

Article

Multi-Channel Representation Learning Enhanced Unfolding Multi-Scale Compressed Sensing Network for High Quality Image Reconstruction

Chunyan Zeng ¹, Shiyan Xia ¹, Zhifeng Wang ^{2,*} and Xiangkui Wan ¹

¹ Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China

² Department of Digital Media Technology, Central China Normal University, Wuhan 430079, China

* Correspondence: zfwang@ccnu.edu.cn

Abstract: Deep Unfolding Networks (DUNs) serve as a predominant approach for Compressed Sensing (CS) reconstruction algorithms by harnessing optimization. However, a notable constraint within the DUN framework is the restriction to single-channel inputs and outputs at each stage during gradient descent computations. This constraint compels the feature maps of the proximal mapping module to undergo multi-channel to single-channel dimensionality reduction, resulting in limited feature characterization capabilities. Furthermore, most prevalent reconstruction networks rely on single-scale structures, neglecting the extraction of features from different scales, thereby impeding the overall reconstruction network's performance. To address these limitations, this paper introduces a novel CS reconstruction network termed the Multi-channel and Multi-scale Unfolding Network (MMU-Net). MMU-Net embraces a multi-channel approach, featuring the incorporation of Adap-SKConv with an attention mechanism to facilitate the exchange of information between gradient terms and enhance the feature map's characterization capacity. Moreover, a Multi-scale Block is introduced to extract multi-scale features, bolstering the network's ability to characterize and reconstruct the images. Our study extensively evaluates MMU-Net's performance across multiple benchmark datasets, including Urban100, Set11, BSD68, and the UC Merced Land Use Dataset, encompassing both natural and remote sensing images. The results of our study underscore the superior performance of MMU-Net in comparison to existing state-of-the-art CS methods.

Keywords: compressed sensing; image reconstruction; deep unfolding network; attention mechanism



Citation: Zeng, C.; Xia, S.; Wang, Z.; Wan, X. Multi-Channel Representation Learning Enhanced Unfolding Multi-Scale Compressed Sensing Network for High Quality Image Reconstruction. *Entropy* **2023**, *25*, 1579. <https://doi.org/10.3390/e25121579>

Academic Editors: Leandro Alves Neves, Marcelo Zanchetta do Nascimento and Thaina Aparecida Azevedo Tosta

Received: 27 October 2023
Revised: 17 November 2023
Accepted: 22 November 2023
Published: 24 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Compressed Sensing (CS) has revolutionized the limitations of the Nyquist sampling theorem, enabling the efficient reconstruction of signals at significantly lower sampling rates than the traditional Nyquist rate [1], particularly for signals exhibiting inherent sparsity or sparsity within specific transform domains [2]. This innovation has profound implications, substantially reducing the cost of sensor data compression, and mitigating the demands on transmission bandwidth and storage capacity in data transmission processes. CS has found wide applications, ranging from single-pixel cameras [3,4] to snapshot compression imaging [5,6] and even magnetic resonance imaging [7,8].

CS reconstruction methods can be broadly categorized into two main classes: traditional CS reconstruction methods [9–16] and deep-learning-based CS reconstruction methods [17–21]. Traditional CS reconstruction methods are designed based on a priori knowledge of image sparsity, presuming that the signal exhibits sparsity within a particular transform domain [22,23]. These methods formulate signal reconstruction as an optimization problem within a sparse model framework [12]. Solving this problem involves iterative approaches employing convex optimization methods, greedy algorithms, or Bayesian-like

techniques to obtain the reconstructed signal. While traditional CS reconstruction methods provide strong convergence and theoretical guidance, they suffer from drawbacks such as computational intensity, slow reconstruction speeds, and limited reconstruction performance [24].

The computational complexity inherent in traditional CS reconstruction methods presents challenges in achieving real-time image reconstruction. To address this, deep learning methods, known for their prowess in image processing, have been introduced into the realm of CS reconstruction. Deep-learning-based CS reconstruction algorithms can be broadly classified into two primary categories: deep non-unfolding networks (DNUNs) [18,19,21,25,26] and deep unfolding networks (DUNs) [8,27–33]. DNUN treats the reconstruction process as a black-box operation, relying on a data-driven approach to build an end-to-end neural network to address the CS reconstruction problem. In this paradigm, the Gaussian random measurement matrix used in traditional CS reconstruction methods is replaced with a learnable measurement network. Subsequently, the reconstruction network framework is constructed around well-established deep learning models such as stacked denoising autoencoders [25], convolutional neural networks (CNNs) [18], or residual networks [26] to learn the mapping from CS measurements to reconstructed signals. Despite the ability of DNUN to achieve real-time reconstruction, surpassing traditional CS reconstruction methods, it has limitations such as high data dependency and poor interpretability, stemming from its entirely data-driven nature and lack of a strong theoretical foundation.

Conversely, DUN combines traditional optimization methods with deep learning techniques, utilizing optimization algorithms as theoretical guides. It employs a fixed-depth neural network to simulate the finite number of iterations of the optimization algorithm, resulting in reconstructed signals. Many optimization algorithms, such as Approximate Message Passing (AMP) [34], Iterative Shrinkage Thresholding Algorithm (ISTA) [35], and the Alternate Direction Multiplier Method (ADMM) [36], have been incorporated into DUN, leading to superior reconstruction performance compared to DNUN. Due to its foundation in theoretically guaranteed optimization algorithms, DUN offers strong reconstruction performance and a degree of interpretability.

Nonetheless, DUN typically operates in a single-channel form in many cases [27–30,37,38], as feature maps within the deep reconstruction network are transmitted between phases and updated within each phase. This structural characteristic limits the characterization ability of the feature maps, ultimately degrading the network's reconstruction performance. Moreover, mainstream DUN methods [28–30,33,37,38] often rely on standard CNNs to build the reconstruction network, with each CNN featuring uniform receptive fields. As the human visual system is a multi-channel model, a series of receptive fields of different sizes are generated in the higher-order areas of the human visual system [39–41]. Therefore, the single receptive field of the standard CNN is inconsistent with the actual observation of the human visual system, which hampers the characterization ability of the CNN.

To address these limitations, this paper introduces two modules within the Deep Reconstruction Subnet (DRS) of our proposed Multi-channel and Multi-scale Unfolding Network (MMU-Net): the Attention-based Multi-channel Gradient Descent Module (AMGDM) and the Multi-scale Proximal Mapping Module (MPMM). These modules are designed to enhance feature characterization and representation in DUN. AMGDM facilitates the transmission of feature maps in a multi-channel format, both intra-stage and inter-stage. This design enhances the feature maps' characterization ability. Moreover, inspired by SK-Net [42], we introduce Adap-SKConv, an attention convolution kernel with a feature fusion mechanism. Adap-SKConv is used to obtain fused gradient terms with attention, further improving the feature representation in AMGDM. To address the limitation of single-scale CNNs, we introduce MPMM, which employs multi-scale CNN. Inspired by the fact that the human visual system has different receptive fields in higher-order areas, in this paper, we utilize the Inception structure [43] and design Multi-scale Block (MB) with multiple parallel convolutional branches in MPMM to simulate the human visual

system using different receptive fields to extract features, thus enhancing the network's representational capability.

The main contributions of this paper are as follows:

- We introduce a novel end-to-end sampling and reconstruction network, named the Multi-channel and Multi-scale Unfolding Network (MMU-Net), comprising three integral components: the Sampling Subnet (SS), Initialize Subnet (IS), and Deep Reconstruction Subnet (DRS).
- Within the Deep Reconstruction Subnet (DRS), the Attention-based Multi-channel Gradient Descent Module (AMGDM) is developed. This module introduces a multi-channel strategy that effectively addresses the challenge of limited feature map characterization associated with the conventional single-channel approach. Additionally, we design the Adap-SKConv attention convolution kernel with a feature fusion mechanism, enhancing the feature characterization of gradient terms. These innovations collectively contribute to a substantial improvement in the network's reconstruction performance.
- In DRS, we introduce the Multi-scale Proximal Mapping Module (MPMM). MPMM incorporates a Multi-scale Block (MB) featuring multiple parallel convolutional branches, facilitating the extraction of features across various receptive fields. This innovation allows for the acquisition of multi-scale features, significantly enhancing the characterization capabilities of the Convolutional Neural Network and thereby leading to an enhanced reconstruction performance.
- Empirical evidence from a multitude of experiments demonstrates the superior performance of the proposed method in comparison to existing state-of-the-art networks. This extensive validation underscores the efficacy and rationality of our approach.

The rest of the paper is organized as follows. Section 2 describes the related work of DNUN and DUN. Section 3 describes the preparatory knowledge for the work of this paper and Section 4 describes the framework and details of MMU-Net. Section 5 describes the experimental parameter settings, baseline, comparison with other state-of-the-art methods and ablation experiments. Section 6 draws the conclusions of the study.

2. Related Work

Deep-learning-based Compressed Sensing (DLCS) reconstruction networks can be categorized into two primary types: Deep Non-unfolding Networks and Deep Unfolding Networks. This section provides an exploration of the relevant work within each classification.

2.1. Deep Non-Unfolding Network (DNUN)

DNUN is characterized by its creation of end-to-end networks designed to execute the CS sampling and reconstruction processes. This approach leverages a data-driven strategy to acquire the knowledge necessary to map CS measurements into reconstructed signals. The initial foray into integrating deep learning into CS reconstruction was led by Mousavi et al. [25]. Their work employed stacked denoising autoencoders and feed-forward deep neural networks for signal reconstruction.

Subsequently, Kulkarni et al. [18] introduced ReconNet, which capitalized on fully connected layers and convolutional neural networks to reconstruct images. By substituting some of the fully connected layers with CNNs, ReconNet achieved superior performance, particularly in the realm of image processing. Yao et al. [26] presented DR2-Net, which initiated image reconstruction from CS measurements using fully connected layers. A residual network was then incorporated to further refine signal reconstruction.

Distinguishing itself from earlier CS reconstruction methods reliant on random Gaussian measurement matrix sampling, Shi et al. proposed CSNet [44]. This innovative approach harnessed CNNs to not only simulate the sampling process but also concurrently construct the sampling network, resulting in commendable reconstruction outcomes.

Building upon the foundation of CSNet, Shi et al. pursued several enhancements, introducing CSNet+ [45] and SCSNet [46]. These iterations further improved network reconstruction performance. However, DNUN's significant drawback lies in its heavy reliance on data, inhibiting its versatility. Moreover, DNUN's network structure is a product of a generic model, lacking theoretical grounding and interpretability due to deep learning's inherent black-box nature, which can impede further optimization.

2.2. Deep Unfolding Network (DUN)

DUN represents a fusion of efficient deep learning models and optimization algorithms to construct deep reconstruction networks with pre-defined stages. Drawing inspiration from the Iterative Shrinkage Thresholding Algorithm, Zhang et al. introduced ISTA-Net and ISTA-Net+ [28]. These models unfolded each iteration into a network stage using CNNs, offering a promising balance between reconstruction performance and interpretability.

Zhang et al. further refined the concept with OPINE-Net+ [30], which replaced the random Gaussian measurement matrix with a learnable sampling matrix. This matrix incorporated orthogonal and binary constraints, while CNNs simulated the sampling and initial reconstruction processes, resulting in an adaptive end-to-end sampling and reconstruction network that notably improved reconstruction performance.

Building on the foundation of ISTA-Net+, You et al. introduced ISTA-Net++ [37]. This dynamic unfolding strategy addressed the challenge of CS sampling and reconstruction at varying sampling rates within a single model. The introduction of a cross-block strategy mitigated the chunking effect and further bolstered reconstruction performance.

