2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

# Lightweight Guidance Sampling and Deep Refinement Reconstruction Network for Adaptive Compressive Sensing

Zhaoxin Cai[1], Yunzhou Zhang[1*], Haoyue Bai[1], Lu Wang[1], Tengda Zhang[1], Sizhan Wang[1], Shibo Zhang[1]

***Abstract*— Adaptive Compressive Sensing (ACS) has attracted increasing attention for its ability to progressively improve image reconstruction quality by dynamically adjusting sampling allocation. Multi-stage sampling is a promising strategy that leverages intermediate reconstructions to guide sampling without relying on image prior information. However, existing multi-stage methods often struggle to capture global structural information, resulting in biased sampling and suboptimal performance. Furthermore, the strong dependency between intermediate reconstruction for sampling guidance and the final reconstruction can hinder targeted optimization. To address these issues, we propose LGDR-Net, a Lightweight Guidance Sampling and Deep Refinement Reconstruction Network. Specifically, the Gradient-Fused Cross-Attention (GFCA) module, embedded within a lightweight guidance network, leverages globally fused information to compensate for incomplete content during multi-stage sampling. Then, sampling resource allocation is driven by inter-stage reconstruction differences, effectively exploiting image sparsity information. Finally, the Deep Refinement Network incorporates a Decoder Dense Feedback Mechanism (DDFM) to reduce cross-layer structural bias and a Multi-Branch Attention Fusion (MBAF) module for improved fine-texture representation. Extensive experiments demonstrate that our proposed LGDR-Net outperforms state-of-the-art methods, achieving an excellent trade-off between computational cost and reconstruction quality.**

## I. Introduction

Compressive Sensing (CS) [1] breaks the NyquistShannon sampling barrier by enabling the recovery of original signals from a limited number of measurements. It emerges as a promising technique for signal acquisition and reconstruction, with wide-ranging applications in Magnetic Resonance Imaging (MRI) [2], [3], single-pixel imaging [4], and wireless communications [5]. According to CS theory, instead of directly acquiring the original signal $\mathbf{x} \in \mathbb{R}^N$, one can obtain a low-dimensional measurement vector $\mathbf{y} \in \mathbb{R}^M$ by applying a designed measurement matrix $\Phi \in \mathbb{R}^{M \times N}$ ($M \ll N$), and undergo the linear mapping $\mathbf{y} = \Phi\mathbf{x}$. $M/N$ is the sampling rate (or CS ratio). Due to the highly underdetermined nature of this problem, directly recovering $\mathbf{x}$ from $\mathbf{y}$ is mathematically ill-posed. Traditional optimization-based methods address this challenge by formulating a regularized minimization problem that models image reconstruction as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\Phi\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R(\mathbf{x}) \tag{1}$$

where the first term is the data fidelity term, and $\lambda R(\mathbf{x})$ is the regularization term. However, traditional methods [6], [7] rely heavily on iterative optimization and handcrafted sparse priors, resulting in high computational cost and poor performance.

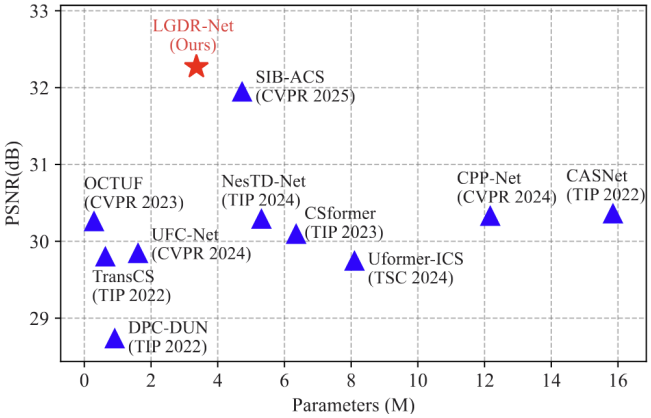



Fig. 1: Comparison of the average PSNR performance of our LGDR-Net with SOTA methods on the BSD68 dataset with sampling rates of 10% and 25%.

Recently, deep unfolding networks (DUNs) have greatly boosted the performance of image reconstruction in terms of speed and quality. [8], [9] integrated iterative optimization algorithms into convolutional networks, achieving performance far beyond that of traditional methods. [10], [11] designed transformer-based networks, which further push the boundaries of reconstruction quality, outperforming prior methods. Furthermore, research attention shifts toward content-aware adaptive sampling for CS, which aims to enhance scene-aware efficiency by prioritizing the allocation of limited sampling resources to complex regions. [12], [13] adopted multi-stage frameworks to allocate sampling resources. In these frameworks, a small sampling rate is used in the initial stage, and the reconstruction process then leverages intermediate measurements [14] or reconstruction-guided images [15] to estimate the sparsity of the original signal. This guides subsequent sampling stages to focus on more informative regions [16], [17]. Through multi-stage sampling allocation, the image reconstruction quality is significantly improved compared with uniform sampling.

Although the multi-stage sampling method has made great progress in performance, it still faces great challenges. In the sampling process, the reconstruction network used for

*The corresponding author of this paper
[1]Zhaoxin Cai, Yunzhou Zhang, Haoyue Bai, Lu Wang, Tengda Zhang, Sizhan Wang, and Shibo Zhang are with College of Information Science and Engineering, Northeastern University, Shenyang 110819, China. zhangyunzhou@mail.neu.edu.cn

This work is funded by National Natural Science Foundation of China(No. 61973066) and Major Science and Technology Projects of Liaoning Province (No. 2021JH1/10400049).

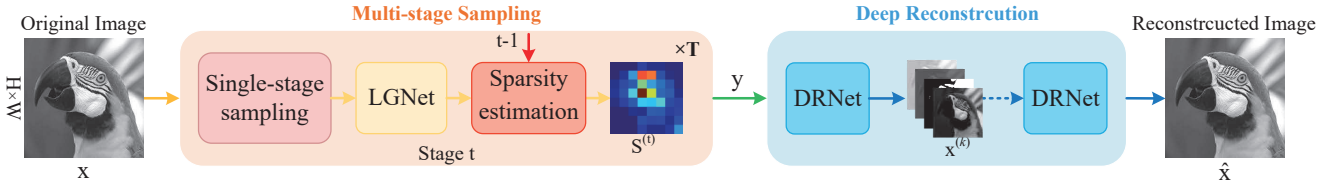



Fig. 2: Overall architecture of our proposed LGDR-Net. Multi-stage adaptive sampling is guided by LGNet with sparsity estimation, followed by progressive reconstruction using DRNet.

sampling guidance relies on convolution stacking, making it difficult to effectively capture global structural information, especially at low sampling rates, leading to sampling bias. Additionally, the tight coupling of sampling guidance and image reconstruction within a single network lacks a targeted and efficient architectural design, making it difficult to simultaneously optimize both the sampling allocation and the final reconstruction accuracy.

