Multi-Cross Sampling and Frequency-Division Reconstruction for Image Compressed Sensing

Heping Song[1], Jingyao Gong[1], Hongying Meng[2], Yuping Lai[3]*

[1]School of Computer Science and Communication Engineering, Jiangsu University, China
[2]Electronic and Electrical Engineering Department, Brunel University London, United Kingdom
[3]School of Cyberspace Security, Beijing University of Posts and Telecommunications, China
songhp@ujs.edu.cn, gongjy@stmail.ujs.edu.cn, hongying.meng@brunel.ac.uk, laiyp@bupt.edu.cn

Abstract

Deep Compressed Sensing (DCS) has attracted considerable interest due to its superior quality and speed compared to traditional CS algorithms. However, current approaches employ simplistic convolutional downsampling to acquire measurements, making it difficult to retain high-level features of the original signal for better image reconstruction. Furthermore, these approaches often overlook the presence of both high- and low-frequency information within the network, despite their critical role in achieving high-quality reconstruction. To address these challenges, we propose a novel Multi-Cross Sampling and Frequency Division Network (MCFD-Net) for image CS. The Dynamic Multi-Cross Sampling (DMCS) module, a sampling network of MCFD-Net, incorporates pyramid cross convolution and dual-branch sampling with multi-level pooling. Additionally, it introduces an attention mechanism between perception blocks to enhance adaptive learning effects. In the second deep reconstruction stage, we design a Frequency Division Reconstruction Module (FDRM). This module employs a discrete wavelet transform to extract high- and low-frequency information from images. It then applies multi-scale convolution and self-similarity attention compensation separately to both types of information before merging the output reconstruction results. The MCFD-Net integrates the DMCS and FDRM to construct an end-to-end learning network. Extensive CS experiments conducted on multiple benchmark datasets demonstrate that our MCFD-Net outperforms state-of-the-art approaches, while also exhibiting superior noise robustness. The code is available at github.com/songhp/MCFD-Net.

Introduction

The theory of Compressed Sensing (CS) (Candes, Romberg, and Tao 2006) has experienced rapid development. CS enables the direct sampling of signals at sub-Nyquist-Shannon rates while preserving the necessary information for accurate reconstruction. This theory finds extensive application in various imaging domains, including remote sensing transmission (Pan et al. 2013; Ghahremani and Ghassemian 2014), medical imaging (Lustig, Donoho, and Pauly 2007; Lustig et al. 2008; Yang et al. 2017; Szczykutowicz and Chen 2010), single-pixel cameras (Duarte et al. 2008; Rousset et al. 2016), owing to its effective image compression and reconstruction capabilities. In the CS sampling process, assuming a signal $x \in \mathbb{R}^N$ and the sensing matrix $\Phi \in \mathbb{R}^{M \times N}$ with $\frac{M}{N}$ being the sampling rate, $M << N$, the CS measurements undergo a linear mapping $y = \Phi x$. The CS theory shows the signal x can be recovered from y by a sparsity-induced optimization problem

$$\min_x \|\Psi x\|_0 \; s.t. \; y = \Phi x \tag{1}$$

where Ψx are the sparse coefficients with respect to domain Ψ, and $\|\cdot\|_0$ denotes the ℓ_0 pseudo norm, i,e., the magnitude of non-zero elements. This gives rise to the two fundamental issues of CS: (1) how to design sampling matrix and (2) how to recover the original signal x based on its measurements y.

Recently, deep compressive sensing (DCS) methods (Sun et al. 2020; Zhang et al. 2020; Chen et al. 2020; You et al. 2021; Song, Chen, and Zhang 2021) have been developed to solve these two issues of CS through an end-to-end learning manner, leveraging the robust learning and representation abilities of neural networks. Zheng et al. (2020) introduced RK-CCSNet, a method that employs sequential convolution modules (SCM) to compress image size by means of filter compression, This approach effectively avoids block

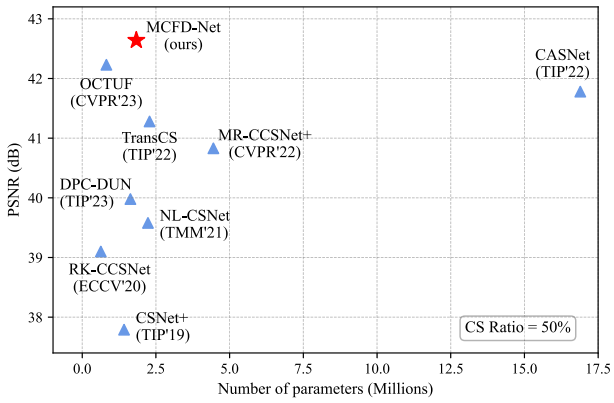


Figure 1: The work advances the state-of-the-art in image CS by achieving an enhanced tradeoff between model complexity (as measured by the number of network parameters) and performance (as indicated by the PSNR metric). Our proposed methods, highlighted in red, attain this balance.

*Corresponding author.

