Breaking Measurement Barriers: From Compressed Sensing to Deep Reconstruction

Gang Qu[1], **Ping Wang**[1*], **Siming Zheng**[2], **Xin Yuan**[1*]

[1]School of Engineering, Westlake University, Hangzhou, Zhejiang, China.
[2]Vivo Mobile Communication Co., Ltd., Hangzhou, Zhejiang, China.
{qugang,wangping,xyuan}@westlake.edu.cn, zhengsiming@vivo.com

Abstract

Deep learning methods have achieved remarkable success in image compressed sensing (CS) task, namely reconstructing a high-fidelity image from its compressed measurement. However, existing methods are deficient in *incoherent compressed measurement* at sensing phase and *implicit measurement representations* at reconstruction phase, limiting the overall performance. In this work, we answer two questions: (i) how to improve the measurement incoherence for decreasing the ill-posedness; (ii) how to learn informative representations from measurements. To this end, we propose a novel asymmetric Kronecker CS (AKCS) model and theoretically present its better incoherence than previous Kronecker CS with minimal increase of complexity. Moreover, apart from the explicit measurement representations in gradient descent projection in unfolding networks, we further propose a measurement-aware cross attention (MACA) mechanism to learn implicit measurement representations. We integrate AKCS and MACA into a widely-used unfolding architecture to get a measurement-enhanced unfolding network (MEUNet). Extensive experiments demonstrate that the proposed MEUNet achieves state-of-the-art (SOTA) performance in reconstruction accuracy with high efficiency.

Code — https://github.com/Gang-Qu/MEUNet-CS

Introduction

Sensing is essential to human perception and understanding of the physical world, yet obtaining high-throughput data that accurately capture optical signals remains a fundamental challenge in imaging. Compressive sensing (CS) addresses this challenge by leveraging the compressibility of natural signals. The core principle of CS lies in acquiring significantly fewer samples than that required by the Nyquist-Shannon sampling theorem through compressed measurements, while still preserving sufficient information for accurate signal reconstruction. This departures from conventional sampling paradigms has broad implications for various applications, including single-pixel cameras (Duarte et al. 2008), lensless imaging (Yuan and Pu 2018; Yuan et al. 2016), hyperspectral imaging (Yuan, Brady, and Katsaggelos 2021; Li et al. 2023; Qu, Wang, and Yuan 2023; Qin et al. 2025; Wang et al. 2025b), high-speed imaging (Wang et al. 2023; Wang, Wang, and Yuan 2023; Wang et al. 2024), information security (Qu et al. 2021b,a)and so on.

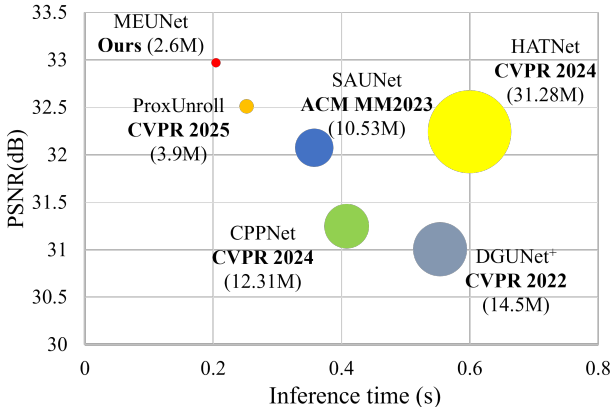

Figure 1: Overall Comparison of different methods.

CS theory has been widely studied in the past decades (Candès, Romberg, and Tao 2006; Donoho 2006; Duarte and Baraniuk 2011). The target 1D signal $\boldsymbol{x} \in \mathbb{R}^N$ can be linearly projected into a compressed measurement $\boldsymbol{y} \in \mathbb{R}^M$ at a sub-Nyquist sampling ratio $\frac{M}{N}$ ($M \ll N$): $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \epsilon$, where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ represents the sensing matrix, and ϵ denotes noise. The reconstruction of $\boldsymbol{x}$ from the compressed measurement $\boldsymbol{y}$ with $\boldsymbol{A}$ is ill-posed, which aims to solve the following optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \boldsymbol{g}(\boldsymbol{x}), \tag{1}$$

where $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ is the data fidelity term, and $\lambda \boldsymbol{g}(\boldsymbol{x})$ is the regularization term. For CS tasks, two important factors heavily affect the performance of reconstruction. In the sensing phase, how to capture the target more efficiently and effectively. In the reconstruction phase, designing advanced algorithms to further explore the correlations among measurement and sensing matrix matters. Deep unfolding networks (DUNs) have been proposed for CS reconstruction in the past few years, which not only enjoy both interpretability and excellent performance in quality and flexibility, but jointly optimized the sensing matrix for detection. However,

*Corresponding author.
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

the introduction of gradient projection in DUN inevitably leads to the decline in efficiency, especially using the large sensing matrix whose size increases exponentially with image scale.