To address the above issues, we propose a Lightweight Guidance Sampling and Deep Refinement Reconstruction Network (LGDR-Net), which decouples sampling guidance from the deep refinement process, as shown in Fig.2. In the multi-stage adaptive sampling stage, we introduce a lightweight guidance network (LGNet) that quickly reconstructs the undersampled measurement data. It applies the Gradient-Fused cross-attention module (GFCA) to fuse gradient information with inter-stage features and complete the undersampled regions from a global perspective, generating a more content-coherent intermediate reconstruction. Subsequently, the energy differences between intermediate reconstructions in the discrete cosine domain are used to estimate the sparsity information to allocate sampling resources. In the deep reconstruction, we design a Deep Refinement Network (DRNet) with an unfolding framework to reconstruct the final high-fidelity image. The proposed decoder dense feedback mechanism (DDFM) establishes a reverse information path, effectively making up for the stage-by-stage feature split caused by the unfolded mapping, bridging the high-level semantics and the low-level details, and thus achieving deep global feature fusion. Meanwhile, a multi-branch attention fusion (MBAF) module is integrated to enhance contextual perception of multi-path features, significantly improving detail restoration. As shown in Fig.1, LGDR-Net achieves a better balance between reconstruction performance and computational cost.

The main contributions of this paper are as follows:

- We propose LGDR-Net, which achieves efficient sampling and high-quality reconstruction by decoupling sampling guidance and reconstruction tasks.
- We design LGNet for multi-stage sampling, with the GFCA module integrated to compensate for the content bias of undersampled regions from a global perspective.
- We design DRNet, which combines DDFM and MBAF modules to compensate for the feature gaps between stages caused by the unfolding process.
- Extensive experiments demonstrate that our LGDR-Net outperforms existing state-of-the-art CS methods.

## II. Related Work

### A. Deep Unfolding Network

The DUNs unfold the iterative process of the traditional optimization algorithm [18], [19] into a neural network structure, combining interpretability with the strong expressive power of deep networks, greatly improving the convergence speed and reconstruction quality. [20]–[23] designed a deep unfolding network based on CNN, and achieved performance far exceeding that of traditional methods by using CNN modules as prior denoisers at each stage. Inspired by the success of vision transformer [24], [10], [25], [26] replaced the original CNN module of DUNs with the modified transformer module, which understands the overall structure of the image from a global perspective and achieved better performance. However, these DUNs adopt uniform sampling, ignoring the problem of uneven sparsity distribution in different scenes.

### B. Adaptive Sampling for CS

Content-aware adaptive sampling can guide more accurate reconstruction. [27]–[29] used image prior information to obtain saliency distribution, proving the effectiveness of adaptive sampling allocation, but the ground truth in the real world is unknown. To solve this, [15], [30] employed two-stage sampling, beginning with a low initial sampling rate, uniformly sampling and reconstructing, and then reallocating the sampling according to the reconstruction error. Furthermore, [13], [31] extend the approach to multi-stage sampling, avoiding the risk of overconfidence in a single sampling decision. In addition, [32], [33] directly uses intermediate measurements, but still needs to design a scoring network to predict the sampling allocation. However, the inherent incompleteness in the measurements leads to substantial errors in guidance [34]. Current methods based on multi-stage reconstruction guided sampling only use CNN-based structures, which makes it difficult to capture the global information of the reconstructed image. Moreover, the sampling guidance and image reconstruction are coupled in the same network, without considering the differences in task complexity at different stages. In this paper, we propose LGDR-Net, which decouples multi-stage sampling guidance from final reconstruction.

## III. Proposed Method

### A. Multi-Stage Adaptive Sampling

We propose a multi-stage adaptive sampling framework that dynamically allocates sampling rates in multiple stages. Specifically, given a sampling rate $sr$, we decompose the sampling process into $T$ stages, each of which has a fixed