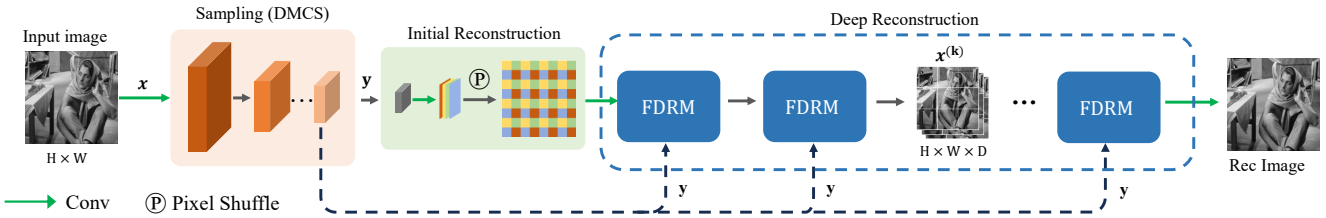


Figure 2: Architecture of our MCFD-Net.

artifacts and high-frequency noise. In a similar vein, Cui et al. (2021) proposed NL-CSNet, which utilizes a deep network with non-local self-similarity prior and designs a new loss function for end-to-end training. Chen, Yang, and Yang (2022) proposed FSOINet, which introduces optimization concepts into feature domain rather than pixel domain, resulting in a significant reduction in network complexity and notable performance improvements. More recently, Fan, Lian, and Quan (2022) proposed MR-CCSNet, which leverages CNN hierarchical features to enhance measurement utilization under challenging conditions, leading to substantial advancements.

Although these DCS methods have shown promising results for image CS, they face two significant challenges. Firstly, during the sensing stage, they utilize a simple convolutional downsampling technique that fails to preserve the high-level features of the original signal. Secondly, during the deep reconstruction stage, the methods rely solely on multi-scale convolution to reconstruct the original image, disregarding global dependencies and frequency information present in the intermediate reconstruction results.

To address aforementioned challenges, we propose a multi-level cross-sampling and frequency-divided reconstruction network (MCFD-Net) to achieve higher quality image CS, as illustrated in Fig. 2. The MCFD-Net consists of two key modules: the Dynamic Multi-Cross Sampling (DMCS) module, serving as the sensing network, and the Frequency-Division Reconstruction Module (FDRM), acting as the deep reconstruction network. DMCS utilizes pyramid convolution to extract features from various scales of receptive fields. It is connected to dual-branch pooling layers through skip connections to mitigate the loss of feature information flow. Moreover, perception blocks are employed to apply a hierarchical adaptive attention mechanism between blocks, thereby enhancing important feature channels. In the deep reconstruction phase, the FDRM incorporates the discrete wavelet transform (DWT) to segment images into high and low frequencies. It guides the reconstructions of different frequencies through multi-scale convolutions, compensates for the local correlations lost during convolution using self-similarity attention, and finely fuses the optimized reconstruction results. Experimental results demonstrate that the MCFD-Net outperforms state-of-the-art image CS methods, striking a superior balance between cost and performance, as depicted in Fig. 1. The main contributions of this paper can be summarized as follows: (1) the development of the DMCS module, which efficiently samples whole images for CS and learns high-level features to guide high-performance reconstruction; (2) the proposal of the FDRM, which introduces the DWT to divide frequencies in the feature domain and compensates for missing multi-scale reconstructions; and (3) the construction of an end-to-end trainable MCFD-Net incorporating the DMCS and FDRM modules.

Related Work

Deep Compressed Sensing

The DCS methods (Shi et al. 2019a; Zhang, Zhao, and Gao 2020; Mdrafi and Gurbuz 2020; Zhou et al. 2020; Shi et al. 2019b) facilitate the rapid acquisition of high-quality image reconstructions by leveraging neural networks to learn the nonlinear mapping between measurements y and source signals x. The DCS network is trained through the minimization of the mean square error

$$\min_{\theta} \frac{1}{2} \sum_{i=1}^{k} \|x_i - F(y_i, \theta)\|_2^2, \tag{2}$$

where F represents the network model parameterized by θ. The DCS network can learn the inverse mapping and achieve high-quality image reconstruction even under low sampling rates. However, current methods exhibit limitations concerning perception network measurements, and the exclusive use of multi-scale convolutional layers for reconstruction proves inadequate in handling complex scene challenges. Consequently, it is essential to investigate a novel DCS model that enhances the efficiency of image perception and modeling.

Attention Mechanism

The attention mechanism is a reasonable allocation enhancement method. Previous studies, Hu, Shen, and Sun (2018), Woo et al. (2018), Roy, Navab, and Wachinger (2018), Chen et al. (2018) have demonstrated that the combination of channel and spatial attention can improve the performance of CNNs. The self-attention mechanism has good contextual correlation ability. Recently, Shen et al. (2022) and Song et al. (2023) propose optimization-inspired network for image CS, where the attention mechanism is extensively employed to construct the foundational Transformer module. These methods employ non-overlapping image blocks sampling strategy, so do not fully exploit the advantages of convolution with the whole images. In this paper, we combine attention mechanism and pyramid cross convolution in sampling network, as well as multi-scale convolution based on

frequency division in reconstruction network to build an efficient CS framework.

Methodology

Overall Architecture

We will present the proposed network in the case of sampling rate is 25%. Fig. 2 illustrates the model of MCFD-Net. MCFD-Net comprises three key modules: A sampling network called DMCS that acquires measurements from the input image. An initial reconstruction network that generates an initial recovered image using a linear mapping. A deep reconstruction network named FDRM that refines the initial recovered image. Multiple FDRMs are stacked within the deep reconstruction network to enable the transition from measurements to high-quality reconstructions.

The sampling network $S(\cdot)$ consists of pyramid cross convolution and multi-level pooling with dual branch sampling, which allows for referencing high-level features of the original signal during inverse mapping reconstruction to guide more effective reconstruction. Furthermore, an attention mechanism is introduced between perception blocks to enhance adaptive learning effects of the sampling network. The sampling process can be represented as follows:

$$y = S(x) \tag{3}$$

where $x \in \mathbb{R}^{1\times H\times W}$ and $y \in \mathbb{R}^{4\times \frac{H}{4}\times \frac{W}{4}}$.