Additionally, Zhang et al. conceived AMP-Net [29] based on the denoising perspective of the Approximate Message Passing algorithm. This model fashioned a sampling network through a random Gaussian matrix and crafted an unfolding network for deep reconstruction employing CNNs. This approach translated into highly efficient image reconstruction.

Song et al. addressed shortcomings in current DUN models related to short-term memory mechanisms. Their proposal, MAPUN [47], incorporated two distinct memory enhancement mechanisms, effectively reducing information loss between phases. This enhancement significantly improved the network's expressive capacity and reconstruction performance.

Summary: DUN surpasses both DNUN and traditional CS reconstruction methods in terms of reconstruction performance and interpretability. Consequently, it has become the prevailing approach in the field of CS reconstruction. Nevertheless, DUN is challenged by the need for multiple multi-channel to single-channel dimensional transformations during the reconstruction process, which can result in a loss of information and reduced feature map characterization capabilities. Additionally, the reliance on single-scale CNNs for reconstruction limits the network's ability to extract image features from a single scale.

3. Preliminaries

This section provides a foundation for understanding the paper's key concepts. It begins with a model of the Compressed Sensing task and subsequently introduces the Iterative Shrinkage Thresholding Algorithm and the Deep Unfolding Network framework based on ISTA. In this paper, vectors are represented using lowercase bold letters, matrices with uppercase bold letters, and parameters with italics. The important mathematical symbols and descriptions in this paper are shown in Table 1:

Table 1. Mathematical notation and description.

Notations	Descriptions
k	Deep reconstruction sub-network stage index
\mathbf{X}	Origin image, $\mathbf{X} \in R^N$
\mathbf{Y}	Measurement, $\mathbf{Y} \in R^M$
r	CS ratio, $r = M/N$
Φ, Φ^\top	The sampling matrix, transpose of the sampling matrix
$\mathbf{F}_\Phi(\bullet), \mathbf{F}_{\Phi^\top}(\bullet)$	Sampling convolutional layer, initialize convolution layer
$\mathbf{X}^{(0)}, \mathbf{X}^{(k)}$	Initialize image, reconstruction image of the k th stage
$\tilde{\mathbf{X}}^{(k-1)}, \hat{\mathbf{X}}^{(k-1)}, \tilde{\mathbf{X}}^{(0)}$	The multi-channel versions of $\mathbf{X}^{(k-1)}, \Phi^\top(\Phi\mathbf{X}^{(k-1)})$, and $\Phi^\top\mathbf{Y}$
$\mathbf{Z}^{(k)}, \tilde{\mathbf{Z}}^{(k)}$	The preliminary instant reconstruction result and the instant reconstruction result of the k th stage
$\mathbf{F}_{gp}, \mathbf{F}_{fc}$	The global average pooling, the two-layered fully connected layer
$\theta^{(k)}$	The threshold for the k th stage soft threshold function
$\rho^{(k)}$	The step size of the k th stage
\mathbf{X}^{final}	Final reconstruction image

3.1. Problem Definition

Definition 1 (Compressed sensing problem). *The CS task encompasses two core components: sampling and reconstruction. Mathematically, the process of CS sampling can be expressed as follows (Equation (1)):*

$$\mathbf{Y} = \Phi\mathbf{X} \tag{1}$$

Here, $\mathbf{X} \in R^N$ signifies the original signal, $\mathbf{Y} \in R^M$ represents the measurement, $\Phi \in R^{M \times N}$ is the random measurement matrix, and $r = M/N$ denotes the sampling rate.

The CS reconstruction problem can be viewed as an ill-posed inverse problem. Traditional CS reconstruction methods approach this by solving Equation (2):

$$\min_{\mathbf{X}} \frac{1}{2} \|\Phi\mathbf{X} - \mathbf{Y}\|_2^2 + \lambda\Psi(\mathbf{X}) \tag{2}$$

Here, $\frac{1}{2} \|\Phi\mathbf{X} - \mathbf{Y}\|_2^2$ represents a data fidelity term, $\Psi(\mathbf{X})$ serves as a regularization term, ensuring that the solution adheres to prior information about the image, and λ denotes a regularization parameter.

3.2. Definitions and Concepts

Definition 2 (ISTA-based DUN framework). *ISTA, a class of gradient algorithms, provides a classical approach for solving linear inverse problems. It accomplishes this by iterating through the following two main steps:*

$$\mathbf{Z}^{(k)} = \mathbf{X}^{(k-1)} - \rho^{(k)} \left(\Phi^\top \left(\Phi\mathbf{X}^{(k-1)} \right) - \Phi^\top\mathbf{Y} \right) \tag{3}$$

$$\mathbf{X}^{(k)} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \mathbf{Z}^{(k)}\|_2^2 - \lambda \|\Psi(\mathbf{X})\|_1 \tag{4}$$

In Equation (3), $\rho^{(k)}$ denotes the step size, k represents the number of iterations, and $\Phi^\top \left(\Phi\mathbf{X}^{(k-1)} \right) - \Phi^\top\mathbf{Y}$ is the gradient of the data fidelity term in Equation (2). Equation (3) demonstrates that $\mathbf{X}^{(k-1)}$ is updated in the direction of the negative gradient of the data fidelity term to produce the instant reconstruction result $\mathbf{Z}^{(k)}$. Equation (4) showcases that the reconstruction result of the k th stage seeks $\mathbf{X}^{(k)}$, approximating it to $\mathbf{Z}^{(k)}$. Equation (4) can be viewed as a specialized form of proximal mapping, which can be converted to:

$$\mathbf{X}^{(k)} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{F}(\mathbf{X}) - \mathbf{F}(\mathbf{Z}^{(k)})\|_2^2 + \theta \|\mathbf{F}(\mathbf{X})\|_1 \tag{5}$$

Here, $\mathbf{F}(\bullet)$ is a nonlinear sparse transform, and ISTA employs a soft threshold function to solve Equation (5):

$$\mathbf{X}^{(k)} = \tilde{\mathbf{F}}\left(\text{soft}\left(\mathbf{F}\left(\mathbf{Z}^{(k)}\right), \theta^{(k)}\right)\right) \tag{6}$$

In Equation (6), $\tilde{\mathbf{F}}(\bullet)$ represents the inverse transformation of $\mathbf{F}(\bullet)$, and $\text{soft}(\bullet, \theta^{(k)})$ denotes the soft threshold function.

The ISTA-based DUN network, based on Equations (3) and (6), establishes the network framework. The reconstruction network comprises N_p stages, each encompassing a Gradient Descent Module (GDM) and Proximal Mapping Module (PMM), as depicted in Figure 1. The GDM corresponds to Equation (3) and simulates ISTA’s iterative step. It accepts the reconstructed image $\mathbf{X}^{(k-1)}$ from the preceding stage as input and generates the instant reconstruction result $\mathbf{Z}^{(k)}$ for the current stage. The GDM involves matrix operations on the feature maps without neural network participation, resulting in single-channel feature maps.

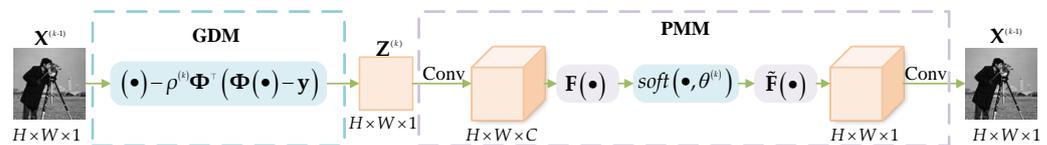


Figure 1. ISTA-based DUN network framework.

In the PMM, two nonlinear transformations, $\mathbf{F}(\bullet)$ and $\tilde{\mathbf{F}}(\bullet)$, designed based on Equation (6), typically consist of CNN modules. The input to PMM is a single-channel $\mathbf{Z}^{(k)}$, initially converted into a multi-channel feature map through convolution. The multi-channel feature maps are then sequentially processed by $\mathbf{F}(\bullet)$, the soft thresholding function, and $\tilde{\mathbf{F}}(\bullet)$ to obtain a multi-channel feature map. Since GDM’s input is single-channel, and it operates solely on feature maps through matrix operations, feature maps remain single-channel throughout. However, PMM’s input and output are restricted to single channels. As a result, the module transforms input from multi-channel to single-channel, which results in information loss and constrains feature map characterization. Additionally, $\mathbf{F}(\bullet)$ and $\tilde{\mathbf{F}}(\bullet)$ are single-scale CNNs, which limits the network’s feature extraction capability.

Definition 3 (CS ratio). In this paper, $\mathbf{X} \in \mathbb{R}^N$ signifies the original signal, $\mathbf{Y} \in \mathbb{R}^M$ represents the measurement. The CS ratio is denoted by r , with $r = M/N$.

Definition 4 (Multi-channel Representation Learning). In this paper, “multi-channel” refers to the presence of multi-channel feature maps, meaning that the output of a network layer consists of feature maps with more than one channel. In contrast, “single-channel” feature maps have only one channel. Multi-channel feature maps can capture more diverse information than their single-channel counterparts.

Definition 5 (Multi-scale CS Network). The term “multi-scale” denotes the structure of a multi-scale network, which employs various convolutional kernels with different receptive fields, constructed in parallel to extract image features from different scales. This differs from a “single-scale” network that relies on a single type of convolutional kernel. Multi-scale networks can extract richer features.

4. Proposed Method

In this section, we introduce the MMU-Net, which consists of three key sub-networks: the Sampling Subnet (SS), Initialize Subnet (IS), and Deep Reconstruction Subnet (DRS). The network’s architectural framework is illustrated in Figure 2, and the complete MMU-Net sampling and reconstruction process is detailed in Algorithm 1. The roles of these three sub-networks are as follows:

- **Sampling Subnet (SS):** The SS emulates the linear sampling of the original image using convolutional layers. It transforms the input image to simulate the measurements obtained from a low-resolution sensor.
- **Initialize Subnet (IS):** The IS operates on the measurements generated by SS. It enhances the dimension of these measurements to match the size of the original image and performs an initial reconstruction of the image.
- **Deep Reconstruction Subnet (DRS):** The DRS unfolds the ISTA and progressively enhances the quality of image reconstruction over multiple stages. It refines the reconstruction in a stepwise manner, gradually approaching a higher fidelity output.