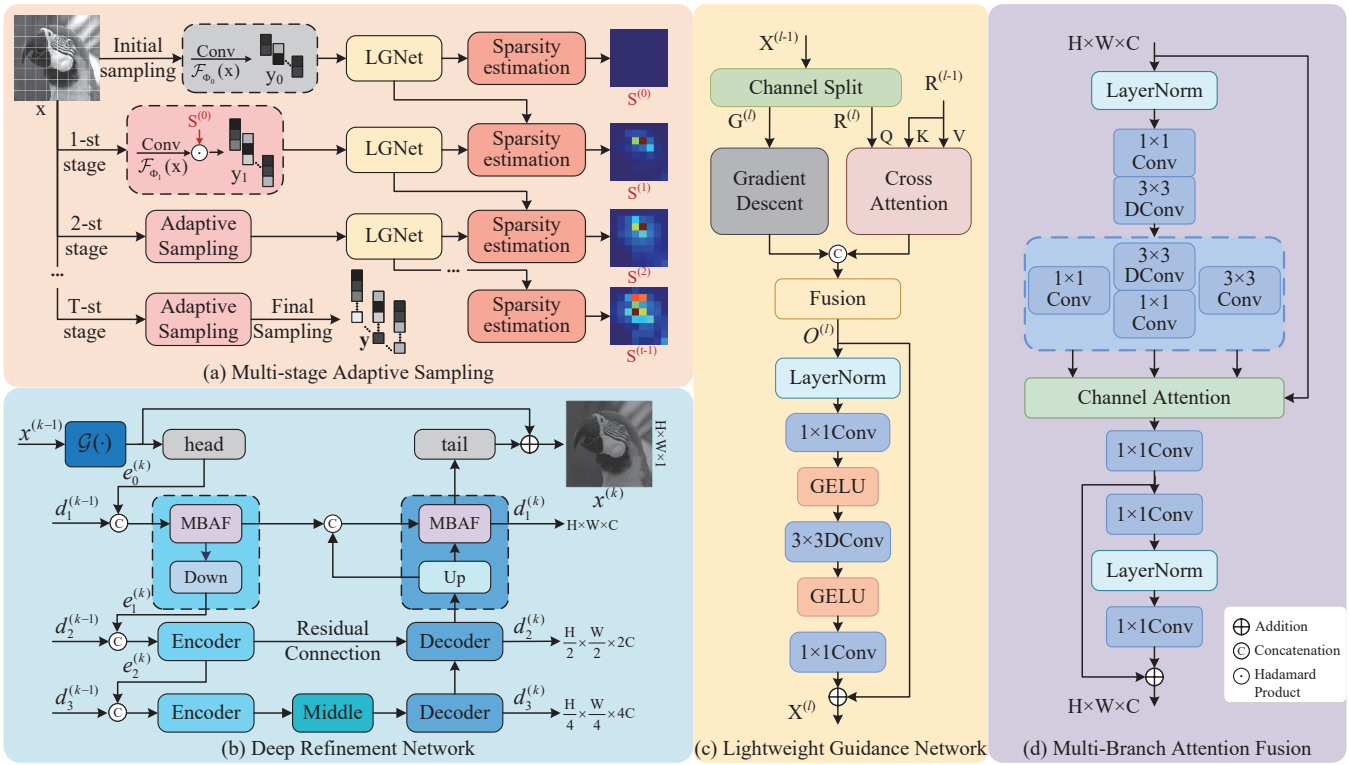



Fig. 3: Details of our proposed LGDR-Net. (a) Multi-stage Adaptive Sampling module; (b) Deep Refinement Network (DRNet) with Decoder Dense Feedback Mechanism (DDFM); (c) Lightweight Guidance Network (LGNet) with Gradient-Fused Cross-Attention (GFCA); (d) Multi-Branch Attention Fusion module (MBAF).

sampling rate of $sr/T$ and consists of an intermediate reconstruction and a corresponding sparsity estimation, as shown in Fig.3(a). During the sampling process, the input image $x \in \mathbb{R}^{H\times W}$ is divided into $N_b$ non-overlapping blocks of size $B \times B$. In the first stage, all blocks are uniformly sampled at a sampling rate of $sr/T$ using a convolution operation $\mathcal{F}_\Phi(\cdot)$, and the degraded image $x_s^{(0)}$ is reconstructed through a lightweight reconstruction network. In each subsequent stage $t \in \{1, 2, \ldots, T\}$, a new sampling budget $\Delta sr = sr/T$ is introduced. The difference in the sparsity distribution of the reconstruction results in the frequency domain between the previous and next stages guides the allocation of $\Delta sr$, i.e., $sr^{(t)} = sr^{(t-1)} + \Delta sr$. For each image block $x_{s,i}^{(t)}$, we assign a dedicated sampling rate $sr_i$:

$$sr_i = \sum_{t=1}^{T} sr_i^{(t)}, \quad \text{s.t.} \frac{1}{N_b}\sum_{i=1}^{N_b} sr_i = sr \tag{2}$$

To enable effective reconstruction during multi-stage adaptive sampling, we design a lightweight guidance network (LGNet) with $L$ layers to provide guidance for subsequent sampling decisions, as shown in Fig.3(c). The entire structure consists of a Gradient-Fused Cross-Attention (GFCA) module and a Feedforward Network. In each layer, the input feature $\mathbf{X_s}^{(l-1)} \in \mathbb{R}^{H\times W\times C}$, extracted from the intermediate reconstruction $x_s^{(l-1)}$, is split into a gradient descent term $\mathbf{G}^{(l-1)} \in \mathbb{R}^{H\times W\times 1}$ (first channel) and a residual feature term $\mathbf{R}^{(l-1)} \in \mathbb{R}^{H\times W\times (C-1)}$. These are fed into the GFCA module, where the former explicitly simulates the gradient descent process ($\mathbf{D}(\cdot)$), and the latter forms a Q-K-V relation with the current stages residual feature $\mathbf{R}^{(l)}$. Then, the Fusion module applies channel attention to enhance the gradient term, capturing channel-wise saliency responses:

$$\begin{aligned}\mathbf{Z}^{(l)} = \text{CrossAttention}(&\text{Norm}(\text{Conv}(\mathbf{R}^{(l)})),\\ &\text{Norm}(\text{Conv}(\mathbf{R}^{(l-1)})), \text{Norm}(\text{Conv}(\mathbf{R}^{(l-1)})))\end{aligned} \tag{3}$$

$$\mathbf{O}^{(l)} = \text{Fusion}(\mathbf{Z}^{(l)}, \mathbf{D}(\mathbf{G}^{(l-1)})) \tag{4}$$

Next, a feedforward network composed of pointwise and depthwise convolutions is applied:

$$\mathbf{X_s}^{(l)} = \text{FFN}(\mathbf{O}^{(l-1)}) \tag{5}$$

Articles [29], [30], [35] have proven that frequency-domain differences are more sensitive than pixel-wise comparisons in capturing structural errors and unstable fine-grained regions. In our adaptive sampling framework, we use LGNet ($\mathcal{F}_{\text{light}}$) to generate the reconstruction at stage $t$, denoted as $x_s^{(t)}$, and compare it with the previous reconstruction $x_s^{(t-1)}$. For the $i$-th image block $x_{s,i}^{(t-1)}$ and $x_{s,i}^{(t)}$, we compute their difference in the Discrete Cosine Transform (DCT) domain to estimate the sparsity information $S_i^{(t)}$:

$$S_i^{(t)} = \frac{\left\| F_i^{(t)} - F_i^{(t-1)} \right\|_2^2}{\sum_{j=1}^{N_b} \left\| F_j^{(t)} - F_j^{(t-1)} \right\|_2^2} \tag{6}$$

where $F_i^{(t)}$ denotes the DCT-transformed block. Then, $S_i^{(t)}$ determines the allocation of the additional sampling rate:

$$sr_i^{(t)} = \min\left(sr_i^{(t-1)} + \Delta sr \cdot S_i^{(t)},\ 1.0\right) \tag{7}$$

To fully utilize the sampling resources, any surplus resulting from clipping $sr_i^{(t)}$ is reallocated to unsaturated blocks, yielding the sparsity map $S^{(t)}$ for the current stage. After $T$ stages, the final measurements $\mathbf{y}$ are acquired based on $S^{(T)}$, effectively concentrating sampling efforts on structurally complex regions.

### B. Deep Refinement Network

In our LGDR-Net, the Deep Refinement Network (DRNet) is designed as the final high-precision reconstruction module. Once the optimized sampling rate map is determined and the final measurements $\mathbf{y}$ are acquired, DRNet performs high-fidelity and high-consistency image reconstruction, as illustrated in Fig.3(b). Following the Deep Unfolding paradigm, we adopt the Proximal Gradient Descent (PGD) [36] algorithm to unfold the optimization in Eq.(1) into $K$ iterative stages. Each iteration updates the reconstruction by combining a data fidelity term with a learned prior. Specifically, a gradient descent step $\mathcal{G}(\cdot)$ ensures consistency with the measurements, followed by a deep prior network $\mathcal{D}_k(\cdot)$ that refines the final result:

$$\mathbf{z}^{(k)} = \mathcal{G}(\mathbf{x}^{(k-1)}) = \mathbf{x}^{(k-1)} - \lambda_k \Phi^\top \left(\Phi \mathbf{x}^{(k-1)} - \mathbf{y}\right) \quad (8)$$

$$\mathbf{x}^{(k)} = \mathcal{D}_k(\mathbf{z}^{(k)}) \quad (9)$$

where $\lambda_k$ is a learnable step size and $\Phi^\top$ denotes the transpose of the measurement matrix.