The initial reconstruction network $I(\cdot)$ performs a depth-wise convolution layer on measurement y, expanding the channel dimension to $16\times \frac{H}{4}\times \frac{W}{4}$. It then apply pixel shuffle layer $PS(\cdot)$ to obtain the initial recovered image denoted by a tensor of size $1\times H\times W$. This result and the measurement are both utilized to guide deep reconstruction.

Within the deep reconstruction model $D(\cdot)$, the initially recovered image $I(y)$ is transformed into a high-dimensional feature vector using a convolutional layer. Subsequently, multiple cascaded FDRM blocks with identical internal structures fuse these features with matched features extracted from the measurements y at various scales and frequency fusion with attention mechanism.

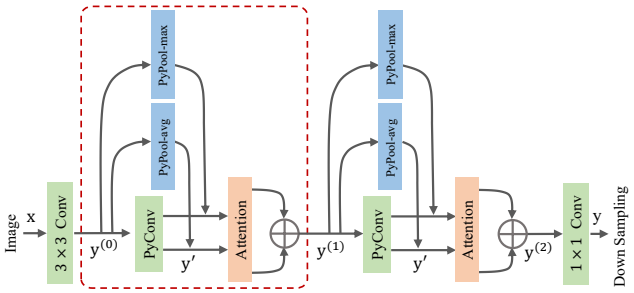
The final reconstruction result $\hat{x}$ represents a joint outcome of the initial reconstruction and deep reconstruction:

$$\hat{x} = D(I(y)) + I(y) \tag{4}$$

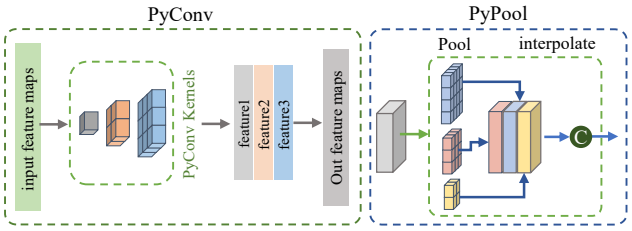
Two novel modules, DMCS and FDRM, will be described below.

Dynamic Multi-Cross Sampling Module

Well-designed feature extraction networks can assist with sampling. In a pyramid structure, the receptive field gradually increases, allowing the network to obtain richer contextual information. This differs from obtaining feature maps solely from convolutions and pooling operations (Fan, Lian, and Quan 2022), which may result in the loss of features at different levels during transmission. Generally, shallower layers can learn low-level features such as lines and textures, while deeper layers in the model can learn higher-order features such as objects and shapes. Based on these principles,



(a) DMCS backbone (illustrated at sampling rate 25%)



(b) Internal structure

Figure 3: Overview of DMCS.

we propose a novel DMCS module to more effectively utilize sampled low-, medium-, and high-level features to guide reconstruction. It utilizes pyramid convolutions for feature extraction and incorporates residual connections with dual-branch pyramid pooling layers to address feature information loss during pooling. Meanwhile, we employ an adaptive attention mechanism between hierarchical levels to enhance important feature channels and maximize useful information. The building blocks of DMCS can be described as follows:

$$y' = PyConv(y^{(k-1)}), \tag{5}$$

$$\begin{aligned} y^{(k)} = {} & mA(y' + PyPool_{\text{max}}(y^{(k-1)})) \\ & \oplus aA(y' + PyPool_{\text{avg}}(y^{(k-1)})), \end{aligned} \tag{6}$$

where $PyConv$ and $PyPool$ correspond to pyramid convolution and pooling layers, $\oplus$ is a concatenation operation, mA and aA represent the high-level channel attention of branches with max-pooling, average-pooling respectively. Specifically, as illustrated in Fig. 3, at a sampling rate of 25%, the input image is first expanded from 1 to n channels using a 3×3 convolutional layer with stride 1, in particular, $y^{(0)}$ is $Conv(x)$. The DMCS block (red box) then subjects the input features to pyramid convolution and multi-scale pooling on the feature map, yielding two branches of fusion-ready features. Next, the current perception layer output is obtained by the fusion of the dual-channel attention features. Finally, we utilize a 1×1 convolutional layer to linearly combine all the features at each level of the network. The sampling ratio is determined by the number K ($k \in [0, K]$) of DMCS blocks, making it highly flexible and easy to employ at diverse sampling ratios through repeating the blocks.