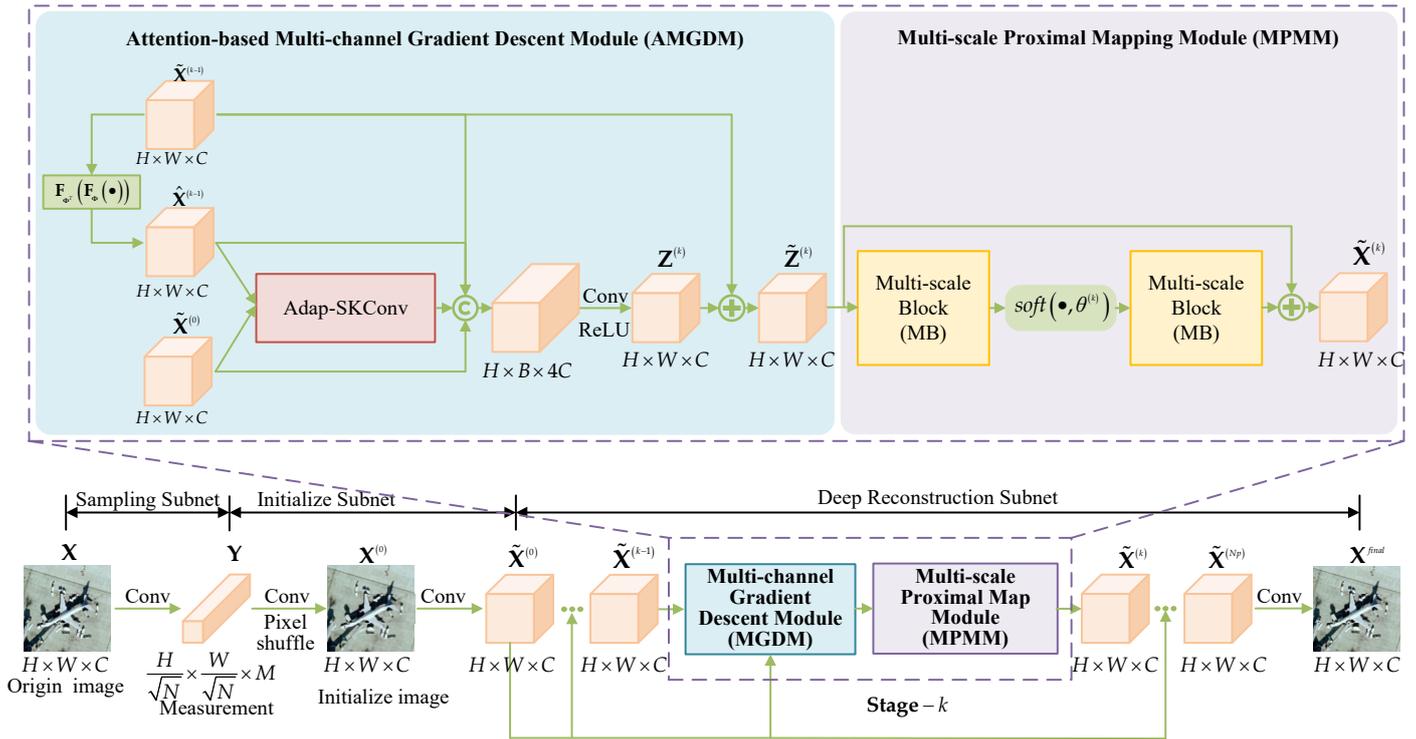


Figure 2. Network framework of the proposed MMU-Net.

4.1. Sampling Subnet (SS)

In the Sampling Subnet, our approach assumes that the original image is represented as $\mathbf{X} \in R^{H \times W}$. To process the image efficiently, it is divided into L blocks of size $\sqrt{N} \times \sqrt{N}$, where $\sqrt{N} \times \sqrt{N} \times L = H \times W$. This paper employs a layer of convolutional operations without biases, represented as $F_{\Phi}(\bullet)$. Notably, we replace the traditional matrix sampling process with this convolutional layer. The sampling matrix Φ is treated as a learnable network parameter and reshaped into M convolutional kernels, each of size $\sqrt{N} \times \sqrt{N}$, with a step size of N . This process yields measurements \mathbf{Y} with dimensions $\frac{H}{\sqrt{N}} \times \frac{W}{\sqrt{N}} \times M$, and it is mathematically expressed as:

$$\mathbf{Y} = F_{\Phi}(\mathbf{X}) \tag{7}$$

4.2. Initialize Subnet (IS)

In the Initialize Subnet, the paper focuses on the initial reconstruction of the measurements \mathbf{Y} into an image denoted as $\mathbf{X}^{(0)}$. This process is facilitated by an unbiased convolutional layer $F_{\Phi^T}(\bullet)$ and a Pixel Shuffle layer. The convolutional layer $F_{\Phi^T}(\bullet)$ operates with a step size of 1 and employs N convolutional kernels of size $1 \times 1 \times N$, derived from the reshape of Φ^T . In IS, the measurements \mathbf{Y} first pass through $F_{\Phi^T}(\bullet)$ to produce a feature map with dimensions $\frac{H}{\sqrt{N}} \times \frac{W}{\sqrt{N}} \times N$. Subsequently, the Pixel Shuffle

layer reorganizes this feature map to generate the initial reconstruction image $\mathbf{X}^{(0)}$ with dimensions $H \times W \times 1$, as represented by the following equation:

$$\mathbf{X}^{(0)} = \text{PixelShuffle}(\mathbf{F}_{\Phi^T}(\mathbf{Y})) \tag{8}$$

Algorithm 1: Algorithm for constructing MMU-Net

Input: Origin image \mathbf{X}

Output: reconstruction image \mathbf{X}^{final}

- 1 The original image \mathbf{X} is sampled using the sampling convolutional layer $\mathbf{F}_{\Phi}(\bullet)$ to obtain the measurement \mathbf{Y}

$$\mathbf{Y} = \mathbf{F}_{\Phi}(\mathbf{X})$$

/* Sampling Subnet */

- 2 The initial reconstruction of the measurements \mathbf{Y} is performed using the initialized convolutional layer \mathbf{F} . The resulting image block is then pixelshuffled to obtain the initial reconstructed image \mathbf{x}

$$\mathbf{X}^{(0)} = \text{PixelShuffle}(\mathbf{F}_{\Phi^T}(\mathbf{Y}))$$

/* Initialize Subnet */

- 3 Conversion of a single-channel into a multi-channel feature map of size

$$\tilde{\mathbf{X}}^{(0)} = \text{Conv}_{(3 \times 3)}(\mathbf{X}^{(0)})$$

/* Deep Reconstruction Subnet */

- 4 **for** $k = 1 : Np$ **do**

- 5 Generate instant reconstruction results $\tilde{\mathbf{Z}}^{(k)}$ based on $\tilde{\mathbf{X}}^{(k-1)}$ via AMGDM

$$\begin{cases} \hat{\mathbf{X}}^{(k-1)} = \mathbf{F}_{\Phi^T}(\mathbf{F}_{\Phi}(\tilde{\mathbf{X}}^{(k-1)})) \\ \mathbf{Z}^{(k)} = \text{ReLU}(\text{Conv}(\text{Concat}(\text{Adap-SKCov}(\tilde{\mathbf{X}}^{(0)}, \hat{\mathbf{X}}^{(k-1)}), \tilde{\mathbf{X}}^{(k-1)}))) \\ \tilde{\mathbf{Z}}^{(k)} = \mathbf{Z}^{(k)} + \tilde{\mathbf{X}}^{(k-1)} \end{cases}$$

- 6 The reconstructed feature map $\tilde{\mathbf{X}}^{(k)}$ for the k th stage is generated by MPMM based on $\tilde{\mathbf{Z}}^{(k)}$

$$\tilde{\mathbf{X}}^{(k)} = \text{MB}(\text{soft}(\text{MB}(\tilde{\mathbf{Z}}^{(k)}), \theta^k))$$

- 7 **end**

- 8 Generate the final reconstructed image \mathbf{X}^{final} based on the reconstructed feature map $\tilde{\mathbf{X}}^{(Np)}$ of the Np th stage

$$\mathbf{X}^{final} = \text{Conv}_{(3 \times 3)}(\tilde{\mathbf{X}}^{(Np)})$$

Return: \mathbf{X}^{final}

4.3. Deep Reconstruction Subnet (DRS)

The Deep Reconstruction Subnet in this paper employs Np stages by unfolding the ISTA. The DRS takes the initial image $\mathbf{X}^{(0)}$ of size $H \times W \times 1$ as its input. Initially, a 3×3 convolutional layer is used to transform the single-channel $\mathbf{X}^{(0)}$ into a multi-channel feature map $\tilde{\mathbf{X}}^{(0)}$ with dimensions $H \times W \times C$. Subsequently, based on the iterative updating steps of ISTA, the network is organized into Np stages, and each stage comprises two modules, namely, AMGDM and MPMM, corresponding to Equations (3) and (4). Finally, the multi-channel feature map $\tilde{\mathbf{X}}^{(Np)}$ from the final stage is reduced to a single-channel image using a 3×3 convolutional layer, resulting in the final reconstructed image \mathbf{X}^{final} .

To address the challenge of limited feature map characterization caused by the single-channel approach within DRS, a multi-channel strategy is incorporated into the AMGDM module. To ensure the rational allocation of weights among different channels, an Adap-SKConv approach with an attention mechanism is introduced to enhance the feature characterization of gradient terms in AMGDM. Additionally, to overcome the limitations of a single-scaled neural network with a restricted receptive field, the MPMM module employs multiple parallel convolutional branching structures (MB) to extract features across various receptive fields. This enables the capture of multi-scale features and enhances the network’s characterization capabilities.

4.3.1. Attention-Based Multi-Channel Gradient Descent Module (AMGDM)

The structure of the AMGDM is designed based on Equation (3) in the iterative step of ISTA, and its position in the network framework is shown in Figure 2. AMGDM makes use of multi-channel versions $\tilde{\mathbf{X}}^{(k-1)}$, $\hat{\mathbf{X}}^{(k-1)}$, and $\tilde{\mathbf{X}}^{(0)}$ of $\mathbf{X}^{(k-1)}$, $\Phi^\top(\Phi\mathbf{X}^{(k-1)})$, and $\Phi^\top\mathbf{Y}$ in Equation (3) to generate an instant reconstruction result $\mathbf{Z}^{(k)}$. Notably, $\hat{\mathbf{X}}^{(k-1)}$ is derived by applying $\mathbf{F}_{\Phi^\top}(\mathbf{F}_\Phi(\bullet))$ channel-by-channel to $\tilde{\mathbf{X}}^{(k-1)}$. The network framework is visually represented in Figure 2.

Specifically, the two gradient terms, $\hat{\mathbf{X}}^{(k-1)}$ and $\tilde{\mathbf{X}}^{(0)}$, are initially processed by the Adap-SKConv module to obtain a fused gradient feature map. Subsequently, this feature map is combined with $\tilde{\mathbf{X}}^{(k-1)}$, $\hat{\mathbf{X}}^{(k-1)}$, and $\tilde{\mathbf{X}}^{(0)}$ to produce a feature map with dimensions $H \times W \times 4C$. This feature map is then downscaled using a 3×3 convolutional layer followed by a ReLU activation function to yield an initial instant reconstruction result $\tilde{\mathbf{Z}}^{(k)}$ of size $H \times W \times C$. Finally, $\tilde{\mathbf{X}}^{(k-1)}$ is added to this result to obtain $\tilde{\mathbf{Z}}^{(k)}$. The AMGDM operation can be represented as shown in Equation (9):

$$\begin{cases} \hat{\mathbf{X}}^{(k-1)} = \mathbf{F}_{\Phi^\top}(\mathbf{F}_\Phi(\tilde{\mathbf{X}}^{(k-1)})) \\ \mathbf{Z}^{(k)} = \text{ReLU}(\text{Conv}(\text{Concat}(\text{Adap-SKConv}(\tilde{\mathbf{X}}^{(0)}, \hat{\mathbf{X}}^{(k-1)}), \tilde{\mathbf{X}}^{(k-1)}))) \\ \tilde{\mathbf{Z}}^{(k)} = \mathbf{Z}^{(k)} + \tilde{\mathbf{X}}^{(k-1)} \end{cases} \quad (9)$$

In AMGDM, drawing inspiration from SKConv with multiple branches in SK-Net [42], Adap-SKConv incorporates an attention mechanism to fuse two feature inputs. The two gradient terms, $\hat{\mathbf{X}}^{(k-1)}$ and $\tilde{\mathbf{X}}^{(0)}$, are processed by Adap-SKConv to enhance the interaction between their information. This fusion enhances the feature characterization of gradient terms. The network structure of Adap-SKConv is visually depicted in Figure 3. Adap-SKConv accepts two inputs, \mathbf{X}_1 and \mathbf{X}_2 . Initially, these inputs are fused, and global average pooling is performed to obtain global information on each channel, represented as the operation \mathbf{F}_{gp} . This operation yields a vector \mathbf{s} for each channel. Subsequently, a two-layered fully connected layer \mathbf{F}_{fc} is employed to obtain compact feature vectors \mathbf{z} . Afterward, \mathbf{z} undergoes softmax and segmentation to derive attentional weights \mathbf{a} and \mathbf{b} , corresponding to \mathbf{X}_1 and \mathbf{X}_2 , respectively. Finally, \mathbf{X}_1 and \mathbf{X}_2 are multiplied and summed with \mathbf{a} and \mathbf{b} , respectively, to yield fused features \mathbf{X}_{out} .

