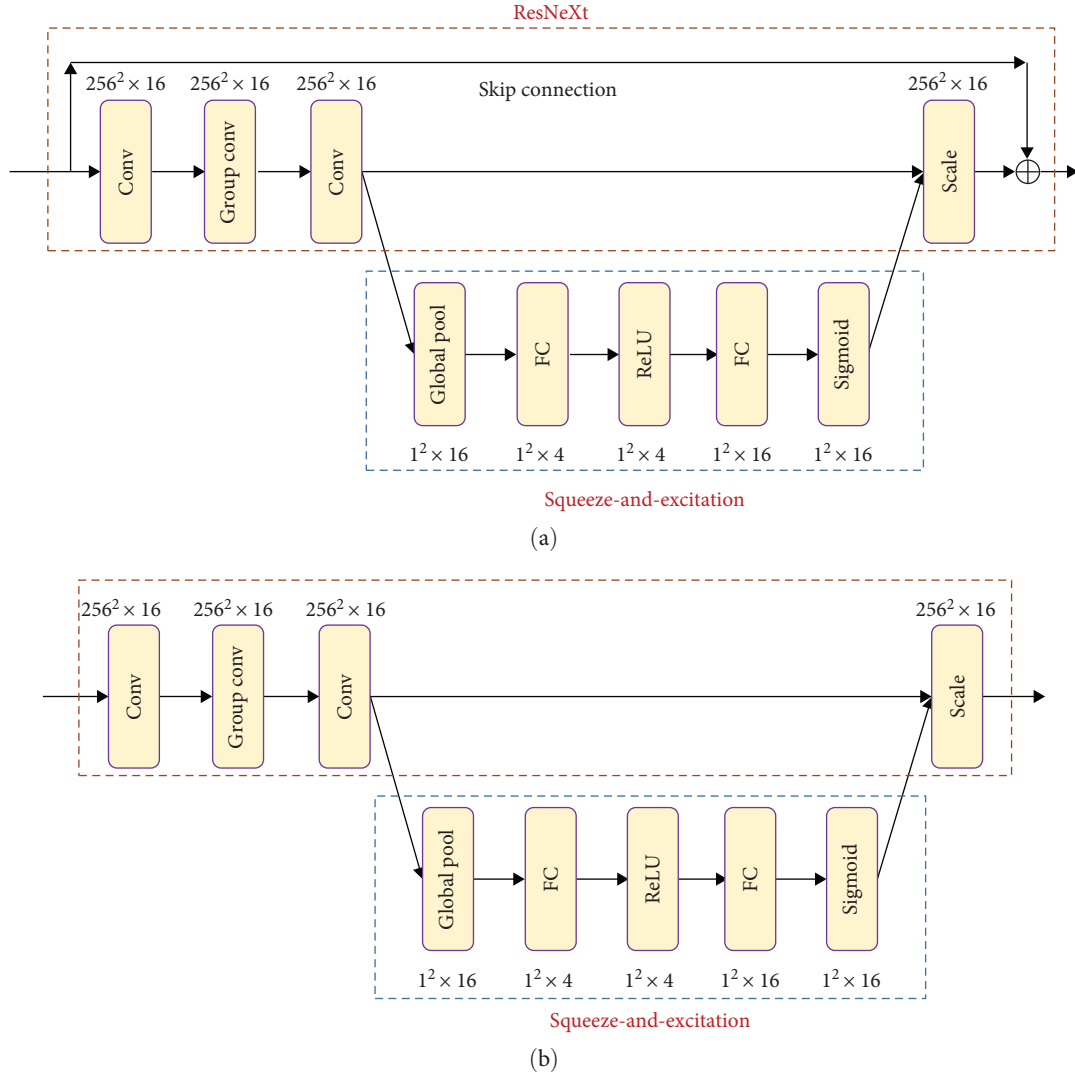


FIGURE 5: Structure of denoise layer and high pass layer: (a) denoise layer  $\mathcal{D}$  [19, 20]; (b) high pass layer  $\mathcal{H}$  [19, 20].