U-shaped architectures have become a standard in image restoration tasks due to their strong feature extraction and multi-scale fusion capability [37]. However, traditional U-Nets use a top-down encoder and bottom-up decoder design, which suffers from inter-stage disconnection in unfolding networks, limiting the influence of high-level semantics on low-level detail reconstruction. To overcome this, we introduce a Decoder Dense Feedback Mechanism (DDFM) to construct the core structure of DRNet, as shown in Fig.3(b). In addition to standard skip connections, this mechanism introduces reverse information paths from decoder to encoder, enabling deep cross-scale information interaction and refinement. Given the input $\mathbf{z}^{(k)} \in \mathbb{R}^{H\times W\times 1}$, we first embed it into a high-dimensional feature space through a head convolution:

$$\mathbf{e}_0^{(k)} = \mathcal{F}_{\text{head}}(\mathbf{z}^{(k)}) \in \mathbb{R}^{H\times W\times C} \quad (10)$$

In the $p$-th encoder layer ($p = 1, ..., P$), the feature $\mathbf{e}_{p-1}^{(k)}$ from the previous encoder is concatenated with the feedback feature $\mathbf{d}_{p-1}^{(k-1)}$ from the corresponding decoder layer. The fused feature is processed by a $1\times 1$ convolution to reduce channel dimensions:

$$\mathbf{e}_{p-1}^{\prime(k)} = \text{Conv}_{1\times 1}(\text{Concat}(\mathbf{e}_{p-1}^{(k)}, \mathbf{d}_{p-1}^{(k-1)})) \quad (11)$$

To capture multi-scale spatial structures, we design a Multi-Branch Attention Fusion (MBAF) module, as shown in Fig.3(d). It first extracts shallow features using depth-wise separable convolution. Then, three parallel convolution branches with kernel sizes of $1\times 1$, $3\times 3$, and $3\times 3$ (depthwise) are applied to extract features with different receptive fields. Their outputs are concatenated and fused with the initial input:

$$f_{b_i}^{(k)} = \text{Conv}_{b_i}(\mathbf{e}_{p-1}^{\prime(k)}) \quad (12)$$

$$f_{\text{concat}}^{(k)} = \text{Concat}(f_{b1}^{(k)}, f_{b2}^{(k)}, f_{b3}^{(k)}, \mathbf{e}_0^{(k)}) \quad (13)$$

The fused feature is processed by a channel attention module to adaptively reweight the importance of channels:

$$\mathbf{e}_{\text{ca}}^{(k)} = f_{\text{concat}}^{(k)} \odot \mathbf{CA}(f_{\text{concat}}^{(k)}) \quad (14)$$

where $\mathbf{CA}(\cdot)$ denotes channel attention. Then, we compress the channels using a $1\times 1$ convolution, apply layer normalization, and downsampling:

$$\mathbf{e}_p^{(k)} = \text{Down}\left(\text{Conv}(\text{LN}(\text{Conv}(\mathbf{e}_{\text{ca}}^{(k)}))) + \mathbf{e}_{\text{ca}}^{(k)}\right) \quad (15)$$

where $\text{Down}(\cdot)$ is a $2\times 2$ strided convolution, and $\text{Up}(\cdot)$ (used in decoder) denotes the transposed convolution for upsampling.

Finally, the decoder output $\mathbf{d}_p^{(k)}$ is passed through a tail convolution and added with the intermediate $\mathbf{z}^{(k)}$ via residual connection to obtain the final reconstruction:

$$\hat{\mathbf{x}}^{(k)} = \mathcal{F}_{\text{tail}}(\mathbf{d}_p^{(k)}) + \mathbf{z}^{(k)} \quad (16)$$

This multi-path fusion and reverse feedback design enables the encoder at each stage to jointly model both local details and global context, yielding richer and more structured features for high-quality image reconstruction.

### C. Loss Function

We design a hybrid loss function, which is composed of Charbonnier Loss [38] for pixel alignment and Structure Similarity Index Measure (SSIM) Loss for structure preservation. Let the training dataset be $\{x_i\}_{i=1}^{N_t}$, where $x_i$ represents the original image and $N_t$ is the total number of training samples. Let $\hat{x}_i$ be the reconstruction result, and $\theta$ be the trainable parameter of the network. We define the loss function $\mathcal{L}(\theta)$ as follows:

$$\mathcal{L}(\theta) = \frac{1}{N_t}\sum_{i=1}^{N_t}\left[\lambda_1 \cdot \mathcal{L}_C(x_i, \hat{x}_i) + \lambda_2 \cdot \mathcal{L}_S(x_i, \hat{x}_i)\right] \quad (17)$$

where $\lambda_1$ and $\lambda_2$ are hyperparameters used to balance the contributions of the two loss terms. $\mathcal{L}_C$ and $\mathcal{L}_S$ are defined as follows:

$$\mathcal{L}_C(x_i, \hat{x}_i) = \left(\|x_i - \hat{x}_i\|_2^2 + \epsilon\right)^{1/2} \quad (18)$$

$$\mathcal{L}_S(x_i, \hat{x}_i) = 1 - \text{SSIM}(x_i, \hat{x}_i) \quad (19)$$

where $\epsilon$ is a small positive constant that guarantees numerical stability.

TABLE I: The performance comparison with SOTA methods in terms of PSNR (dB) and SSIM on Set11 and BSD68 datasets at different sampling rates (4% - 50%). The top two results at each sampling rate are bolded and underlined, respectively.