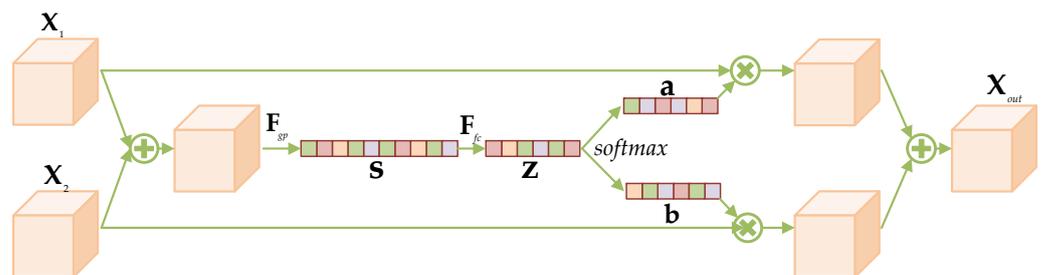


Figure 3. The network structure of Adap-SKConv.

4.3.2. Multi-Scale Proximal Mapping Module (MPMM)

The Multi-scale Proximal Mapping Module corresponds to Equation (6) and is responsible for solving proximal mapping through a soft threshold function and a nonlinear transformation. Its structure is depicted in Figure 2, and the operation can be expressed as shown in Equation (10):

$$\tilde{\mathbf{X}}^{(k)} = MB\left(\text{soft}\left(MB\left(\tilde{\mathbf{Z}}^{(k)}\right), \theta^{(k)}\right)\right) \tag{10}$$

In this paper, the Multi-scale Block is employed to perform nonlinear transformations $\mathbf{F}(\bullet)$ and $\tilde{\mathbf{F}}(\bullet)$. MB leverages multiple parallel convolutional branching structures, inspired by Inception [43], to extract multi-scale features and enhance the characterization capabilities of the network. Notably, unlike classical ISTA-based Deep Unrolling Networks, the inputs and outputs of the Proximal Mapping Module in this paper are multi-channel feature maps rather than single-channel feature maps. Therefore, there is no need for a pre- $\mathbf{F}(\bullet)$ dimensional increase operation or a post- $\tilde{\mathbf{F}}(\bullet)$ dimensional reduction operation in MPMM to avoid information loss.

The Multi-scale Block in MPMM adopts a parallel convolutional multi-branching structure inspired by Inception [43] to extract multi-scale features and enhance the network’s characterization abilities. The network structure of MB is visually presented in Figure 4, and the operation can be expressed as shown in Equation (11):

$$\mathbf{X}^{out} = \text{Conv}_{(3 \times 3)}\left(\text{Concat}\left(\mathbf{X}^{b_1}, \mathbf{X}^{b_2}, \mathbf{X}^{b_3}, \mathbf{X}^{b_4}\right)\right)$$

$$\text{with : } \begin{cases} \mathbf{X}^{b_1} = \text{AvgPool}\left(\text{Conv}_{(1 \times 1)}\left(\mathbf{X}^{in}\right)\right) \\ \mathbf{X}^{b_2} = \text{Conv}_{(1 \times 1)}\left(\mathbf{X}^{in}\right) \\ \mathbf{X}^{b_3} = \text{Conv}_{(1 \times 1)}\left(\left(\text{Conv}_{(3 \times 3)}\left(\mathbf{X}^{in}\right)\right)\right) \\ \mathbf{X}^{b_4} = \text{Conv}_{(1 \times 1)}\left(\text{Conv}_{(3 \times 3)}\left(\left(\text{Conv}_{(3 \times 3)}\left(\mathbf{X}^{in}\right)\right)\right)\right) \end{cases} \tag{11}$$

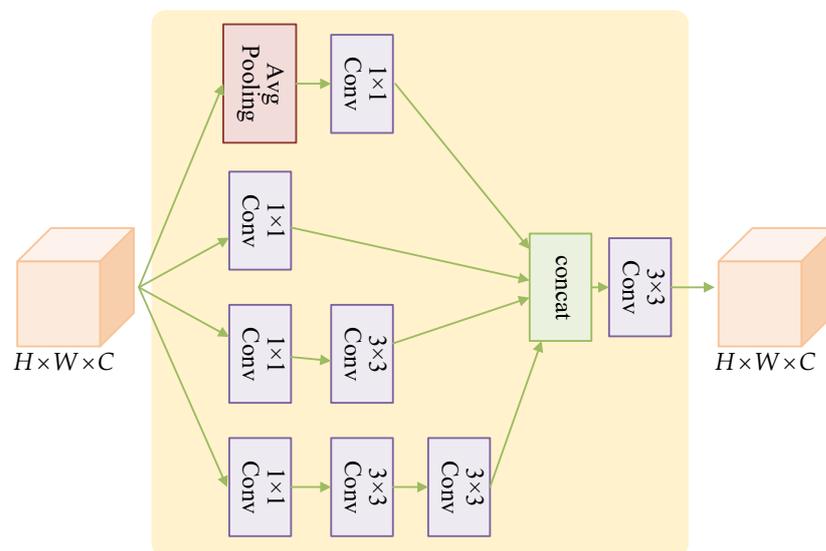


Figure 4. The network structure of Multi-scale Block.

The MB module is designed with four convolutional branches operating at different scales. The first branch includes a global average pooling layer and a convolutional layer with a kernel size of 1×1 and a ReLU activation function. The second branch consists of a convolutional layer with a kernel size of 1×1 . The third branch comprises a convolutional layer with a kernel size of 1×1 and a convolutional layer with a kernel size of 3×3 . The fourth branch consists of one convolution layer with a kernel size of 1×1 and two convolution layers with a kernel size of 3×3 . The use of two 3×3 convolution kernels

5.1. Experimental Parameter Settings

In our experiments, we employ a dataset comprising 91 images, consistent with previous work [30]. These images are utilized for training, with the luminance components of 88,912 randomly extracted image blocks, each of size 33×33 , forming the training set. Our testing set encompasses three natural image datasets and a remote sensing image dataset. The nature image dataset consists of three widely recognized benchmark nature image datasets: Set11 [18], BSD100 [48], and Urban100 [49], and the remote sensing image dataset consists of eight images from the UC Merced Land Use Dataset [50].

For MMU-Net's configuration, we set $Np = 13$, use a batch size of 32, establish a learning rate of 1×10^{-4} , and run the training process for 300 epochs. During training, the network is optimized using an Adam optimizer [51] with a momentum of 0.9 and a weight decay of 0.999.

Our experiments are conducted using the Pytorch 1.11, and the hardware setup comprises an Intel Core i7-12700F processor and an RTX 3070 GPU. To evaluate the reconstruction quality, we utilize the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [52], focusing on the luminance components. In the results tables, the highest-performing method is indicated in bold, and the second-best is underlined.

5.2. Evaluation Metrics

5.2.1. Peak Signal to Noise Ratio (PSNR)

PSNR is a widely-used metric for evaluating image quality at the pixel level. It measures the quality of a reconstructed image in decibels (dB), with higher values indicating superior image quality. For images \mathbf{X} and \mathbf{Y} , both of size $m \times n$, the PSNR is computed as shown in Equation (13):

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_X^2}{\text{MSE}} \right) \quad (13)$$

Here, MAX_X^2 is the maximum possible pixel value of image \mathbf{X} , and MSE denotes the mean square error between images \mathbf{X} and \mathbf{Y} .

5.2.2. Structural Similarity Index Measure (SSIM)

SSIM is a metric that assesses image quality by quantifying structural similarity between two images. It provides insights into brightness, contrast, and structure, with SSIM values ranging from 0 to 1, where larger values indicate greater similarity between images. The SSIM between images \mathbf{X} and \mathbf{Y} is calculated according to Equation (14):

$$\text{SSIM}(\mathbf{X}, \mathbf{Y}) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (14)$$

Here, μ_X and μ_Y represent the mean values of images \mathbf{X} and \mathbf{Y} , while σ_X^2 and σ_Y^2 represent their variances. The covariance between \mathbf{X} and \mathbf{Y} is denoted as σ_{XY} . Additionally, c_1 and c_2 are constant terms.

5.3. Baselines

To gauge the effectiveness of MMU-Net, we conducted comparative evaluations by contrasting it with five well-established baseline methods. In this section, we provide an overview of these baseline techniques and their specific characteristics:

AdapReconNet [18]: AdapReconNet adopts a matrix sampling approach for chunked image sampling. It utilizes a fully connected layer for initial image reconstruction, while employing a variant of the ReconNet for deep reconstruction. Notably, the sampling matrix remains unaltered during the training phase, and the initial reconstruction subnetwork and deep reconstruction subnetwork are jointly trained.

CSNet+ [45]: CSNet+ employs a convolutional neural network to accomplish chunked uniform sampling and chunked initial image reconstruction. Furthermore, it integrates a deep reconstruction sub-network. During the training phase, the sampling sub-network,

initial reconstruction sub-network, and deep reconstruction sub-network are collectively trained.

ISTA-Net+ [28]: ISTA-Net+ utilizes a fixed random Gaussian matrix for chunked image sampling and initial reconstruction. Deep image reconstruction is performed using an ISTA-based deep unfolding network. Similar to AdapReconNet, ISTA-Net+ maintains the sampling matrix constant throughout training and jointly trains the initial reconstruction and deep reconstruction sub-networks.

OPINE-Net+ [30]: OPINE-Net+ integrates a CNN for chunked uniform sampling and chunked initial image reconstruction. It employs an ISTA-based deep unfolding network for the final image reconstruction. OPINE-Net+ extends the architecture of ISTA-Net+ by jointly training the look-alike network, the initial reconstruction sub-network, and the deep reconstruction sub-network.

AMP-Net [29]: AMP-Net initiates image reconstruction with a sampling matrix, initially set as a random Gaussian matrix. It performs chunked image sampling and initial reconstruction using this matrix. For the deep reconstruction phase, AMP-Net follows a denoising perspective, where a deep unfolding network is constructed based on the Approximate Message Passing algorithm. The sampling network, initial reconstruction sub-network, and deep reconstruction sub-network are collectively trained during the training phase.

5.4. Comparison with State-of-the-Art Methods (RQ1)

5.4.1. Comparison in Natural Images

In this section, we compare MMU-Net with five state-of-the-art deep-learning-based CS reconstruction methods using four CS ratios: 0.04, 0.1, 0.25, and 0.3, under natural image datasets. The compared methods include AdapReconNet, CSNet+, ISTA-Net+, AMP-Net, and OPINE-Net+. AdapReconNet and CSNet+ belong to DNUNs, ISTA-Net+ and OPINE-Net+ are ISTA-based DUNs, and AMP-Net is an AMP-based DUN.