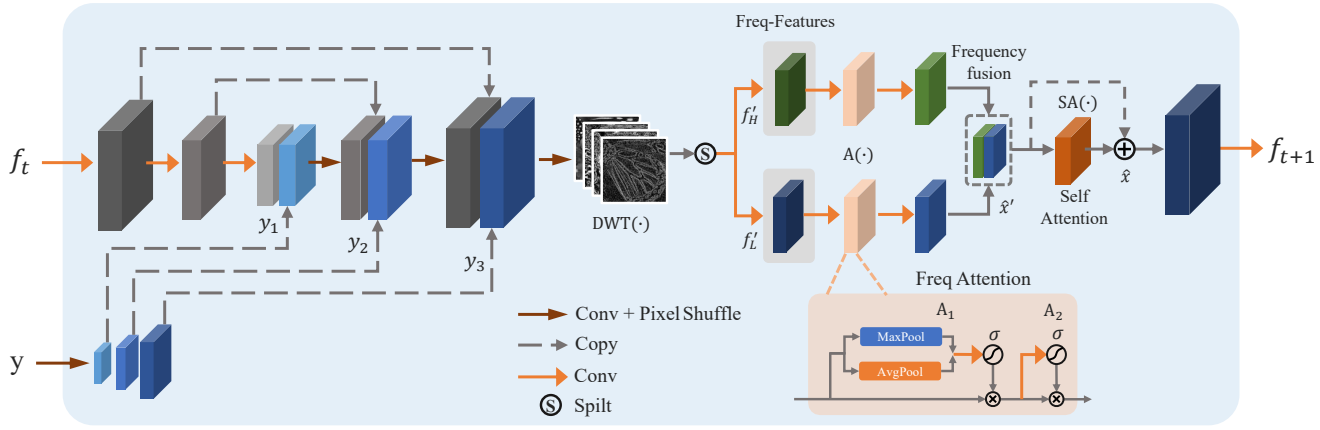


Figure 4: Overview of FDRM schematic diagram, including two submodules of frequency division convolution. The reconstructed feature map output is measured three times to supplement the reconstructed image details.

Frequency Division Reconstruction Module

We employ a residual learning-based deep reconstruction network to implement the nonlinear signal reconstruction process for improved reconstruction accuracy. Following the method proposed by Fan, Lian, and Quan (2022), it leverages the advantages of utilizing multiple scales and multiple measurements by using the measurement reuse block (MRB). However, only performing convolution at different scales in the feature domain means that MRB treats low-frequency and high-frequency feature maps equally, resulting in a significant loss of detail benefits in high frequencies. Moreover, MRB has tailored approximately 20 convolutional block layers for specific sampling rates, which substantially increases complexity and lacks flexibility. Therefore, there is a need to explore an approach that can dynamically adapt to the sampling rate, further leverage the benefits of measurement reuse, and considerably reduce complexity without compromising the quality of reconstructed features across both high and low frequencies.

Based on the above analysis, we propose the FDRM, as shown in Figure 4. Firstly, we perform convolution and pixel rearrangement on phased recovered result f_t and measurements y. Unlike MRB, which customizes the number of convolution layers and different dimensions for pixel shuffling, we uniformly expand the dimension of measurements y to $\mathbb{R}^{C\times H\times W}$. Then, we use three cascaded convolutional (CV) layers to obtain compressed feature maps:

$$f^{\downarrow} = CV_1(f_t),\ f^{\uparrow\uparrow\uparrow} = PS(CV_6(f^{\downarrow} \oplus y_3)) \quad (7)$$

$$f^{\downarrow\downarrow} = CV_2(f^{\downarrow}),\ f^{\uparrow\uparrow} = PS(CV_5(f^{\downarrow\downarrow} \oplus y_2)) \quad (8)$$

$$f^{\downarrow\downarrow\downarrow} = CV_3(f^{\downarrow\downarrow}),\ f^{\uparrow} = PS(CV_4(f^{\downarrow\downarrow\downarrow} \oplus y_1)) \quad (9)$$

where $f^{\downarrow} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, $f^{\downarrow\downarrow} \in \mathbb{R}^{C\times\frac{H}{4}\times\frac{W}{4}}$, and $f^{\downarrow\downarrow\downarrow} \in \mathbb{R}^{C\times\frac{H}{8}\times\frac{W}{8}}$. By performing multi-scale measurements on them and then extracting three types of matching information $y_1 \in \mathbb{R}^{C\times\frac{H}{8}\times\frac{W}{8}}, y_2 \in \mathbb{R}^{C\times\frac{H}{4}\times\frac{W}{4}}, y_3 \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$, To preserve the current reconstruction results, we copy $f^{\downarrow}$, $f^{\downarrow\downarrow}$ and $f^{\downarrow\downarrow\downarrow}$, and fuse them with $f^{\uparrow\uparrow\uparrow}$, $f^{\uparrow\uparrow}$ and $f^{\uparrow}$ through convolutional layers. Finally, we use a pixel shuffle layer to expand the fused feature maps to obtain f_3 for further processing.

We further use the Discrete Wavelet Transform $DWT(\cdot)$ to divide the frequency components of the reconstructed result f_3 into four frequency sub-bands: low-low (LL), low-high (LH), high-low (HL), and high-high (HH). We combine the three high-frequency components (LH, HL, HH) into a unified channel component f_H through convolutional merging, and restore their size using Pixel Shuffle. Then, we merge them separately with the low-frequency component f_L for next operation. The process can be represented as follows:

$$f'_L, f'_H = DWT(f_3)_{LL}, DWT(f_3)_{LH\oplus HL\oplus HH} \quad (10)$$

$$f_L, f_H = Conv_L(f'_L), Conv_H(f'_H) \quad (11)$$

Low-frequency signals typically represent structural information, while high-frequency signals represent the details but also contain noise. By dividing the feature maps according to frequency division, it is enable to compensate the defects of the convolution operation with fixed frequencies. This promotes denoising of high-frequency feature maps and restoration of edge details. Therefore, in feature fusion, a dynamic attention block A is assigned to learn the contribution of both low and high frequency features, including channel attention A_1 and spatial attention A_2. This results in fused feature maps that enhance important feature representations. The attention mechanism can be represented as follows:

$$A_1 = \sigma(Conv(MaxPool(x) + AvgPool(x))) \quad (12)$$

$$A_2 = \sigma(Conv_{(7\times7)}(x_{avg}^{dim=1} \oplus x_{max}^{dim=1})) \quad (13)$$

where σ denotes the sigmoid function, $MaxPool, AvgPool$ represent the max-pooling, average-pooling respectively, $\otimes$ represents element-wise multiplication. In the process of feature fusion and reconstruction, we introduce a self-attention block SA in different stages to compensate for the limitations of convolutional kernel receptive fields and