| Dataset | Methods | Sampling Rate | | | | | |
|---|---|---|---|---|---|---|---|
| | | 4% | 10% | 25% | 30% | 40% | 50% |
| Set11 | AMP-Net (TIP'21) | 25.26/0.7722 | 29.40/0.8779 | 34.63/0.9481 | 36.03/0.9586 | 38.28/0.9715 | 40.34/0.9804 |
| | TransCS (TIP'22) | 25.41/0.7883 | 29.54/0.8877 | 35.06/0.9548 | 35.62/0.9588 | 38.46/0.9737 | 40.49/0.9815 |
| | CASNet (TIP'22) | 26.41/0.8153 | 30.36/0.9014 | 35.67/0.9591 | 36.92/0.9662 | 39.04/0.9760 | 40.93/0.9826 |
| | OCTUF (CVPR'23) | 26.31/0.8095 | 30.70/0.9030 | 36.10/0.9604 | 37.21/0.9673 | 39.41/0.9773 | 41.34/0.9838 |
| | CSformer (TIP'23) | 26.28/0.8062 | 30.66/0.9027 | 35.46/0.9570 | -/- | -/- | 41.04/0.9831 |
| | DPC-DUN (TIP'23) | 24.38/0.7498 | 29.42/0.8801 | 34.75/0.9483 | 36.02/0.9577 | 38.06/0.9698 | 39.90/0.9791 |
| | Uformer-ICS (TSC'23) | 26.90/0.8173 | 30.07/0.8904 | 34.38/0.9470 | 35.40/0.9550 | 37.24/0.9661 | 39.03/0.9749 |
| | NesTD-Net (TIP'24) | 26.73/0.8238 | 30.91/0.9099 | 36.27/0.9622 | 37.17/0.9677 | 39.47/0.9778 | -/- |
| | UFC-Net (CVPR'24) | 25.92/0.7943 | 30.15/0.8960 | 35.42/0.9567 | -/- | -/- | -/- |
| | CPP-Net (CVPR'24) | 27.23/0.8337 | 31.27/0.9135 | 36.35/0.9631 | 37.55/0.9696 | 39.53/0.9781 | 41.39/0.9842 |
| | SIB-ACS (CVPR'25) | 26.56/0.7857 | 32.30/**0.9178** | 37.73/0.9645 | 38.91/0.9700 | 41.06/0.9785 | 43.10/0.9841 |
| | **LGDR-Net(Ours)** | **28.39/0.8411** | **32.37**/0.9143 | **37.96/0.9650** | **39.08/0.9701** | **41.38/0.9789** | **43.55/0.9845** |
| BSD68 | AMP-Net (TIP'21) | 25.47/0.6534 | 27.89/0.7901 | 31.77/0.9031 | 32.84/0.9240 | 34.86/0.9508 | 36.82/0.9680 |
| | TransCS (TIP'22) | 25.28/0.6881 | 27.88/0.7885 | 31.87/0.9041 | 32.56/0.9276 | 34.82/0.9540 | 36.81/0.9699 |
| | CASNet (TIP'22) | 25.73/0.7079 | 28.41/0.8230 | 32.31/0.9196 | 33.39/0.9358 | 35.44/0.9581 | 37.49/0.9728 |
| | OCTUF (CVPR'23) | 25.57/0.6983 | 28.28/0.8177 | 32.24/0.9185 | 33.32/0.9348 | 35.35/0.9578 | 37.41/0.9729 |
| | CSformer (TIP'23) | 25.91/0.7045 | 28.28/0.8078 | 31.91/0.9102 | -/- | -/- | 37.16/0.9714 |
| | DPC-DUN (TIP'23) | 23.86/0.6293 | 26.77/0.7608 | 30.70/0.8832 | 31.75/0.9050 | 33.70/0.9364 | 35.62/0.9574 |
| | Uformer-ICS (TSC'23) | 25.50/0.6926 | 28.13/0.8027 | 31.74/0.9006 | 32.62/0.9179 | 34.15/0.9425 | 37.18/0.9618 |
| | NesTD-Net (TIP'24) | 25.65/0.7057 | 28.28/0.8231 | 32.30/0.9203 | 33.22/0.9354 | 35.37/0.9586 | 37.45/0.9735 |
| | UFC-Net (CVPR'24) | 25.30/0.6877 | 27.95/0.8086 | 31.74/0.9093 | -/- | -/- | -/- |
| | CPP-Net (CVPR'24) | 25.81/0.7068 | 28.41/0.8227 | 32.25/0.9188 | 33.34/0.9353 | 35.33/0.9575 | 37.30/0.9722 |
| | SIB-ACS (CVPR'25) | 25.84/0.6826 | 29.54/0.8401 | 34.35/**0.9312** | 35.72/0.9455 | 38.38/0.9653 | 41.14/0.9779 |
| | **LGDR-Net(Ours)** | **26.94/0.7407** | **29.77/0.8410** | **34.79**/0.9307 | **36.32/0.9457** | **39.21/0.9664** | **42.16/0.9793** |

TABLE II: PSNR (dB) comparison with the latest SOTA methods at extremely low sampling rates of 1% and 4%.

| Methods | Urban100 | | DIV2K | | Avg. |
|---|---|---|---|---|---|
| | 1% | 4% | 1% | 4% | |
| OCTUF | 19.88 | 23.68 | 23.00 | 26.61 | 23.29 |
| UFC-Net | 19.69 | 19.69 | 22.75 | 26.21 | 22.09 |
| CPP-Net | 20.55 | 24.66 | 23.28 | 26.97 | 23.87 |
| SIB-ACS | -/- | 23.97 | -/- | 26.50 | -/- |
| Ours | **21.97** | **25.67** | **24.75** | **27.97** | **25.09** |

## IV. Experiments

### A. Implementation Details

For training, we follow [26] and use the WED dataset [39] without data augmentation. Randomly cropped patches of size $128 \times 128$ are used as input. The batch size is set to 16, and the number of feature channels $C$ is 32. The block size $B$ is 8. The number of adaptive sampling stages $T$ is 5, with 8 unfolding stages in LGNet and 6 in DRNet. The loss balance hyperparameter $\lambda_1$ and $\lambda_2$ is set to 1 and 0.1.

For performance evaluation, we test on four benchmark datasets: Set11 [40], BSD68 [41], Urban100 [42], and DIV2K [43]. Evaluation metrics include Peak Signal-to-Noise Ratio (PSNR) and SSIM. All experiments are conducted on an NVIDIA RTX 4090 GPU using PyTorch.

### B. Comparison with the State-of-the-Arts

To comprehensively evaluate the effectiveness of our proposed LGDR-Net, we compare it against several state-of-the-art (SOTA) methods, including AMP-Net [22], TransCS [10], CASNet [30], OCTUF [25], CSformer [44], DPC-DUN [45], Uformer-ICS [15], NesTD-Net [46], UFC-Net [47], CPP-Net [48] and SIB-ACS [34].

For quantitative comparison, our results are shown in Tab.I (regular sampling rate) and Tab.II (extremely low sampling rate). As shown in Tab.I, on the BSD68 dataset, LGDR-Net achieves PSNR (dB) / SSIM improvements of 0.23 dB (0.78%) / 0.0009 (0.11%) over the previous SOTA method at a 10% sampling rate. At sampling rates of 30% and 40%, LGDR-Net surpasses the previous SOTA method by 0.60 dB (1.68%) and 0.83 dB (2.16%), respectively. As shown in Tab.II, at extremely low sampling rates of 4% and 1%, LGDR-Net consistently outperforms the previous SOTA method. Specifically, at 4%, it achieves improvements of 1.01 dB (4.10%) on Urban100 and 1.00 dB (3.71%) on DIV2K; at 1%, the gains increase to 1.42 dB (6.91%) on Urban100 and 1.47 dB (6.31%) on DIV2K.