Table 2 presents the average PSNR and SSIM results of the five CS reconstruction methods on three datasets: Set11, BSDS68, and Urban100. The table illustrates that, across all four sampling rates, MMU-Net consistently outperforms the existing state-of-the-art CS reconstruction methods on Set11, BSDS68, and Urban100. This result confirms the efficacy of MMU-Net's network structure. Notably, the DUN-based CS reconstruction methods demonstrate significantly better average PSNR and SSIM results compared to DNUN-based methods, suggesting the superiority of the DUN framework in enhancing reconstruction performance.

Table 2. Average PSNR and SSIM of reconstructed images for the six CS reconstruction methods across three datasets: Set11, BSDS68, and Urban100, and four sampling rates: 0.04, 0.1, 0.25, and 0.3. Bold indicates the best reconstruction performance, while underline represents the second-best reconstruction performance.

Dataset	Methods	CS Ratio			
		0.04	0.1	0.25	0.3
Set11	AdapReconNet [18]	23.87/0.7279	27.39/0.8521	31.75/0.9257	33.16/0.9379
	CSNet+ [45]	24.83/0.7480	28.34/0.8580	33.34/0.9387	34.30/0.9490
	ISTA-Net+ [28]	21.32/0.6037	26.64/0.8087	32.59/0.9254	33.74/0.9386
	AMP-Net [29]	24.64/0.7527	28.84/0.8765	34.42/0.9513	36.03/0.9586
	OPINE-Net+ [30]	25.65/0.7911	29.79/0.8905	34.81/0.9503	36.04/0.9600
	ours	25.91/0.8008	30.17/0.8961	35.38/0.9555	36.62/0.9635
BSDS68	AdapReconNet [18]	24.30/0.6491	26.72/0.7821	30.10/0.8901	30.54/0.9044
	CSNet+ [45]	25.43/0.6706	27.91/0.7938	31.12/0.9060	31.66/0.9152
	ISTA-Net+ [28]	22.17/0.5486	25.32/0.7022	29.36/0.8525	30.20/0.8771
	AMP-Net [29]	25.40/0.6534	27.79/0.7853	31.46/0.9053	32.84/0.9240
	OPINE-Net+ [30]	25.20/0.6818	27.72/0.8014	31.56/0.9121	32.50/0.9236
	ours	25.29/0.6915	27.98/0.8097	31.76/0.9102	32.69/0.9259

Table 2. Cont.

Dataset	Methods	CS Ratio			
		0.04	0.1	0.25	0.3
Urban100	AdapReconNet [18]	21.92/0.6390	24.55/0.7801	28.21/0.8841	29.71/0.9043
	CSNet+ [45]	21.96/0.6430	24.76/0.7899	28.13/0.8827	29.90/0.9162
	ISTA-Net+ [28]	19.83/0.5377	24.04/0.7378	29.78/0.8954	30.15/0.9070
	AMP-Net [29]	22.80/0.6814	26.04/0.8283	30.89/0.9202	32.19/0.9365
	OPINE-Net+ [30]	22.97/0.7018	26.51/0.8362	31.36/0.9216	32.58/0.9414
	ours	23.35/0.7189	27.06/0.8474	31.96/0.9335	32.95/0.9442

Figure 5 displays the original images of lena256 and Parrots from the Set11 dataset, along with the images reconstructed by the seven CS reconstruction methods at a sampling rate of 0.1. The zoomed-in details of the reconstructed images are provided. The visual comparison reveals that the images reconstructed by MMU-Net exhibit minimal block artifacts and superior visual quality. A closer examination of the magnified image details of lena256 and Parrots underscores the richness of details and textures in the MMU-Net's reconstructed images. In summary, MMU-Net outperforms the five state-of-the-art CS reconstruction methods in terms of average PSNR and SSIM while delivering superior visual quality.

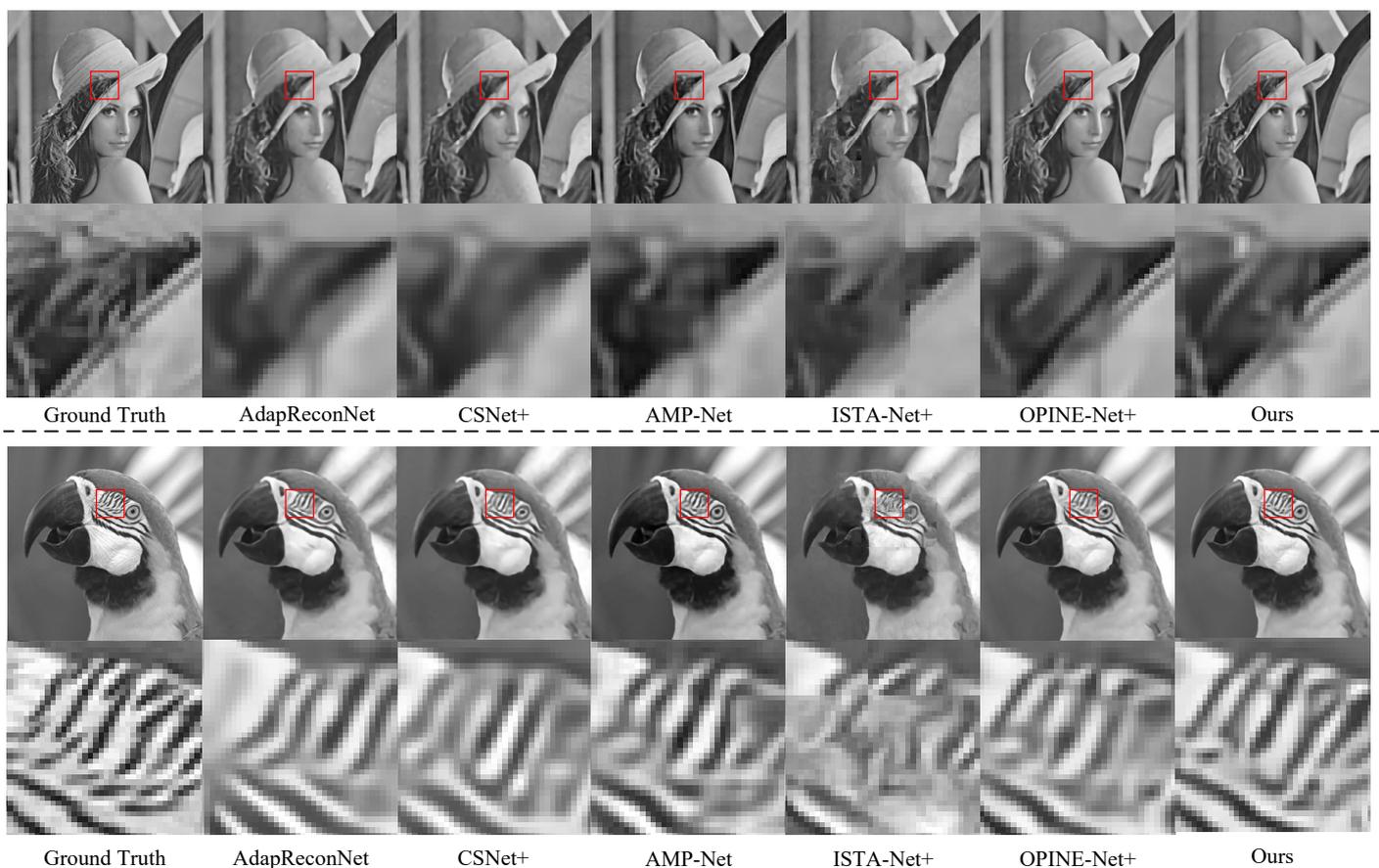


Figure 5. Reconstructed images generated by lena256 and Parrots in Set11 using six reconstruction methods at a sampling rate of 0.1, along with original images. Zoomed-in details are provided below each image.

5.4.2. Comparison in Remote Sensing Images

In this section, we assess the performance of MMU-Net using the UC Merced Land Use Dataset, a remote sensing image dataset. Based on our earlier findings favoring DUNs over DNUNs, we benchmark MMU-Net against three state-of-the-art DUNs: ISTA-Net+, AMP-Net, and OPINE-Net+. We evaluate the reconstruction quality at four different sampling rates: 0.04, 0.1, 0.25, and 0.3, with results visualized in Figure 6 and presented in Table 3.

Table 3. Average PSNR and SSIM of the reconstructed images for the four CS reconstruction methods applied to the remote sensing image dataset at sampling rates of 0.04, 0.1, 0.25, and 0.3. Bold indicates the best reconstruction performance, while underline represents the second-best reconstruction performance.

Image	Methods	CS Ratio			
		0.04	0.1	0.25	0.3
airplane	ISTA-Net+ [28]	22.88/0.6401	28.83/0.8383	33.73/0.9029	35.04/0.9193
	AMP-Net [29]	26.69/0.8146	32.68/0.9258	<u>39.08/0.9786</u>	<u>40.66/0.9852</u>
	OPINE-Net+ [30]	27.18/0.8179	32.74/0.9220	<u>38.87/0.9754</u>	<u>40.43/0.9819</u>
	ours	<u>27.49/0.8289</u>	<u>32.89/0.9263</u>	<u>39.45/0.9776</u>	<u>40.84/0.9836</u>
buildings	ISTA-Net+ [28]	18.49/0.5271	24.13/0.7977	32.25/0.9459	33.89/0.9594
	AMP-Net [29]	23.03/0.7782	28.94/0.9213	36.06/0.9812	37.93/0.9873
	OPINE-Net+ [30]	23.19/0.7689	29.19/0.9206	35.87/0.9783	37.69/0.9846
	ours	<u>23.44/0.7797</u>	<u>29.37/0.9242</u>	<u>36.89/0.9816</u>	<u>38.42/0.9862</u>
dense residential	ISTA-Net+ [28]	19.43/0.5557	24.69/0.7896	31.82/0.9434	33.47/0.9599
	AMP-Net [29]	23.40/0.7487	28.49/0.9132	35.63/0.9800	37.56/0.9867
	OPINE-Net+ [30]	23.88/0.7667	29.11/0.9182	35.80/0.9793	37.62/0.9855
	ours	<u>24.15/0.7793</u>	<u>29.71/0.9276</u>	<u>36.69/0.9822</u>	<u>38.38/0.9874</u>
freeway	ISTA-Net+ [28]	21.29/0.5380	27.05/0.8132	33.21/0.9401	34.49/0.9533
	AMP-Net [29]	24.48/0.7296	29.54/0.9018	36.07/0.9757	37.67/0.9832
	OPINE-Net+ [30]	25.46/0.7640	30.64/0.9148	<u>36.37/0.9742</u>	<u>37.86/0.9814</u>
	ours	<u>25.89/0.7827</u>	<u>30.91/0.9177</u>	<u>36.99/0.9767</u>	<u>38.31/0.9824</u>
intersection	ISTA-Net+ [28]	20.50/0.5483	26.40/0.7763	33.12/0.9211	34.49/0.9381
	AMP-Net [29]	24.82/0.7433	29.84/0.8904	36.67/0.9706	38.42/0.9801
	OPINE-Net+ [30]	25.06/0.7496	30.43/0.8906	<u>36.61/0.9671</u>	<u>38.29/0.9766</u>
	ours	<u>25.20/0.7539</u>	<u>30.58/0.8919</u>	<u>37.42/0.9703</u>	<u>38.94/0.9788</u>
mobile home park	ISTA-Net+ [28]	17.37/0.4904	22.33/0.7338	29.75/0.9269	31.68/0.9480
	AMP-Net [29]	21.10/0.6985	25.94/0.8808	32.40/0.9683	34.11/0.9772
	OPINE-Net+ [30]	21.54/0.7213	26.53/0.8896	<u>32.86/0.9674</u>	<u>34.40/0.9750</u>
	ours	<u>21.94/0.7439</u>	<u>26.81/0.8954</u>	<u>33.66/0.9712</u>	<u>34.99/0.9777</u>
overpass	ISTA-Net+ [28]	22.87/0.5520	27.11/0.7481	34.19/0.9326	36.14/0.9550
	AMP-Net [29]	25.34/0.7140	29.69/0.8590	36.54/0.9695	38.36/0.9789
	OPINE-Net+ [30]	25.57/0.7182	30.69/0.8860	37.82/0.9739	39.09/0.9796
	ours	<u>26.07/0.7182</u>	<u>31.53/0.9069</u>	<u>38.54/0.9772</u>	<u>39.92/0.9826</u>
tennis court	ISTA-Net+ [28]	20.80/0.4438	23.71/0.6024	28.63/0.8407	30.29/0.8806
	AMP-Net [29]	23.68/0.5943	25.98/0.7494	29.22/0.8891	30.32/0.9124
	OPINE-Net+ [30]	23.71/0.5956	26.54/0.7763	30.83/0.9075	31.93/0.9243
	ours	<u>23.84/0.6075</u>	<u>27.19/0.8079</u>	<u>31.61/0.9176</u>	<u>32.43/0.9293</u>