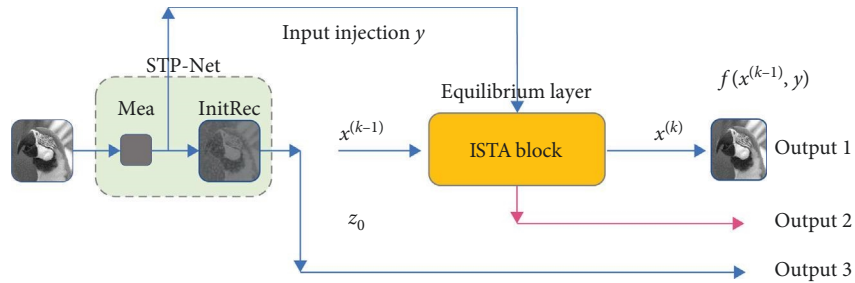


FIGURE 6: MsDC-DEQ-Net.

proposed technique, Figures 5(a) and 5(b) show that we have incorporated aspects of ResNeXt [19] and the SENet [20] into our model. These two networks have good performance, so incorporating these into our proposed network is expected to perform well compared with techniques like ISTA-NET<sup>+</sup>. From Table 1, it is evident that our proposed model outperforms the other methods with higher PSNR and SSIM scores, particularly at the extremely low CS ratio of 1%. Even at CS

ratios of 4% and 10%, our model still achieves superior PSNR. Although our proposed model has slightly lower performance at a CS ratio of 25%, it remains competitive. By increasing the number of iterations in the Anderson acceleration method, we can obtain even better results.

To assess the generalizability of our model, we also compared it with other methods on the larger BSD68 dataset. As shown in Table 2, our proposed model achieves the best

TABLE 1: Mean PSNR (dB) and SSIM comparisons on Set11 at different CS ratios [16].

Algorithm	CR = 1%		CR = 4%		CR = 10%		CR = 25%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
STP-Net [25]	20.73	0.5111	23.39	0.6310	26.08	0.7679	30.06	0.8843
DPA-Net [22]	18.05	0.5011	23.50	0.7205	26.99	0.8354	31.74	0.9238
SWDGAN [23]	21.01	0.5410	24.67	0.6410	28.45	0.8570	33.45	0.9300
PE-Net [24]	20.74	0.4909	25.19	<b>0.7586</b>	28.58	0.8701	33.23	0.9407
ISTA-Net <sup>+</sup> [26]	17.42	0.4029	21.32	0.6037	26.64	0.8087	32.59	0.9254
OPINE-Net [26]	19.87	0.5070	25.04	0.7730	29.33	<b>0.8825</b>	34.44	<b>0.9491</b>
AMP-Net [30]	20.20	0.5581	25.26	0.7722	29.40	0.8779	<b>34.63</b>	0.9481
STP-ISTA-Net [16]	21.32	0.5529	25.47	0.7152	29.01	0.8438	33.72	0.9327
STP-DEQ-Net [16]	21.24	0.5565	24.67	0.7028	28.96	0.8495	32.63	0.9331
MsDC-DEQ-Net	<b>21.43</b>	<b>0.5642</b>	<b>25.55</b>	0.7303	<b>29.96</b>	0.8723	34.09	0.9446

The best performance is labeled in bold.

TABLE 2: Mean PSNR (dB) and SSIM comparisons on BSD68 at different CS ratios.

Algorithm	CR = 1%		CR = 4%		CR = 10%		CR = 25%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
STP-Net [25]	21.06	0.4639	23.11	0.5804	25.11	0.7091	28.48	0.8473
DPA-Net [22]	18.98	0.4643	23.27	0.6096	25.57	0.7267	29.68	0.8763
SWDGAN [23]	22.28	<b>0.5600</b>	24.88	0.6700	27.29	0.7890	30.96	0.9100
PENet [24]	22.10	0.4820	25.06	0.6711	27.38	0.7960	30.97	0.9053
ISTA-Net <sup>+</sup> [26]	19.14	0.4158	22.17	0.5486	25.32	0.7022	29.36	0.8525
OPINE-Net [26]	21.80	0.4972	24.87	0.6709	27.54	0.7966	31.28	0.9034
AMP-Net [30]	22.28	0.5387	<b>25.26</b>	<b>0.6760</b>	27.86	0.7926	31.74	0.9048
STP-ISTA-Net [16]	21.87	0.5046	25.12	0.6639	27.94	0.7932	31.35	0.8956
STP-DEQ-Net [16]	21.38	0.4882	23.74	0.6209	26.58	0.7615	30.48	0.8966
MsDC-DEQ-Net	<b>22.53</b>	0.5316	25.18	0.6724	<b>28.24</b>	<b>0.7988</b>	<b>32.17</b>	<b>0.9136</b>

The best performance is labeled in bold.

performance at CS ratios of 10% and 25%. It achieves the second-best performance at a CS ratio of 4%. At the extremely low CS ratio of 1%, our proposed model exhibits higher PSNR but slightly inferior SSIM compared to SWDGAN and AMP-Net.