For qualitative comparison, we further visualize results on the BSD68 and Urban100 datasets at 25% and 4% sampling rates, as shown in Fig.6 and Fig.4. The results demonstrate that LGDR-Net significantly improves the reconstruction of complex textures and fine structures. In particular, it preserves clear building edges and natural wall textures, showing notable advantages in both structural fidelity and detail restoration. To highlight the effectiveness of our multi-stage adaptive sampling, Fig. 5 shows three Urban100 images at a 25% sampling rate, showing the original image, the LGDR-Net reconstruction with PSNR/SSIM, and sampling allocation from Stage 35. The sampling resource allocation gradually shifts from uniform coverage to high-frequency, structured regionssuch as edges, windows, enabling finer detail and higher-quality reconstruction. The refined allocation aligning with image edges indicates that the network leverages intermediate feedback, validating the multi-stage designs interpretability and robustness.

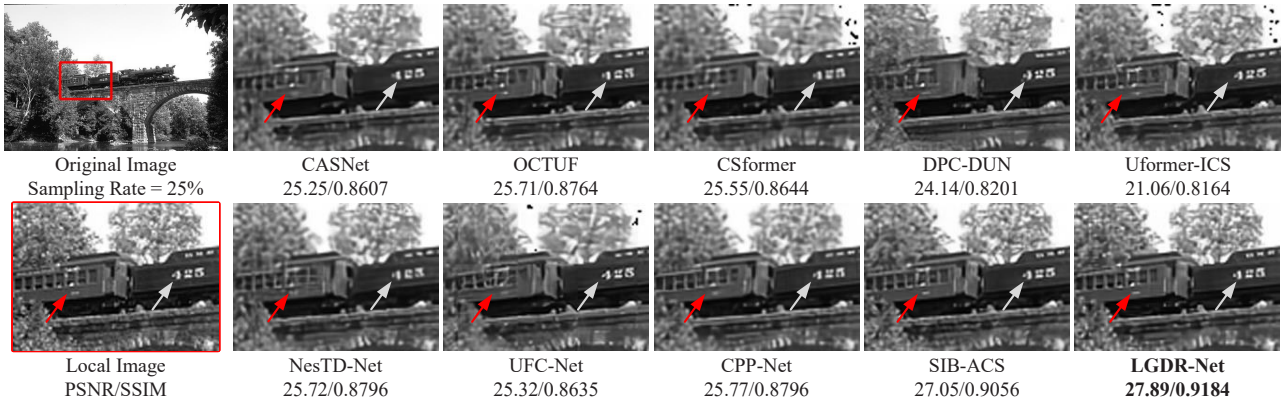



Fig. 4: Visual comparison of reconstruction results on the BSD68 dataset at sampling rate = 25%. Regions of interest are zoomed in and indicated by arrows to highlight noticeable differences in structural details and texture preservation.

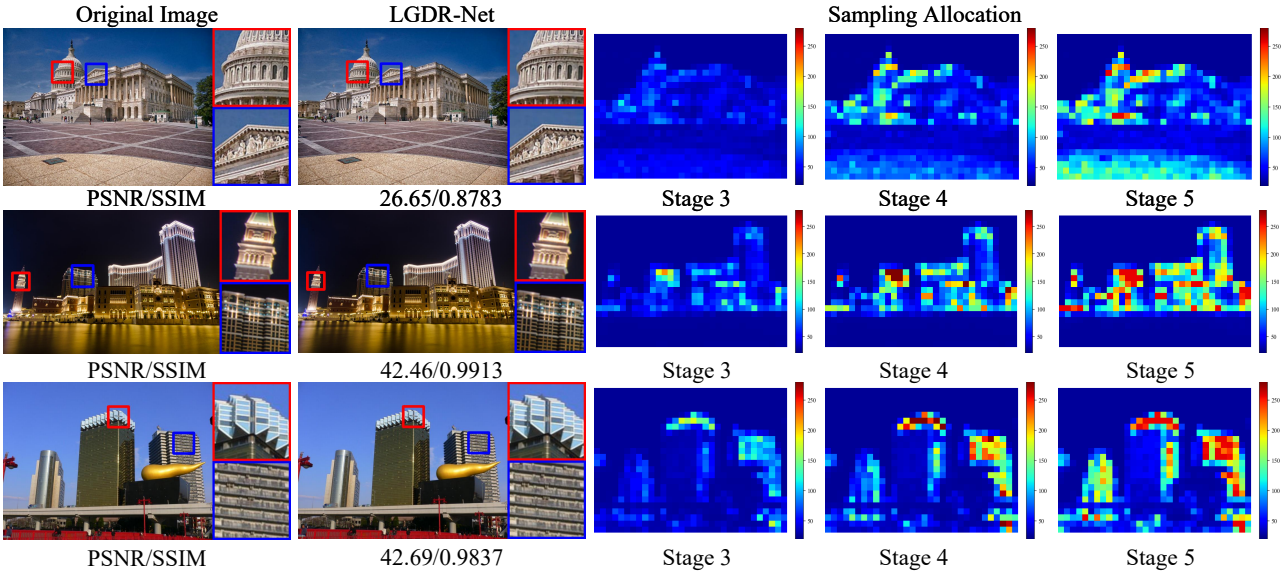



Fig. 5: Visualization of multi-stage adaptive sampling and reconstruction results on the Urban100 dataset under a sampling rate = 25%.


Fig. 6: Comparisons of visual results and PSNR/SSIM performance on the Urban100 dataset at an extremely low sampling rate = 4%.

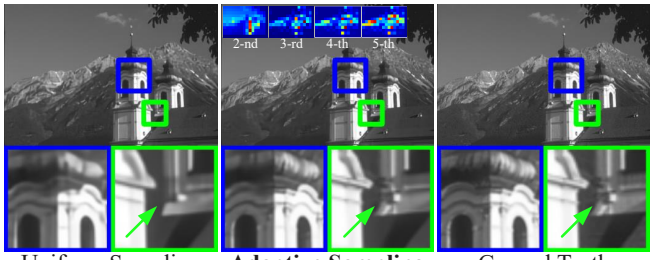
### C. Ablation Studies

*1) Effectiveness of Adaptive Sampling.:* To validate the effectiveness of the proposed adaptive sampling strategy, we conduct ablation studies on different sampling settings, including uniform sampling (using only DRNet) and adaptive sampling with varying numbers of stages (2/4/5/8). As shown in Tab.III, the adaptive strategy significantly improves reconstruction quality over uniform sampling. However, increasing the number of sampling stages beyond 5 yields marginal gains while introducing higher computational complexity and model parameters. Therefore, we adopt a 5-Stage sampling configuration. Furthermore, Fig.7 compares the visual results. LGDR-Net preserves more structural edges and fine textures than uniform sampling, highlighting the benefits of multi-stage adaptive sampling.