The table showcases the average PSNR and SSIM values of reconstructed images for the four CS reconstruction methods across eight different remote sensing images. The results presented in Table 3 indicate that the PSNR of MMU-Net's reconstructed images surpasses the second-best result by an average of 0.48 dB. Moreover, MMU-Net exhibits significantly better performance compared to the other three state-of-the-art CS reconstruction methods, underscoring the effectiveness of the MMU-Net's network structure.

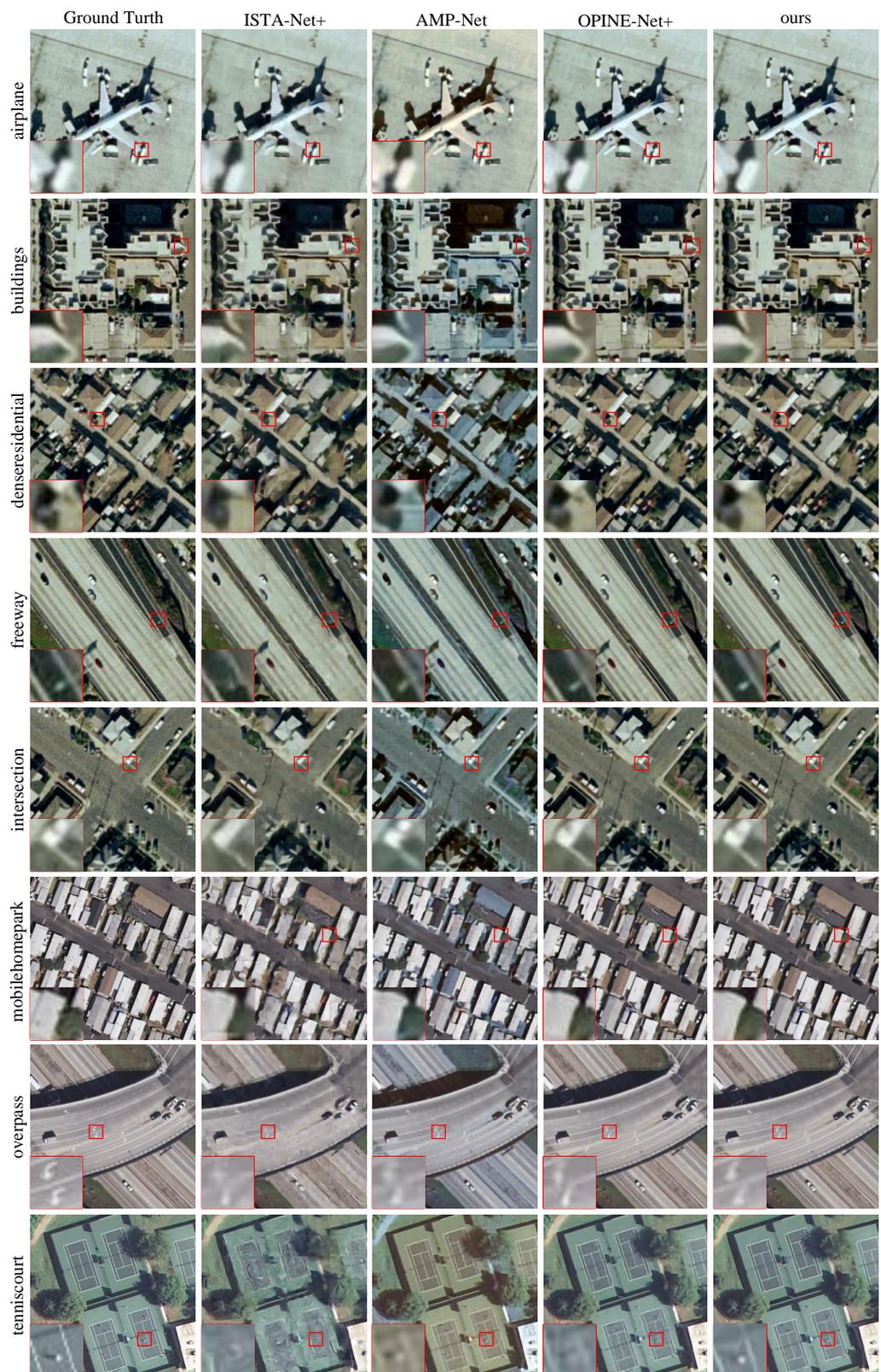


Figure 6. Eight different remote sensing images from the UC Merced Land Use Dataset are compared using the four methods at a sampling rate of 0.1. A zoomed-in view of the details is provided in the lower left corner of each image.

In Figure 6, we visually compare the reconstructed images and their corresponding originals at a sampling rate of 0.1 for various land-use classes. The lower-left corner of

each image provides a magnified view of the selected area in the red box. As depicted in Figure 6, MMU-Net generates reconstructed images with clear contours and rich texture information. Importantly, it maintains the fidelity of small foreground targets even at lower sampling rates, ensuring that the target positions and shapes remain undistorted. In summary, the proposed MMU-Net excels in terms of both the average PSNR, SSIM, and visual quality, making it well-suited for demanding tasks such as target recognition in remote sensing images.

5.5. Study of Computational Time

In the context of CS reconstruction, the model's reconstruction time and the number of parameters are crucial performance metrics. Typically, more complex network structures entail higher time complexity and a higher number of network parameters. In this section, two experiments are designed to validate the network performance of MMU-Net. The first compares the average GPU running time and the number of network parameters of MMU-Net with five other CS reconstruction algorithms. Comparison data are obtained by testing the same dataset in the same environment using the source code provided by the authors. The second explores the average GPU running time of MMU-Net on images of different sizes and the trend of the running time as the image size increases.

Table 4 provides the average GPU running times required by six CS reconstruction methods to reconstruct a 512×512 image at a sampling rate of 0.25. From the table, it is evident that the DNUN models, AdapReconNet and CSNet+, with relatively straightforward network architectures, exhibit shorter average running times in comparison to the DUN methods. In contrast, MMU-Net, the method proposed in this paper, has more expensive computation and preservation costs due to its multi-scale network structure and higher network complexity compared to other DUN methods. However, it still falls within the same order of magnitude as the other methods. Importantly, MMU-Net's reconstruction performance surpasses that of the other methods.

Table 4. Average GPU runtime of six CS reconstruction algorithms for reconstructing 512×512 images at a sampling rate of 0.25.

Methods	AdapReconNet	CSNet+	ISTA-Net+	AMP-Net	OPINE-Net+	Ours
Time	0.0027 s	0.0007 s	0.0143 s	0.1270 s	0.0101 s	0.1950 s
#Para	1.15 M	1.17 M	0.34 M	0.58 M	1.10 M	2.23 M

Figure 7 and Table 5 give the average GPU running time of MMU-Net, reconstructing images of sizes 64×64 , 128×128 , 256×256 , 512×512 and 1024×1024 , respectively. From the right panel of Figure 7, it can be seen that there is a near linear correlation between the average GPU running time of MMU-Net and the image size. When the input image size is large, the average GPU runtime of MMU-Net does not surge.

Table 5. Average GPU runtime required to reconstruct images of 5 different sizes on MMU-Net.

Size	64×64	128×128	256×256	512×512	1024×1024
Time	0.0278 s	0.0350 s	0.0761 s	0.1950 s	0.7250 s

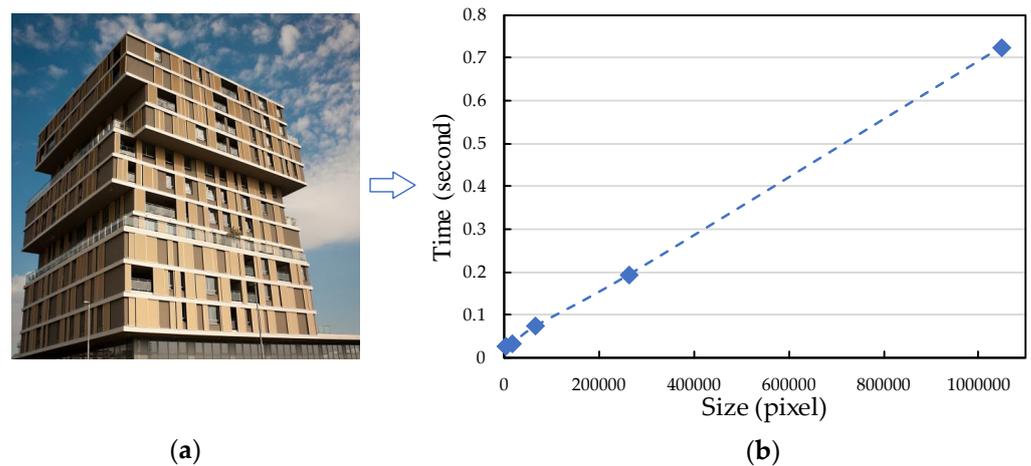


Figure 7. Visualization results of the average GPU runtime required to reconstruct the image on MMU-Net for five different sizes. (a) shows a building image in Urban100 of size 1024×1024 , which is downsampled to obtain a series of images of 512×512 , 256×256 , 128×128 and 64×64 . (b) shows a scatter plot of the average GPU runtime obtained by reconstructing the five image sizes on MMU-Net.

5.6. Ablation Studies and Discussions

In this section, we conduct ablation experiments to validate the effectiveness of the multi-channel strategy, Adap-SKConv, and the multi-scale strategy (MB).

5.6.1. Effectiveness of AMGDM (RQ2)

To assess the effectiveness of the multi-channel strategy and Adap-SKConv within the AMGDM module, we utilize four network modules: GDM-(a), GDM-(b), GDM-(c), and GDM-(d), which replace the gradient descent modules at the locations shown in Figure 1. These modules allow us to compare network performance in different scenarios.

GDM-(a) represents a single-channel module without an attention mechanism, similar to the GDM used in most ISTA-based DUNs. GDM-(b) is a multi-channel module without an attention mechanism. GDM-(c) incorporates a multi-channel module with the CBAM (Convolutional Block Attention Module) attention mechanism, which replaces the Adap-SKConv proposed in this paper. GDM-(d) is a multi-channel module with Adap-SKConv, i.e., the AMGDM proposed in this paper. The network structure of each module is illustrated in Figure 8.