Dataset	Methods	CS Ratio					
		50%	25%	12.5%	6.25%	3.125%	1.5625%
Set11	CSNet$^+$ (TIP'19)	36.08/0.9469	31.86/0.9023	28.00/0.8178	25.33/0.7140	23.20/0.5916	21.35/0.4780
	RK-CCSNet (ECCV'20)	37.67/0.9620	31.85/0.9106	28.26/0.8427	25.63/0.7342	23.41/0.6188	21.60/0.4944
	NL-CSNet (TMM'21)	37.78/0.9641	32.37/0.9223	28.72/0.8601	25.68/0.7583	23.45/0.6289	21.69/0.4951
	MR-CCSNet$^+$ (CVPR'22)	39.94/0.9686	34.21/0.9290	30.20/0.8738	26.94/0.7827	24.19/0.6582	22.13/0.5337
	TransCS (TIP'22)	40.88/0.9752	34.73/0.9361	-/-	-/-	-/-	-/-
	CASNet (TIP'22)	41.46/0.9706	35.17/0.9319	30.88/0.8765	27.42/0.7909	24.37/0.6679	22.22/0.5413
	DPC-DUN (TIP'23)	41.04/0.9831	33.94/0.9309	-/-	-/-	-/-	-/-
	OCTUF (CVPR'23)	42.37/0.9769	35.67/**0.9420**	-/-	-/-	-/-	-/-
	MCFD-Net (ours)	**42.92/0.9870**	**36.72**/0.9415	**31.72/0.8865**	**28.47/0.8100**	**25.26/0.6932**	**22.82/0.5680**
BSDS	CSNet$^+$ (TIP'19)	36.52/0.9627	32.65/0.9155	29.35/0.8312	26.93/0.7284	25.00/0.6256	23.25/0.5306
	RK-CCSNet (ECCV'20)	37.69/0.9742	32.77/0.9248	29.54/0.8496	26.97/0.7393	25.12/0.6437	23.36/0.5368
	NL-CSNet (TMM'21)	38.17/0.9785	33.14/0.9360	29.84/0.8685	27.30/0.7713	25.24/0.6555	23.52/0.5419
	MR-CCSNet$^+$ (CVPR'22)	39.39/0.9806	33.96/0.9373	30.46/0.8677	27.86/0.7725	25.63/0.6664	23.90/0.5681
	TransCS (TIP'22)	39.83/0.9827	34.23/0.9435	-/-	-/-	-/-	-/-
	CASNet (TIP'22)	40.25/0.9813	34.57/0.9387	30.85/0.8687	28.02/0.7727	25.80/0.6701	23.92/0.5725
	DPC-DUN (TIP'23)	38.56/0.9733	33.55/0.9214	-/-	-/-	-/-	-/-
	OCTUF (CVPR'23)	40.64/0.9847	34.79/0.9480	-/-	-/-	-/-	-/-
	MCFD-Net (ours)	**41.12/0.9848**	**35.71/0.9484**	**31.67/0.8823**	**29.03/0.7950**	**26.62/0.6925**	**24.65/0.5956**

Table 1: The performance of comparison results with advanced methods in terms of PSNR (dB) and SSIM. The top two results are highlighted in bold and underlined.

weight sharing. We also incorporate residual connections to pass the self-attention block, thereby preserving the flow of information in the feature domain. The process can be represented as follows:

$$\hat{x}' = Conv(A(f_H) \oplus A(f_L)) \tag{14}$$

$$\hat{x} = SA(\hat{x}') \oplus \hat{x}' \tag{15}$$

Inspired by the idea of residual learning (Zhang and Ghanem 2018; Chen, Yang, and Yang 2022), we further utilize the measurements y, sampling network $S(\cdot)$, and initial reconstruction network $I(\cdot)$ to learn the residual compensation of recovered features $\hat{x}$ This process can be defined as:

$$\Delta = I(S(\hat{x}) - y) \tag{16}$$

$$\hat{x} \longleftarrow \hat{x} + \Delta \tag{17}$$

where Δ represents the difference in the feature domain. By applying $I(\cdot)$, we obtain the residual information of the current reconstructed features projected onto the observation domain. It continuously adjusts the feature maps to make them conform as closely as possible to the inverse downsampling process. Finally, the phased result $f_{t+1} = \hat{x}$ is utilized as the input for the next phase. Subsequent experiments have demonstrated that compared to MRB, FDRM significantly reduces redundant computational costs while achieving better reconstruction quality.

Loss Function

MCFD-Net is an end-to-end network for reconstructing the original image x from the measurements y. In MCFD-Net, we use mean squared error (MSE) to measure reconstruction quality. The loss function of overall reconstruction network can be expressed as:

$$L = L_I + L_D \tag{18}$$

where L_I and L_D represent initial reconstruction loss and deep reconstruction loss respectively, specifically:

$$L_i = \frac{\sum_{j=1}^{n} \|I(S(x_j, \theta), \phi_I) - x_j\|_F^2}{2n} \tag{19}$$

$$L_d = \frac{\sum_{j=1}^{n} \|D(I(S(x_j, \theta), \phi_I), \phi_D) - x_j\|_F^2}{2n} \tag{20}$$

Where n is the number of training samples, x_j represents the j-th image in the training set, θ, ϕ_I and ϕ_D refer to the network parameters of the sampling network $S(\cdot)$, the initial reconstruction network $I(\cdot)$, the deep reconstruction network $D(\cdot)$, respectively.