For visualization purposes, Figure 7 shows an original image of a parrot. Figure 8 shows the CS reconstruction with different techniques. By zooming in on the local area around the parrot’s eye, it is seen that the proposed method has a more realistic reconstruction than the other methods shown. For instance, the stripes around the eye appear to be better reconstructed in the proposed approach, which we expect is due to the multiscale dilated convolution network introduced into the model.

**4.2. Ablation Studies.** The multiscale dilated convolution model has demonstrated remarkable performance due to its ability to extract features at different scales from an image, where the combination of these features contributes to better reconstruction [35]. Previous studies [32–35] have often utilized dilation factors of 1, 2, and 3. In our proposed model, we intentionally incorporated seven branches with different dilation factors, as depicted in Figure 4, to better observe the impact of these factors. The branches are labeled from 1 to 7 based on their corresponding dilation factor values. To

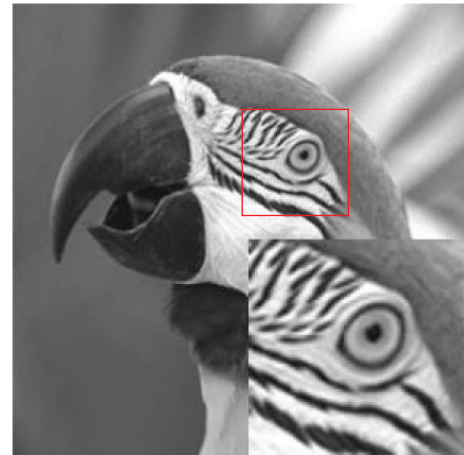


FIGURE 7: Original image of the parrot with zoomed-in region around the eye.

facilitate training, at the outset, we assigned the seven branches with the same parameter values as a pre-trained STP-DEQ-Net, with the exception of the dilation factor. We conducted several ablation studies to analyze the effects, as shown in Tables 3–7 [48].



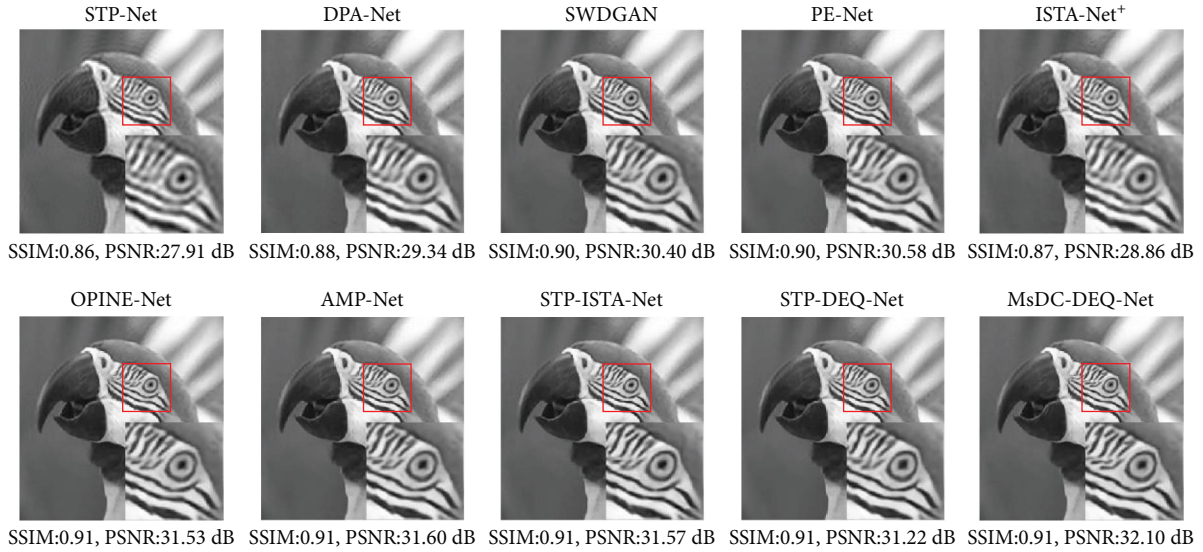


FIGURE 8: Reconstruction of parrot image with 10 different methods (corresponds with Table 1 when CR = 10%).