GDM-(b), GDM-(c), and GDM-(d) all adopt multi-channel structures, thereby eliminating the need for subsequent PMMs to perform single-channel and multi-channel transformations, which reduces information loss. GDM-(c) and GDM-(d) utilize different attention mechanisms. Table 6 presents the average PSNR of these three methods on Set11 and the UC Merced Land Use Dataset at three different sampling rates.

From Table 6, we observe that the PSNR of the reconstructed images by GDM-(b) is, on average, 0.19 dB higher than that of GDM-(a) for the three sampling rates. This demonstrates that the multi-channel strategy proposed in this paper enhances the feature map characterization capability by mitigating the information loss resulting from dimensionality reduction, ultimately improving network performance. Additionally, when comparing GDM-(b) and GDM-(d), it is evident that the Adap-SKConv proposed in this paper contributes to an average gain of 0.17 dB in network performance. This confirms that Adap-SKConv effectively enhances the information exchange between gradient terms, thereby improving the quality of reconstruction through a well-designed attention mechanism. Lastly, when comparing GDM-(c) and GDM-(d) between Adap-SKConv proposed in this paper and the state-of-the-art CBAM attention mechanism, we find that the two-input structure of Adap-SKConv outperforms the single-input structure of CBAM in facilitating information exchange between the gradient terms. This enhances feature map characterization and, consequently, improves network reconstruction results.

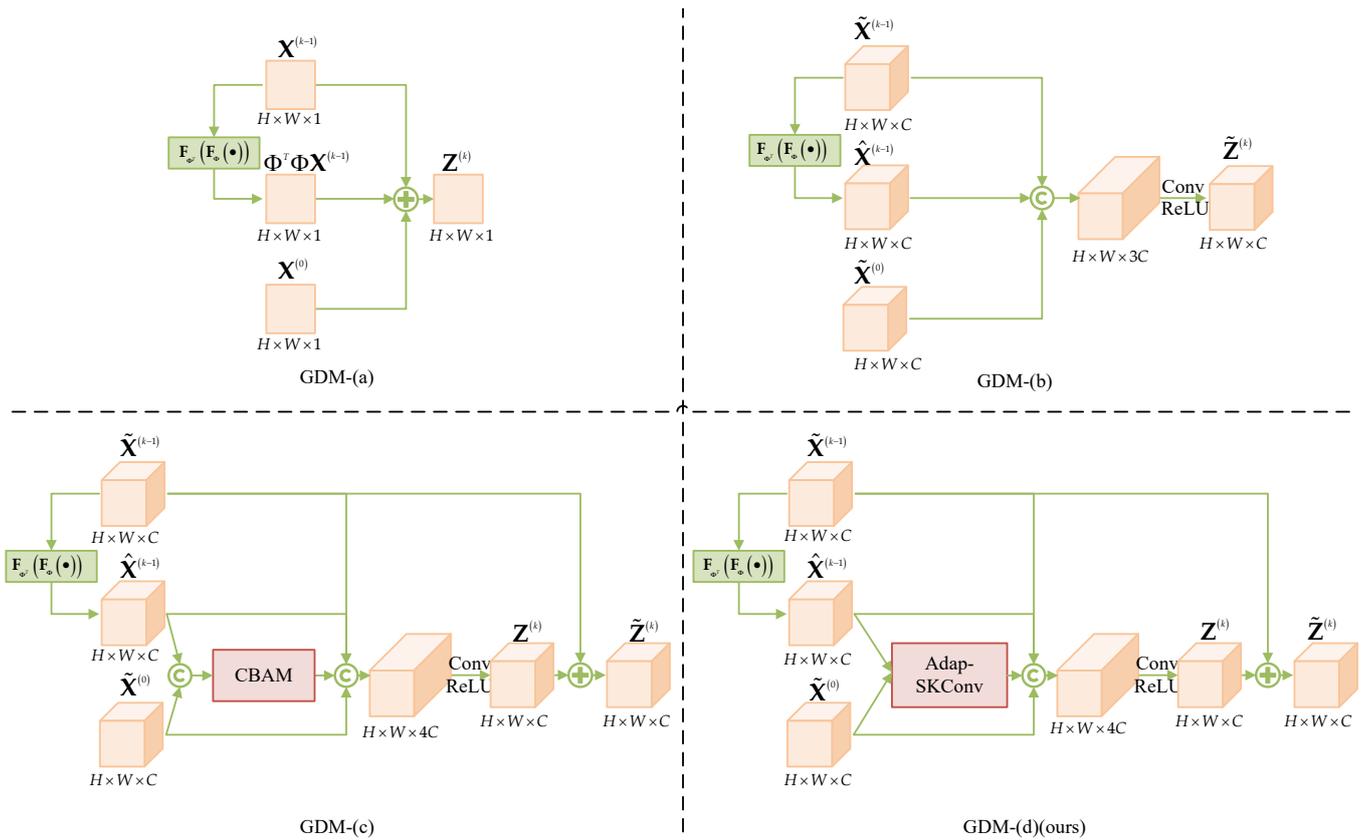


Figure 8. Network framework of GDM-(a), GDM-(b), GDM-(c) and GDM-(d).

Table 6. Average PSNR and SSIM of the reconstructed images of the four CS reconstruction methods under the remote sensing image dataset with four sampling rates of 0.04, 0.1, 0.25, and 0.3. Bold indicates the best reconstruction performance, while underline represents the second-best reconstruction performance.

Dataset	Methods	Multi-Channel	Attention	0.1	0.25	0.3
Set11	GDM-(a)	×	×	29.79	34.81	36.04
	GDM-(b)	✓	×	29.90	35.01	36.17
	GDM-(c)	✓	✓	29.95	35.06	36.26
	GDM-(d)(ours)	✓	✓	30.05	35.16	36.41
UC Merced Land Use Dataset	GDM-(a)	×	×	29.48	35.62	37.16
	GDM-(b)	✓	×	29.60	35.90	37.30
	GDM-(c)	✓	✓	29.67	35.97	37.40
	GDM-(d)(ours)	✓	✓	29.76	36.11	37.56

5.6.2. Effectiveness of MB (RQ3)

In this section, we conduct ablation experiments on the Multi-scale Blocks to assess the effectiveness of the multi-scale strategy, and the experimental results are included in Table 7.

We design and examine single-scale module Block-(1) and multi-scale modules Block-(2), Block-(3), and Block-(4), which comprise two, three, and four branches, respectively. Each of these modules is integrated into the network structure illustrated in Figure 1, replacing sections with $F(\bullet)$ and $\tilde{F}(\bullet)$. Among these modules, Block-(4) represents the MB designed in this paper. The structures of these four Blocks are visualized in Figure 9.

Table 7. Average PSNR of reconstructed images for four network branching structures at three sampling rates (0.1, 0.25, and 0.3) on Set11 and the UC Merced Land Use Dataset, demonstrating the effectiveness of the multi-scale strategy. Bold indicates the best reconstruction performance, while underline represents the second-best reconstruction performance.

Dataset	Methods	0.1	0.25	0.3	0.25	0.3
Set11	Block-(1)	29.79	34.81	36.04	34.81	36.04
	Block-(2)	29.86	34.89	36.14	35.01	36.17
	Block-(3)	29.92	34.98	36.25	35.06	36.26
	Block-(4)	29.98	35.10	36.35	35.16	36.41
UC Merced Land Use Dataset	Block-(1)	29.48	35.62	37.16	35.62	37.16
	Block-(2)	29.56	35.72	37.15	35.90	37.30
	Block-(3)	29.62	35.84	37.27	35.97	37.40
	Block-(4)	29.70	35.86	37.36	36.11	37.56

As shown in Table 7, the average Peak Signal-to-Noise Ratio of the reconstructed images increases with the number of branches. This observation confirms that the multi-scale strategy enhances network performance by increasing the network’s representation capability. However, as the number of branches increases, network complexity also rises, leading to longer training and reconstruction times. To strike a balance between performance and network complexity, this paper selects Block-(4) with four branches as the network structure for the proposed MB.

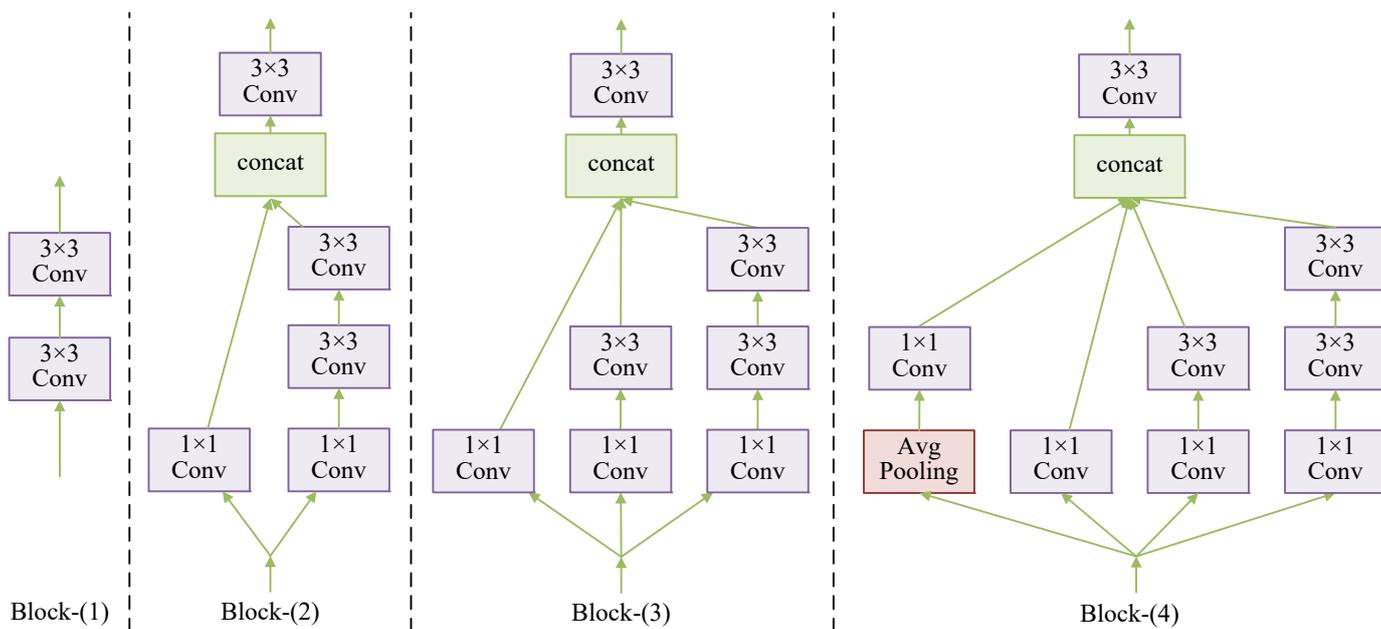


Figure 9. Network structure of Block-(1), Block-(2), Block-(3) and Block-(4).