Experiments

Implementation Details

For training purposes, we selected 40,000 sub-images from the COCO 2014 dataset, each with a size of 96×96. These sub-images were randomly cropped and flipped. To enhance computational efficiency and model robustness, we converted the images to the YCbCr color space and utilized only the Y channel during both training and testing phases. The experimental results were evaluated on three benchmark datasets: Set11 (Kulkarni et al. 2016), BSDS100 (Martin et al. 2001), and Urban100 (Huang, Singh, and Ahuja 2015).

During training, we used Adam optimizer (Kingma and Ba 2015) to update model parameters with momentum and weight decay set at 0.9 and 0.999 respectively. By conveniently stacking the DMCS Block in the sampling network, we trained six different sampling rates for our model: 50%, 25%, 12.5%, 6.25%, 3.125%,a nd 1.5625%. We assessed the reconstruction performance of network models using two metrics: PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity).

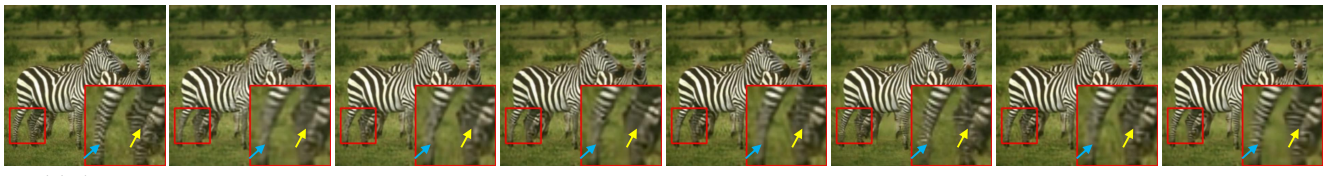


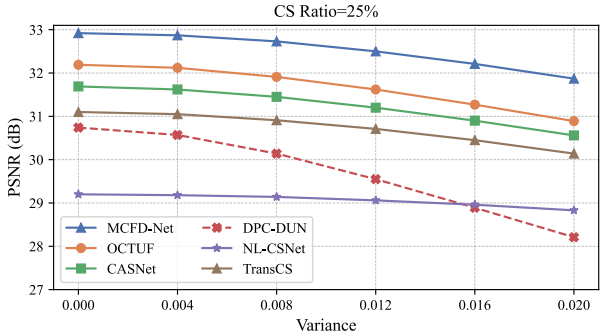
Figure 5: The reconstruction results at 25% sampling rate, competitive methods have produced varying degrees of noise and ringing effects in restoring zebra stripe details, resulting in obvious distortion. Our method reconstructs details most accurately.

Figure 6: The reconstruction results at an extremely low (1.5625%) sampling rate show that competitive methods produce distorted or erroneous results in the restoration of hand and car window contours, while our method still achieves the highest quality.

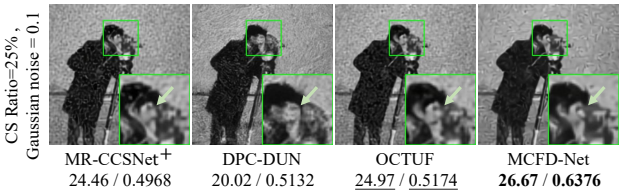
Comparison with the State-of-the-Arts

To assess the effectiveness of our proposed MCFD-Net, we conducted a comparative analysis against state-of-the-art methods, namely CSNet$^+$ (Shi et al. 2019a), RK-CCSNet (Zheng et al. 2020), NL-CSNet (Cui et al. 2021), MR-CCSNet$^+$ (Fan, Lian, and Quan 2022), TransCS (Shen et al. 2022), CASNet (Chen and Zhang 2022), DPC-DUN (Song, Chen, and Zhang 2023), and OCTUF (Song et al. 2023). The results, as presented in Table 1, demonstrate that MCFD-Net consistently outperforms these methods. Specifically, on the Set11 dataset, MCFD-Net achieves average PSNR gains of 1.72dB (5.80%) and 1.07dB (3.53%) over MR-CCSNet$^+$ and CASNet, respectively. Furthermore, when evaluating a slightly larger sample set, BSDS100, at a sampling rate of 25%, MCFD-Net exhibits PSNR improvements of 1.48dB (4.32%) and 0.92dB (2.64%) over TransCS and OCTUF, respectively.

For visual comparison, we conducted tests by randomly selecting images from BSDS100 and Urban100 datasets. Figure 5 demonstrates the superior performance of our method in terms of artifact and noise control, as well as sharpness and detail quality, particularly at extremely low sampling rates (1.5625%). Our approach, leveraging the multi-level cross-measurement capability of DMCS and FDRM sub-band enhancement reconstruction, addresses the issue of detail loss evident in Figure 6. Competing methods tend to overlook hand details, resulting in significant loss, whereas our MCFD-Net preserves them more effectively. As a result, MCFD-Net exhibits enhanced texture and detail effects, particularly in low sampling rate scenarios and complex scenes, where this advantage becomes more pronounced.



(a) Performance degradation under different levels of noise.



(b) Visual assessment of noise degradation.

Figure 7: We assess the robustness of our method by comparing its performance under varying levels of Gaussian noise.