TABLE 3: Mean SSIM and mean PSNR (dB) on Set11 with one scale dilated convolution model (after [48]).

Connected branch	All	None	1	2	3	4	5	6	7
SSIM	<b>0.8723</b>	0.6756	0.8595	0.7198	0.6973	0.7007	0.7006	0.7000	0.6989
PSNR	<b>29.96</b>	25.63	29.36	26.13	25.88	25.91	25.90	25.90	25.89

Bold indicates best performance.

TABLE 4: Mean SSIM and mean PSNR (dB) on Set11 with two scales dilated convolution model (after [48]).

Connected branch	1 + 2	1 + 3	1 + 4	1 + 5	1 + 6	1 + 7
SSIM	<b>0.8650</b>	0.8624	0.8628	0.8625	0.8586	0.8605
PSNR	<b>29.57</b>	29.51	29.51	29.50	29.32	29.41

Bold indicates best performance.

TABLE 5: Mean SSIM and mean PSNR (dB) on Set11 with three scales dilated convolution model (after [48]).

Connected branch	1 + 2 + 3	1 + 2 + 4	1 + 2 + 5	1 + 2 + 6	1 + 2 + 7
SSIM	0.8679	<b>0.8683</b>	0.8680	0.8647	0.8664
PSNR	29.71	<b>29.72</b>	29.71	29.55	29.63

Bold indicates best performance.

TABLE 6: Mean SSIM and mean PSNR (dB) on Set11 with six scales dilated convolution model (after [48]).

Disconnected branch	None	All	1	2	3	4	5	6	7
SSIM	<b>0.8723</b>	0.6756	0.6635	0.8674	0.8720	0.8715	0.8713	0.8724	0.8703
PSNR	<b>29.96</b>	25.63	25.49	29.77	29.93	29.92	29.90	29.94	29.84

Bold indicates best performance.

TABLE 7: Mean SSIM and mean PSNR (dB) on Set11 with multiscale dilated convolution model (after [48]).

Disconnected branch	None	All	3 + 6	3 + 6 + 1	3 + 6 + 2	3 + 6 + 4	3 + 6 + 5	3 + 6 + 7
SSIM	<b>0.8723</b>	0.6756	0.8718	0.6644	0.8665	0.8703	0.8692	0.8699
PSNR	<b>29.96</b>	25.63	29.91	25.50	29.69	29.83	29.76	29.80

Bold indicates best performance.

The results in Table 3 indicate that when all seven parallel branches are connected (denoted as “under all”), the model effectively learns the residual and achieves good performance. Conversely, when no branches are connected (denoted as “under none”), the model essentially produces an immediate reconstruction  $r^{(k)}$ , which requires improvement in terms of quality. Notably, when only one branch is connected, branch 1, with a dilation factor of 1, outperforms branches 2–7.

Tables 4 and 5 present additional ablation studies where 2 or 3 branches in Figure 4 are connected while the remaining branches are removed. Branch 1 is retained in all cases due to its outstanding performance, as demonstrated in Table 3. When fusing the features of another branch with branch 1, it is observed that branch 2 has a greater influence compared to the other branches. Furthermore, by incorporating branch 4, the overall performance is further enhanced.

Tables 6 and 7 present the performance of the proposed model when some of the seven branches in Figure 4 are removed. From Table 6, it is evident that branch 1 is the most important, as its removal leads to significant performance degradation. In contrast, the removal of branch 3 and branch 6 has a much smaller impact on model performance. Table 7 provides further insights, demonstrating that branch 2 is more important than branches 4, 5, and 7. Removing branch two results in more degradation compared to the removal of the other branches. Overall, these findings highlight the varying importance of the different branches, with branch 1 and branch 2 playing crucial roles in the model’s performance.