6. Conclusions

In this paper, we introduced a novel approach for Compressed Sensing image reconstruction. Our proposed MMU-Net leverages innovative strategies to enhance feature map characterization and gradient term representation, ultimately improving reconstruction performance. Specifically, MMU-Net incorporates a multi-channel strategy, bolstering the network’s ability to characterize feature maps effectively. In addition, the introduction of Adap-SKConv within the attention mechanism in Gradient Descent Modules facilitates the exchange of information between gradient terms, leading to improved representation capabilities. Furthermore, we introduced the Multi-scale Block, which enhances network characterization by introducing a multi-scale structure capable of extracting features at different scales. Our extensive experimental results demonstrate the superior performance

of MMU-Net compared to state-of-the-art reconstruction algorithms. We have achieved a harmonious balance between algorithmic complexity and reconstruction quality, especially in the context of CS for natural and remote sensing images. The MMU-Net framework, as proposed in this paper, not only offers an effective solution for CS reconstruction in these domains but also opens up possibilities for enhancing a broad spectrum of applications, including image processing and computer vision. However, the MMU-Net proposed in this paper also has some limitations. First, due to the use of multi-channel and multi-scale strategy to build the network, resulting in more parameters in the model, the model requires further compression. Second, the method proposed in this paper adopts the block sampling strategy to improve sampling efficiency, and cannot realize the global pixel interaction, which limits the overall performance, and the feasibility of whole-map sampling needs to be further studied. For future research, we can direct our efforts toward further enhancing the performance of MMU-Net and exploring its applicability in diverse fields, promising continued advancements in image reconstruction techniques and their broader utility.

Author Contributions: Conceptualization, Z.W. and C.Z.; methodology, Z.W. and C.Z.; software, S.X.; validation, Z.W., X.W. and S.X.; formal analysis, Z.W.; investigation, Z.W. and S.X.; resources, Z.W.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, Z.W.; visualization, Z.W. and S.X.; supervision, Z.W.; project administration, Z.W.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: The research work in this paper was supported by the National Natural Science Foundation of China (No. 62177022, 61901165), Natural Science Foundation of Hubei Province (No. 2022CFA007), Wuhan Knowledge Innovation Project (No.2022020801010258), AI and Faculty Empowerment Pilot Project (No. CCNUAI&FE2022-03-01), Collaborative Innovation Center for Informatization and Balanced Development of K-12 Education by MOE and Hubei Province (No. xtzd2021-005), and the National Natural Science Foundation of China (No. 61501199).

Informed Consent Statement: This study did not involve humans.

Data Availability Statement: Data will be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CS	Compressed Sensing
MMU-Net	Multi-channel and Multi-scale Unfolding
DNUN	Deep Non-unfolding Network
DUN	Deep Unfolding Network
CNN	convolutional neural network
AMP	Approximate Message Passing
ISTA	Iterative Shrinkage Thresholding Algorithm
ADMM	Alternate Direction Multiplier Method
AMGDM	Attention-based Multi-channel Gradient Descent Module
MPMM	Multi-scale Proximal Mapping Module
MB	Multi-scale Block
SS	Sampling Subnet
IS	Initialize Subnet
DRS	Deep Reconstruction Subnet
GDM	Gradient Descent Module
PMM	Proximal Mapping Module
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural similarity index Measure

References

1. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
2. Candes, E.J.; Wakin, M.B. An Introduction to Compressive Sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30. [[CrossRef](#)]

3. Duarte, M.F.; Davenport, M.A.; Takhar, D.; Laska, J.N.; Sun, T.; Kelly, K.F.; Baraniuk, R.G. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 83–91. [[CrossRef](#)]
4. Rousset, F.; Ducros, N.; Farina, A.; Valentini, G.; D'Andrea, C. Adaptive Basis Scan by Wavelet Prediction for Single-Pixel Imaging. *IEEE Trans. Comput. Imaging* **2016**, *3*, 36–46. [[CrossRef](#)]
5. Wu, Z.; Zhang, Z.; Song, J.; Zhang, M. Spatial-temporal synergic prior driven unfolding network for snapshot compressive imaging. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
6. Wt, Z.; Zhang, J.; Mou, C. Dense Deep Unfolding Network with 3D-CNN Prior for Snapshot Compressive Imaging. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 4872–4881.
7. Lustig, M.; Donoho, D.; Pauly, J.M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2007**, *58*, 1182–1195. [[CrossRef](#)] [[PubMed](#)]
8. Yang, Y.; Sun, J.; Li, H.; Xu, Z. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 521–538. [[CrossRef](#)]
9. Chen, C.; Tramel, E.W.; Fowler, J.E. Compressed-sensing recovery of images and video using multihypothesis predictions. In Proceedings of the 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 6–9 November 2011; pp. 1193–1198.
10. Zhang, J.; Zhao, D.; Gao, W. Group-based sparse representation for image restoration. *IEEE Trans. Image Process.* **2014**, *23*, 3336–3351. [[CrossRef](#)]
11. Zhao, C.; Zhang, J.; Wang, R.; Gao, W. CREAM: CNN-REgularized ADMM framework for compressive-sensed image reconstruction. *IEEE Access* **2018**, *6*, 76838–76853. [[CrossRef](#)]
12. Zhao, C.; Ma, S.; Zhang, J.; Xiong, R.; Gao, W. Video Compressive Sensing Reconstruction via Reweighted Residual Sparsity. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 1182–1195. [[CrossRef](#)]
13. Zhao, C.; Zhang, J.; Ma, S.; Fan, X.; Zhang, Y.; Gao, W. Reducing Image Compression Artifacts by Structural Sparse Representation and Quantization Constraint Prior. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2057–2071. [[CrossRef](#)]
14. Zhang, J.; Zhao, C.; Zhao, D.; Gao, W. Image compressive sensing recovery using adaptively learned sparsifying basis via L0 minimization. *Signal Process.* **2014**, *103*, 114–126 [[CrossRef](#)]
15. Elad, M. *Sparse and Redundant Representations—From Theory to Applications in Signal and Image Processing*; Springer: New York, NY, USA, 2010.
16. Nam, S.; Davies, M.E.; Elad, M.; Gribonval, R. The Cosparsity Analysis Model and Algorithms. *Appl. Comput. Harmon. Anal.* **2013**, *34*, 30–56. [[CrossRef](#)]
17. Gilton, D.; Ongie, G.; Willett, R. Neumann networks for linear inverse problems in imaging. *IEEE Trans. Comput. Imaging* **2019**, *6*, 328–343. [[CrossRef](#)]
18. Kulkarni, K.; Lohit, S.; Turaga, P.; Kerviche, R.; Ashok, A. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 449–458.
19. Sun, Y.; Chen, J.; Liu, Q.; Liu, B.; Guo, G. Dual-path attention network for compressed sensing image reconstruction. *IEEE Trans. Image Process.* **2020**, *29*, 9482–9495. [[CrossRef](#)] [[PubMed](#)]
20. Zeng, C.; Ye, J.; Wang, Z.; Zhao, N.; Wu, M. Cascade neural network-based joint sampling and reconstruction for image compressed sensing. *Signal Image Video Process.* **2022**, *16*, 47–54. [[CrossRef](#)]
21. Wang, Z.; Wang, Z.; Zeng, C.; Yu, Y.; Wan, X. High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network. *Circuits Syst. Signal Process.* **2023**, *42*, 1593–1616. [[CrossRef](#)]
22. Kim, Y.; Nadar, M.S.; Bilgin, A. Compressed sensing using a Gaussian Scale Mixtures model in wavelet domain. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 3365–3368. [[CrossRef](#)]
23. He, L.; Carin, L. Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing. *IEEE Trans. Signal Process.* **2009**, *57*, 3488–3497. [[CrossRef](#)]
24. Song, J.; Chen, B.; Zhang, J. Memory-Augmented Deep Unfolding Network for Compressive Sensing. In Proceedings of the 29th ACM International Conference on Multimedia, New York, NY, USA, 15 July 2021; MM '21, p. 4249–4258. [[CrossRef](#)]
25. Mousavi, A.; Patel, A.B.; Baraniuk, R.G. A deep learning approach to structured signal recovery. In Proceedings of the 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 29 September–2 October 2015; pp. 1336–1343.
26. Yao, H.; Dai, F.; Zhang, S.; Zhang, Y.; Tian, Q.; Xu, C. Dr2-net: Deep residual reconstruction network for image compressive sensing. *Neurocomputing* **2019**, *359*, 483–493. [[CrossRef](#)]
27. Su, Y.; Lian, Q. iPiano-Net: Nonconvex optimization inspired multi-scale reconstruction network for compressed sensing. *Signal Process. Image Commun.* **2020**, *89*, 115989. [[CrossRef](#)]
28. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1828–1837.

29. Zhang, Z.; Liu, Y.; Liu, J.; Wen, F.; Zhu, C. AMP-Net: Denoising-based deep unfolding for compressive image sensing. *IEEE Trans. Image Process.* **2020**, *30*, 1487–1500. [[CrossRef](#)]
30. Zhang, J.; Zhao, C.; Gao, W. Optimization-inspired compact deep compressive sensing. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 765–774. [[CrossRef](#)]
31. Xiang, J.; Dong, Y.; Yang, Y. FISTA-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Trans. Med. Imaging* **2021**, *40*, 1329–1339. [[CrossRef](#)] [[PubMed](#)]
32. Chen, W.; Yang, C.; Yang, X. FSOINET: Feature-space optimization-inspired network for image compressive sensing. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 2460–2464.
33. Zhang, J.; Zhang, Z.; Xie, J.; Zhang, Y. High-Throughput Deep Unfolding Network for Compressive Sensing MRI. *IEEE J. Sel. Top. Signal Process.* **2022**, *16*, 750–761. [[CrossRef](#)]
34. Donoho, D.L.; Maleki, A.; Montanari, A. Message passing algorithms for compressed sensing: I. motivation and construction. In Proceedings of the 2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo), Cairo, Egypt, 6–8 January 2010; pp. 1–5.
35. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [[CrossRef](#)]
36. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
37. You, D.; Xie, J.; Zhang, J. ISTA-NET++: Flexible Deep Unfolding Network for Compressive Sensing. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6. [[CrossRef](#)]
38. Zhang, H.; Yang, C. Dual-Domain Update and Double-Group Optimization Network for Image Compressive Sensing. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 1286–1290. [[CrossRef](#)]
39. Amir, Y.; Harel, M.; Malach, R. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J. Comp. Neurol.* **1993**, *334*, 19–46. [[CrossRef](#)] [[PubMed](#)]
40. DeYoe, E.A.; Carman, G.J.; Bandettini, P.; Glickman, S.; Wieser, J.; Cox, R.; Miller, D.; Neitz, J. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2382–2386. [[CrossRef](#)]
41. Barranca, V.J. Neural network learning of improved compressive sensing sampling and receptive field structure. *Neurocomputing* **2021**, *455*, 368–378. [[CrossRef](#)]
42. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 510–519. [[CrossRef](#)]
43. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
44. Shi, W.; Jiang, F.; Zhang, S.; Zhao, D. Deep networks for compressed image sensing. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 877–882.
45. Shi, W.; Jiang, F.; Liu, S.; Zhao, D. Image compressed sensing using convolutional neural network. *IEEE Trans. Image Process.* **2019**, *29*, 375–388. [[CrossRef](#)]
46. Shi, W.; Jiang, F.; Liu, S.; Zhao, D. Scalable convolutional neural network for image compressed sensing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12290–12299.
47. Song, J.; Chen, B.; Zhang, J. Deep Memory-Augmented Proximal Unrolling Network for Compressive Sensing. *Int. J. Comput. Vis. (IJCV)* **2023**, *131*, 1477–1496. [[CrossRef](#)]
48. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Proceedings Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 416–423. [[CrossRef](#)]
49. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5197–5206. [[CrossRef](#)]
50. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the Sigspatial International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 3–5 November 2010; p. 270.
51. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
52. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.