Noise Robustness

We utilized the Urban100 dataset and introduced Gaussian random noise to the images at different levels, specifically five standard deviations of 0.004, 0.008, 0.012, 0.016, and 0.020. The reconstruction results were then evaluated against the original noise-free images. Table 2 and Figure 7 show statistical results for five noise levels and two sampling rates. The results show that our method exhibits minimal performance degradation across all noise variances. Among them, DPC-DUN, TransCS, CASNet and OCTUF show noise attenuation of 2.52dB, 0.94dB, 1.14dB and 1.32dB respectively while MCFD-Net achieves a lower degradation of only 1.05dB. Notably, MCFD-Net consistently outperforms all competing methods in terms of PSNR and exhibits lower

ratio	Methods	Gaussian noise variance			decay
		0.000	0.008	0.020	(dB)
25%	MCFD-Net	32.92	32.73	31.87	1.05
	OCTUF	32.19	31.91	30.87	1.32
	CASNet	31.69	31.45	30.55	1.14
	DPC-DUN	30.74	30.14	28.22	2.52
	TransCS	31.10	30.93	30.16	0.94
	NL-CSNet	29.20	29.14	28.83	0.37
12.5%	MCFD-Net	28.05	28.01	27.86	0.19
	CASNet	27.61	27.55	27.28	0.33
	MR-CCSNet$^+$	26.89	26.85	26.66	0.23
	RK-CCSNet	25.75	25.73	25.62	0.13
	CSNet$^+$	25.53	25.51	25.42	0.11
	NL-CSNet	26.01	25.99	25.90	0.11

Table 2: Comparison of noise robustness results based on Urban100 dataset.

performance degradation among high-performance methods, highlighting its excellent noise robustness.

Comparisons of Running Times

We conducted a comparative analysis of parameter quantities and algorithm inference runtime for several competitive methods. The comparison of parameter quantities is presented in Figure 1, while the runtime statistics are displayed in Table 3. The CPU utilized was an Intel(R) Core(TM) i9-10980XE, and the GPU employed was an NVIDIA GeForce RTX 3090. Notably, MCFD-Net can reconstruct a 128×128 size image in an average of 65 milliseconds of GPU time, which is comparable to that of methods with similar parameter quantities. The results indicate that the FDRM exhibits a lower parameter quantity than MRB while still achieving effective performance improvement. Often, there exists a clever trade-off relationship between computational complexity and CS quality. Considering its superior CS performance improvement over competing methods, a slight increase in computing parameters is both justifiable and acceptable.

Methods	GPU / CPU (s)		
	Ratio=50%	Ratio=25%	Ratio=12.5%
MCFD-Net	0.058/0.471	0.067/0.492	0.071/0.591
DPC-DUN	0.052/0.147	0.051/0.154	-/-
CASNet	0.221/13.490	0.219/13.582	0.226/13.365
OCTUF	0.066/0.503	0.064/0.437	-/-
TransCS	0.130/0.583	0.128/0.588	-/-

Table 3: The comparison of the inference time between GPU and CPU using the BSDS100 dataset, which consists of 128×128 sized images.

Ablation Studies

We conducted ablation experiments on the DMCS and FDRM sub-modules on the BSDS100 dataset. The experimental results are presented in Table 4, where (a) corresponds to MR-CCSNet$^+$ (Fan, Lian, and Quan 2022). It is evident that both DMCS and FDRM effectively enhance the quality of compressed sensing. However, DMCS plays a more crucial role than FDRM. Therefore, when the complexity of reconstruction methods encounters limitations, improving sampling methods can yield greater benefits compared to reconstruction methods. This implies that high-performance sampling can guide high-performance reconstruction, and both aspects should be balanced. In summary, DMCS significantly improves downsampling performance and reduces feature loss. We conducted a visual analysis of the DMCS module, as depicted in Figure 8. On the other hand, FDRM separately combines high-frequency and low-frequency information from images to enhance reconstruction, resulting in notable performance improvements and a substantial reduction in network parameters.

cases	DMCS	FDRM	CS Ratio	
			25.00%	1.5625%
(a)	-	-	33.96/0.9373	23.90/0.5681
(b)	-	√	34.37/0.9401	24.26/0.5794
(c)	√	-	35.40/0.9463	24.34/0.5854
(d)	√	√	35.71/0.9484	24.65/0.5956

Table 4: The ablation experiments of MCDF-Net on BSDS100 dataset.

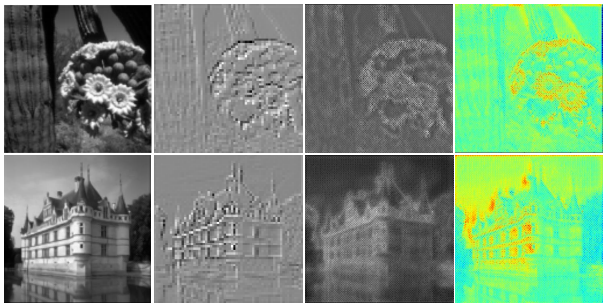


Figure 8: The feature map visualization of MCFD-Net during the reconstruction block processing includes high-frequency components features for the 2-en and 4-th layers of iterative blocks, as well as a heatmap of self-attention mechanism features.

Conclusions

In this paper, we propose a novel multi-level cross-perception and frequency division reconstruction method (MCFD-Net) to achieve higher quality image CS. The sampling module DMCS effectively captures and retains fine image details to support high-quality reconstruction. The deep reconstruction module FDRM employs a frequency division and joint learning strategy to restore high-performance image quality while suppressing corruption from high-frequency noise. Experimental results demonstrate that our proposed MCFD-Net outperforms current state-of-the-art methods across various sampling rates and scenarios, maintaining excellent performance particularly at medium-low sampling rates and in noisy cases. In the future, We plan to extend MCFD-Net to video and other inverse problems as a way to advance the great potential of this work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62172193, 62272051).