## 5. Conclusion and Future Work

Inspired by the concepts of DUM, we present a novel approach called MsDC-DEQ-Net, which combines the DEQ based on the ISTA algorithm with multiscale dilated convolution for image CS. By mapping a single iteration of the ISTA algorithm to a deep learning block and using it as a deep equilibrium layer, our model maintains a clear interpretability. Extensive experiments demonstrate that the proposed model achieves competitive performance when compared to state-of-the-art CS methods.

In order to leverage the structural diversity originating from the CS domain [13], we incorporate ResNeXt to enhance performance and the SE block to eliminate redundancy and enhance valuable information. Compared to DUM, our proposed model significantly reduces the number of learnable parameters by utilizing only one optimization iteration block.

In future research, we plan to explore the robustness of the proposed model and its application in other fields. Additionally, we recognize the importance of high-throughput methods that facilitate information transition by incorporating multiple channels in the input and output of the iteration block.

## Data Availability

Two datasets were used in the experimental study by Russakovsky et al. [46] and Martin et al. [47].

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was partially funded by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

- [1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] J. Huang and F. Yang, “Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 968–971, IEEE, Barcelona, Spain, 2012.
- [4] A. M. Assem and R. M. Dansereau, “Compressive sensing using S-transform in pulse radar,” in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 488–492, IEEE, Barcelona, Spain, 2017.
- [5] P. Firoozi, S. Rajan, and I. Lambadaris, “Efficient Kronecker-based sparse one-time sensing matrix for compressive sensing cryptosystem,” in *2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, pp. 354–359, IEEE, Athens, Greece, 2021.
- [6] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, “Rank minimization for snapshot compressive imaging,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2990–3006, 2019.
- [7] B. Huang, X. Yan, J. Zhou, and Y. Fan, “CSMCNet: scalable video compressive sensing reconstruction with interpretable motion estimation,” pp. 1–12, 2021, [Online]. Available: <http://arxiv.org/abs/2108.01522>.
- [8] E. Belyaev, “An efficient compressive sensed video codec with inter-frame decoding and low-complexity intra-frame encoding,” *Sensors*, vol. 23, no. 3, Article ID 1368, 2023.
- [9] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [10] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [11] D. Needell and J. A. Tropp, “CoSaMP: iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [12] 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. 18914–18919, 2009.
- [13] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [14] J. Zhang and B. Ghanem, “ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, IEEE, Salt Lake City, UT, USA, 2018.