TABLE III: Ablation studies of different sampling settings, including uniform sampling and multi-stage sampling.

| Setting | Sampling Rate | |
|---|---|---|
| | 10% | 25% |
| Uniform Sampling | 28.18/0.8288 | 32.04/0.9220 |
| 2-Stage Sampling | 28.62/0.8297 | 32.19/0.9184 |
| 4-Stage Sampling | 29.32/0.8386 | 34.22/0.9293 |
| 5-Stage Sampling | 29.77/0.8410 | 34.79/0.9307 |
| 8-Stage Sampling | 29.81/0.8416 | 34.84/0.9313 |

*2) Effectiveness of Deep Refinement Network.:* To analyze the effectiveness of the proposed DRNet, we compare four network variants, as shown in Tab.IV. The results indicate that adding either the DDFM or MBAF module individually improves performance over the baseline (a). When both modules are combined, the model achieves the best perfor-



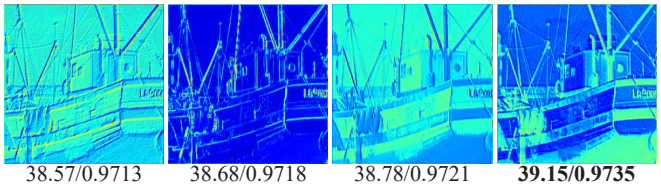

Uniform Sampling 34.89/0.9471 | **Adaptive Sampling 38.23/0.9535** | Ground Truth PSNR/SSIM

Fig. 7: Comparison of uniform sampling and adaptive sampling at sampling rate = 25%. The sampling rate allocation for each stage is shown at the top of the middle column.

mance, demonstrating their complementarity and synergistic effect. Furthermore, the visualized feature maps in Fig.8 illustrate the contribution of each module to feature learning. The baseline model (w/o Both) suffers from high-frequency noise and low-quality features. DDFM enhances the transmission of structural information, while MBAF significantly improves the representation of key details. The full LGDR-Net produces the most uniform and strongly activated feature maps, confirming the effectiveness of incorporating both modules into the architecture.

TABLE IV: Ablation studies of different components in DRNet on the BSD68 dataset at different sampling rates.

| cases | DDFM | MBAF | Sampling Rate | |
|---|---|---|---|---|
| | | | 10% | 25% |
| (a) | - | - | 29.42/0.8361 | 34.44/0.9278 |
| (b) | ✓ | - | 29.47/0.8367 | 34.51/0.9285 |
| (c) | - | ✓ | 29.61/0.8377 | 34.57/0.9287 |
| (d) | ✓ | ✓ | **29.77/0.8410** | **34.79/0.9307** |




38.57/0.9713 | 38.68/0.9718 | 38.78/0.9721 | **39.15/0.9735**

Fig. 8: Visual analysis of feature maps in the 4th iteration of our proposed LGDR-Net.

### D. Noise Robustness

To simulate measurement noise in real-world scenarios, we add Gaussian noise to the Set11 dataset, setting five levels of noise variance from 0 to 0.008, and conduct comparative evaluations between our method and multiple SOTA methods. As shown in Fig.9, across the noise range from noise-free ($\sigma^2 = 0$) to high noise ($\sigma^2 = 0.008$), LGDR-Net consistently achieves the best performance, demonstrating that LGDR-Net can maintain robust reconstruction quality even when facing severe noise interference, demonstrating strong stability and resilience even under heavy noise interference.

### E. Complexity Study

We assess the complexity of LGDR-Net and compare it with representative SOTA methods in terms of MACs and

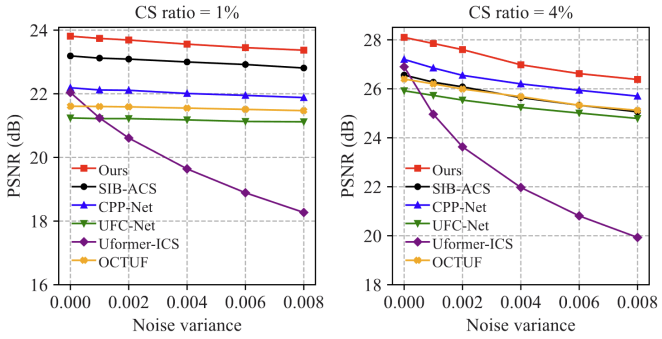



Fig. 9: Comparison of PSNR (dB) performance degradation on Set11 dataset at sampling rates of 1% and 4% under different Gaussian noise levels.

the number of parameters. All statistics are measured on $256 \times 256$ input images with a 25% sampling rate, as summarized in Tab.V. LGDR-Net requires only 97.18 GMACs per reconstruction, the lowest among all compared methods. Although UFC-Net has the smallest model size, its reconstruction performance (as shown in Tab.I) is significantly inferior to ours. Overall, LGDR-Net achieves an excellent balance between reconstruction quality and computational cost, making it a high-performance CS solution.

TABLE V: Comparison of Parameters (M) and MACs (G) for reconstructing a $256 \times 256$ image at sampling rate = 25%.

| Methods | UFC-Net | CPP-Net | SIB-ACS | LGDR-Net |
|---|---|---|---|---|
| Params (M) | **1.61** | 12.17 | 4.73 | 3.36 |
| MACs (G) | 121.09 | 162.91 | 308.97 | **97.18** |

## V. Conclusions

In this paper, we propose LGDR-Net, a novel adaptive compressive sensing framework that decouples sampling guidance from final reconstruction. The proposed lightweight guidance network utilizes more accurate intermediate reconstructions to facilitate refined sampling allocation. For the final image reconstruction, the proposed deep refinement network is employed to produce high-fidelity reconstruction results. Extensive experiments demonstrate LGDR-Net's superior performance over SOTA methods in image reconstruction quality with lower computational complexity.

## References

[1] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[2] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008.

[3] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[4] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021.

[5] Chengbo Li, Hong Jiang, Paul Wilford, Yin Zhang, and Mike Scheutzow. A new compressive video sensing framework for mobile broadcast. *IEEE Transactions on Broadcasting*, 59(1):197–205, 2013.

[6] Weisheng Dong, Guangming Shi, Xin Li, Yi Ma, and Feng Huang. Compressive sensing via nonlocal low-rank regularization. *IEEE transactions on image processing*, 23(8):3618–3632, 2014.