References

Candes, E. J.; Romberg, J.; and Tao, T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2): 489–509.

Chen, B.; and Zhang, J. 2022. Content-aware scalable deep compressed sensing. *IEEE Transactions on Image Processing*, 31: 5412–5426.

Chen, J.; Sun, Y.; Liu, Q.; and Huang, R. 2020. Learning memory augmented cascading network for compressed sensing of images. In *European Conference Computer Vision*, 513–529.

Chen, W.; Yang, C.; and Yang, X. 2022. FSOINET: Feature-space optimization-inspired network for image compressive sensing. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2460–2464.

Chen, Y.; Kalantidis, Y.; Li, J.; Yan, S.; and Feng, J. 2018. A^2-nets: Double attention networks. In *Advances in Neural Information Processing Systems*, 350–359.

Cui, W.; Liu, S.; Jiang, F.; and Zhao, D. 2021. Image compressed sensing using non-local neural network. *IEEE Transactions on Multimedia*, 25: 816–830.

Duarte, M. F.; Davenport, M. A.; Takhar, D.; Laska, J. N.; Sun, T.; Kelly, K. F.; and Baraniuk, R. G. 2008. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2): 83–91.

Fan, Z.-E.; Lian, F.; and Quan, J.-N. 2022. Global sensing and measurements reuse for image compressed sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8954–8963.

Ghahremani, M.; and Ghassemian, H. 2014. Remote sensing image fusion using ripplet transform and compressed sensing. *IEEE Geoscience and Remote Sensing Letters*, 12(3): 502–506.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-Excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Huang, J.-B.; Singh, A.; and Ahuja, N. 2015. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5197–5206.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Kulkarni, K.; Lohit, S.; Turaga, P.; Kerviche, R.; and Ashok, A. 2016. ReconNet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 449–458.

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): 1182–1195.

Lustig, M.; Donoho, D. L.; Santos, J. M.; and Pauly, J. M. 2008. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2): 72–82.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, 416–423.

Mdrafi, R.; and Gurbuz, A. C. 2020. Joint learning of measurement matrix and signal reconstruction via deep learning. *IEEE Transactions on Computational Imaging*, 6: 818–829.

Pan, Z.; Yu, J.; Huang, H.; Hu, S.; Zhang, A.; Ma, H.; and Sun, W. 2013. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9): 4864–4876.

Rousset, F.; Ducros, N.; Farina, A.; Valentini, G.; d'Andrea, C.; and Peyrin, F. 2016. Adaptive basis scan by wavelet prediction for single-pixel imaging. *IEEE Transactions on Computational Imaging*, 3(1): 36–46.

Roy, A. G.; Navab, N.; and Wachinger, C. 2018. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Transactions on Medical Imaging*, 38(2): 540–549.

Shen, M.; Gan, H.; Ning, C.; Hua, Y.; and Zhang, T. 2022. TransCS: A transformer-based hybrid architecture for image compressed sensing. *IEEE Transactions on Image Processing*, 31: 6991–7005.

Shi, W.; Jiang, F.; Liu, S.; and Zhao, D. 2019a. Image compressed sensing using convolutional neural network. *IEEE Transactions on Image Processing*, 29: 375–388.

Shi, W.; Jiang, F.; Liu, S.; and Zhao, D. 2019b. Scalable convolutional neural network for image compressed sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 12290–12299.

Song, J.; Chen, B.; and Zhang, J. 2021. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the ACM International Conference on Multimedia*, 4249–4258.

Song, J.; Chen, B.; and Zhang, J. 2023. Dynamic path-controllable deep unfolding network for compressive sensing. *IEEE Transactions on Image Processing*, 32(5): 2202–2214.

Song, J.; Mou, C.; Wang, S.; Ma, S.; and Zhang, J. 2023. Optimization-inspired cross-attention transformer for compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6174–6184.

Sun, Y.; Chen, J.; Liu, Q.; Liu, B.; and Guo, G. 2020. Dual-path attention network for compressed sensing image reconstruction. *IEEE Transactions on Image Processing*, 29: 9482–9495.

Szczykutowicz, T. P.; and Chen, G.-H. 2010. Dual energy CT using slow kVp switching acquisition and prior image

constrained compressed sensing. *Physics in Medicine & Biology*, 55(21): 6411.

Woo, S.; Park, J.; Lee, J.-Y.; and Kweon, I. S. 2018. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, 3–19.

Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P. L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; et al. 2017. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, 37(6): 1310–1321.

You, D.; Zhang, J.; Xie, J.; Chen, B.; and Ma, S. 2021. COAST: Controllable arbitrary-sampling network for compressive sensing. *IEEE Transactions on Image Processing*, 30: 6066–6080.

Zhang, J.; and Ghanem, B. 2018. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1828–1837.

Zhang, J.; Zhao, C.; and Gao, W. 2020. Optimization-inspired compact deep compressive sensing. *IEEE Journal of Selected Topics in Signal Processing*, 14(4): 765–774.

Zhang, Z.; Liu, Y.; Liu, J.; Wen, F.; and Zhu, C. 2020. AMP-Net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing*, 30: 1487–1500.

Zheng, R.; Zhang, Y.; Huang, D.; and Chen, Q. 2020. Sequential convolution and runge-kutta residual architecture for image compressed sensing. In *European Conference Computer Vision*, 232–248.

Zhou, S.; He, Y.; Liu, Y.; Li, C.; and Zhang, J. 2020. Multi-channel deep networks for block-based image compressive sensing. *IEEE Transactions on Multimedia*, 23: 2627–2640.