- [15] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1123–1133, 2021.
- [16] Y. Yu and R. M. Dansereau, "STP-DEQ-Net: a deep equilibrium model based on ISTA method for image compressive sensing," in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 887–891, IEEE, Belgrade, Serbia, 2022.
- [17] D. Duvenaud, J. Z. Kolter, and M. Johnson, "Deep implicit layers: neural ODEs, equilibrium models, and beyond," NeurIPS workshop. [Online]. Available: <http://implicit-layers-tutorial.org/>.
- [18] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for video snapshot compressive imaging," 2022, [Online]. Available: <http://arxiv.org/abs/2201.06931>.
- [19] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, IEEE, Honolulu, HI, USA, 2017.
- [20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, CVF, [http://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Hu\\_Squeeze-and-Excitation\\_Networks\\_CVPR\\_2018\\_paper.html](http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html), 2018.
- [21] J. Zhang, Z. Zhang, J. Xie, and Y. Zhang, "High-throughput deep unfolding network for compressive sensing MRI," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 750–761, 2022.
- [22] Y. Sun, J. Chen, Q. Liu, B. Liu, and G. Guo, "Dual-path attention network for compressed sensing image reconstruction," *IEEE Transactions on Image Processing*, vol. 29, pp. 9482–9495, 2020.
- [23] Y. Tian, X. Chai, Z. Gan, Y. Lu, Y. Zhang, and S. Song, "SWDGAN: GAN-based sampling and whole image denoising network for compressed sensing image reconstruction," *Journal of Electronic Imaging*, vol. 30, no. 6, pp. 1–22, 2021.
- [24] W. Ma, X. Wu, S. Sirkemaa, and M. O. Agyeman, "Parallel enhanced network for image compressed sensing," *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2022)*, vol. 12247, Article ID 122470D, 2022.
- [25] Y. Yu and R. M. Dansereau, "STP-Net: semi-tensor product neural network for image compressive sensing," in *The Seventh International Conference on Advances in Signal, Image and Video Processing (SIGNAL 2022)*, pp. 7–12, Carleton University, 2022.
- [26] J. Zhang, C. Zhao, and W. Gao, "Optimization-inspired compact deep compressive sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 765–774, 2020.
- [27] E. Belyaev, M. Codreanu, M. Juntti, and K. Egiazarian, "Compressive sensed video recovery via iterative thresholding with random transforms," *IET Image Processing*, vol. 14, no. 6, pp. 1187–1199, 2020.
- [28] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.
- [29] M. S. Asif, F. Fernandes, and J. Romberg, "Low-complexity video compression and compressive sensing," in *2013 Asilomar Conference on Signals, Systems and Computers*, pp. 579–583, IEEE, Pacific Grove, CA, USA, 2013.
- [30] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "AMP-Net: denoising-based deep unfolding for compressive image sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 1487–1500, 2021.
- [31] S. Bai, J. Z. Kolter, and V. Koltun, "Deep equilibrium models," *Advances in Neural Information Processing Systems*, vol. 32, pp. 1–12, 2019.
- [32] Y. Wang, G. Wang, C. Chen, and Z. Pan, "Multi-scale dilated convolution of convolutional neural network for image denoising," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19945–19960, 2019.
- [33] N. K. Chowdhury, M. M. Rahman, and M. A. Kabir, "PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–14, 2020.
- [34] G. Lin, Q. Wu, L. Qiu, and X. Huang, "Image super-resolution using a dilated convolutional neural network," *Neurocomputing*, vol. 275, pp. 1219–1230, 2018.
- [35] Z. Wang, Z. Wang, C. Zeng, Y. Yu, and X. Wan, "High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network," *Circuits, Systems, and Signal Processing*, vol. 42, pp. 1593–1616, 2023.
- [36] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [37] D. Cheng, H. Qi, and A. Xue, "A survey on semi-tensor product of matrices," *Journal of Systems Science and Complexity*, vol. 20, no. 2, pp. 304–322, 2007.
- [38] D.-Z. Cheng and L.-J. Zhang, "On semi-tensor product of matrices and its applications," *Acta Mathematicae Applicatae Sinica, English Series*, vol. 19, no. 2, pp. 219–228, 2003.
- [39] D. Cheng and Y. Dong, "Semi-tensor product of matrices and its some applications to physics," *Methods and Applications of Analysis*, vol. 10, no. 4, pp. 565–588, 2003.
- [40] R. G. Baraniuk, M. A. Davenport, M. F. Duarte, and C. Hegde, "An introduction to compressive sensing," 2011, Connexions e-textbook [Online]. Available: <http://cnx.org/content/col11133/latest>.
- [41] J. Wang, Z. Xu, Z. Wang, S. Xu, and J. Jiang, "Rapid compressed sensing reconstruction: a semi-tensor product approach," *Information Sciences*, vol. 512, pp. 693–707, 2020.
- [42] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: non-iterative reconstruction of images from compressively sensed measurements," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 449–458, IEEE, Las Vegas, NV, USA, 2016.
- [43] H. Yao, F. Dai, S. Zhang, Y. Zhang, Q. Tian, and C. Xu, "DR2-Net: deep residual reconstruction network for image compressive sensing," *Neurocomputing*, vol. 359, pp. 483–493, 2019.
- [44] N. Li and C. C. Zhou, "AMPA-Net: optimization-inspired attention neural network for deep compressed sensing," in *2020 IEEE 20th International Conference on Communication Technology (ICCT)*, pp. 1338–1344, IEEE, Nanning, China, 2020.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, IEEE, Las Vegas, NV, USA, 2016.
- [46] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

- [47] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, pp. 416–423, IEEE, Vancouver, BC, Canada, 2001.
- [48] Y. Yu, *Reconstruction of compressive sensed (CS) images with deep equilibrium model (DEQ) based on iterative shrinkage-thresholding algorithm (ISTA)*, Ph.D. Thesis, Carleton University, Ottawa, Canada, 2023.