[7] Jian Zhang, Chen Zhao, Debin Zhao, and Wen Gao. Image compressive sensing recovery using adaptively learned sparsifying basis via l0 minimization. *Signal Processing*, 103:114–126, 2014.

[8] Jian Sun, Huibin Li, Zongben Xu, et al. Deep admm-net for compressive sensing mri. *Advances in neural information processing systems*, 29, 2016.

[9] Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1828–1837, 2018.

[10] Minghe Shen, Hongping Gan, Chao Ning, Yi Hua, and Tao Zhang. Transcs: A transformer-based hybrid architecture for image compressed sensing. *IEEE Transactions on Image Processing*, 31:6991–7005, 2022.

[11] Hongping Gan, Minghe Shen, Yi Hua, Chunyan Ma, and Tao Zhang. From patch to pixel: A transformer-based hierarchical framework for compressive image sensing. *IEEE Transactions on Computational Imaging*, 9:133–146, 2023.

[12] Noam Elata, Tomer Michaeli, and Michael Elad. Adaptive compressed sensing with diffusion-based posterior sampling. In *European Conference on Computer Vision*, pages 290–308. Springer, 2024.

[13] Yujun Huang, Bin Chen, Naiqi Li, Baoyi An, Shu-Tao Xia, and Yaowei Wang. Mb-racs: Measurement-bounds-based rate-adaptive image compressed sensing network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[14] Chen Hui, Shengping Zhang, Wenxue Cui, Shaohui Liu, Feng Jiang, and Debin Zhao. Rate-adaptive neural network for image compressive sensing. *IEEE Transactions on Multimedia*, 26:2515–2530, 2023.

[15] Kuiyuan Zhang, Zhongyun Hua, Yuanman Li, Yushu Zhang, and Yicong Zhou. Uformer-ics: A u-shaped transformer for image compressive sensing service. *IEEE Transactions on Services Computing*, 17(5):2974–2988, 2023.

[16] Ying Yu, Bin Wang, and Liming Zhang. Saliency-based compressive sampling for image signals. *IEEE signal processing letters*, 17(11):973–976, 2010.

[17] Ali Taimori and Farokh Marvasti. Adaptive sparse image sampling and recovery. *IEEE Transactions on Computational Imaging*, 4(3):311–325, 2018.

[18] Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3):265–274, 2009.

[19] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[20] Wuzhen Shi, Feng Jiang, Shaohui Liu, and Debin Zhao. Image compressed sensing using convolutional neural network. *IEEE Transactions on Image Processing*, 29:375–388, 2019.

[21] Jian Zhang, Chen Zhao, and Wen Gao. Optimization-inspired compact deep compressive sensing. *IEEE Journal of Selected Topics in Signal Processing*, 14(4):765–774, 2020.

[22] Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. Amp-net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing*, 30:1487–1500, 2020.

[23] Di You, Jingfen Xie, and Jian Zhang. Ista-net++: Flexible deep unfolding network for compressive sensing. *arXiv preprint arXiv:2103.11554*, 2021.

[24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[25] Jiechong Song, Chong Mou, Shiqi Wang, Siwei Ma, and Jian Zhang. Optimization-inspired cross-attention transformer for compressive sensing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6174–6184, 2023.

[26] Bin Chen and Jian Zhang. Practical compact deep compressed sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[27] Yan Qian, Ruiqing He, Qian Chen, Guohua Gu, Feng Shi, and Wenwen Zhang. Adaptive compressed 3d ghost imaging based on the variation of surface normals. *Optics Express*, 27(20):27862–27872, 2019.

[28] R Monika, Samiappan Dhanalakshmi, R Kumar, and R Narayanamoorthi. Coefficient permuted adaptive block compressed sensing for camera enabled underwater wireless sensor nodes. *IEEE Sensors Journal*, 22(1):776–784, 2021.

[29] Kuiyuan Zhang, Zhongyun Hua, Yuanman Li, Yongyong Chen, and Yicong Zhou. Ams-net: Adaptive multi-scale network for image compressive sensing. *IEEE Transactions on Multimedia*, 25:5676–5689, 2022.

[30] Bin Chen and Jian Zhang. Content-aware scalable deep compressed sensing. *IEEE Transactions on Image Processing*, 31:5412–5426, 2022.

[31] Chenxi Qiu and Xuemei Hu. Adacs: Adaptive compressive sensing with restricted isometry property-based error-clamping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4702–4719, 2024.

[32] Chenxi Qiu, Tao Yue, and Xuemei Hu. Adaptive and cascaded compressive sensing. *arXiv preprint arXiv:2203.10779*, 2022.

[33] Chenxi Qiu, Tao Yue, and Xuemei Hu. Reconstruction-free cascaded adaptive compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2620–2630, 2024.

[34] Zhifu Tian, Tao Hu, Chaoyang Niu, Di Wu, and Shu Wang. Sampling innovation-based adaptive compressive sensing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2387–2397, 2025.

[35] Heping Song, Jingyao Gong, Hongying Meng, and Yuping Lai. Multi-cross sampling and frequency-division reconstruction for image compressed sensing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4909–4917, 2024.

[36] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

[37] Michael Elad, Bahjat Kawar, and Gregory Vaksman. Image denoising: The deep learning revolution and beyonda survey paper. *SIAM Journal on Imaging Sciences*, 16(3):1594–1654, 2023.

[38] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

[39] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016.

[40] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 449–458, 2016.

[41] David Martin, Charless Fowlkes, Doron Tal, and Jitendra 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*, volume 2, pages 416–423. IEEE, 2001.

[42] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015.

[43] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

[44] Dongjie Ye, Zhangkai Ni, Hanli Wang, Jian Zhang, Shiqi Wang, and Sam Kwong. Csformer: Bridging convolution and transformer for compressive sensing. *IEEE Transactions on Image Processing*, 32:2827–2842, 2023.

[45] Jiechong Song, Bin Chen, and Jian Zhang. Dynamic path-controllable deep unfolding network for compressive sensing. *IEEE Transactions on Image Processing*, 2023.

[46] Hongping Gan, Zhen Guo, and Feng Liu. Nestd-net: Deep nesta-inspired unfolding network with dual-path deblocking structure for image compressive sensing. *IEEE Transactions on Image Processing*, 33:1923–1937, 2024.

[47] Xiaoyang Wang and Hongping Gan. Ufc-net: Unrolling fixed-point continuous network for deep compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25149–25159, 2024.

[48] Zhen Guo and Hongping Gan. Cpp-net: Embracing multi-scale feature fusion into deep unfolding cp-ppa network for compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25086–25095, 2024.