Bearing these concerns in mind, Kronecker CS (KCS) has been adopted in DUN methods in the most recent works (Wang and Yuan 2023; Qu, Wang, and Yuan 2024), which aims to solve the problems of efficiency. KCS is a full-size image modulation method with no need to separate the image into patches, but uses the Kronecker product of two sub-matrices (image scale) to replace the conventional large sensing matrix (the square of image scale). Based on the properties of the Kronecker product, the Eq. (2) is equivalent to the vectorized form: $\mathrm{vec}\,(\boldsymbol{Y}) = (\boldsymbol{\Psi} \otimes \boldsymbol{\Phi})\,\mathrm{vec}\,(\boldsymbol{X})$, where vec$(\cdot)$ denotes the vectorization operation, and $\otimes$ is the Kronecker product. This kind of transformation indicates that the KCS can replace traditional CS in most situations when the large sensing matrix is not necessary to be predetermined, so we can use two sub-matrices to replace the large one. Thus, KCS solves the problems the block CS introduced, which maintains efficiency and is also applicable in real-world CS systems, e.g., SPI system (Qu, Wang, and Yuan 2024). KCS also provides high efficiency because the forward and inverse gradient projection process in DUN are also implemented at image level.

However, the drawbacks of applying KCS in DUN have not been discussed yet in previous works. The rigid structure of the Kronecker product imposes a strong coupling relation between sub-matrices, leading to the high mutual coherence and suboptimal restricted isometry property (RIP) constants. This structural prior, while computationally efficient, fundamentally limits the expressiveness of the learned sensing operator in DUN and, consequently, limits the reconstruction quality. Thus, there is a measurement barrier at the sensing phase when KCS is adopted.

To address this limitation, we propose asymmetric Kronecker CS (AKCS), a novel sensing paradigm that retains the computational efficiency of KCS while dramatically promoting its capability of expressiveness in DUN. AKCS conditions the column-wise sensing matrix on the choice of the row-wise sensing basis, effectively assigning a unique, learnable column operator for each row measurement, which breaks the restrictive symmetry of standard Kronecker CS. The good incoherent property introduced by AKCS inspires us to further explore the implicit prior of CS measurement, rather than simply explicit representations as a data fidelity term in DUN. Overall, our contributions are summarized as follows:

- We propose AKCS model, which breaks the limitations of standard Kronecker CS paradigm by introducing asymmetric row-adaptive sub-matrices. We theoretically analyze the coherence of the sensing matrices generated by the standard KCS and AKCS model, which demonstrates the superiority of AKCS in the capability of expressiveness in DUN.
- We demonstrate the compressive measurement is an important prior for CS reconstruction in DUN model. A measurement-aware cross-attention (MACA) module is specifically designed to capture and fuse the global feature prior embedded in measurement, leading to an efficient and powerful DUN model for CS reconstruction.
- The combination of AKCS model and MACA in DUN leads to MEUNet, which shows superiority in both efficiency and performance. Further experiments on simulated dataset and real data demonstrate the performance of the proposed method, indicating its potential to be applied in various real-world CS systems for efficient and high-quality reconstruction.

Related Works

Image CS Reconstruction

CS reconstruction methods can be classified into two categories: model-based methods (Kim, Nadar, and Bilgin 2010; Dong et al. 2014; Metzler, Maleki, and Baraniuk 2016) and learning-based methods (Zhang and Ghanem 2018; Shen et al. 2022; Wang and Yuan 2023; Zheng et al. 2024; Qu, Wang, and Yuan 2024; Wang et al. 2025a; Qu et al. 2025). The conventional model-based methods mainly rely on some hand-crafted priors to recover the original image from its sub-sampling measurement in an iterative manner, which enjoy high generalization and robustness, but meanwhile suffer from high computational cost and limited reconstruction quality. Earlier deep learning-bsed methods (Kulkarni et al. 2016; Metzler, Mousavi, and Baraniuk 2017; Qu et al. 2022) treat DNN as a black box and directly build a mapping from compressed measurement to the image. Recently, DUNs are proposed to incorporate DNN with conventional model-based methods, and train the unfolding network with multiple stages in an end-to-end manner, which enjoys good interpretability and has become the mainstream for CS reconstruction. Different optimization methods lead to different optimization-inspired DUNs, e.g., proximal gradient descent (PGD) algorithms (Chen and Zhang 2022; Song, Chen, and Zhang 2021), AMP (Zhu et al. 2020), ADMM (Wang et al. 2025a) and so on. Although DUN presents superiority compared to the end-to-end design, the introduction of gradient projection inevitably increases the computational burden, especially for large-scale image with its corresponding large sensing matrix.

Kronecker CS in DUN

KCS is a possible solution to solve the high computational complexity of gradient descent into DUN, which uses two independent sub-matrices to replace the large sensing matrix in vectorized CS and shows better performance due to the retaining of global information compared to block CS methods (Zhang and Ghanem 2018; Zhang et al. 2020; Mou, Wang, and Zhang 2022; Guo and Gan 2024a; Shen and Gan 2025). KCS-based DUN leads to SOTA performance for CS reconstruction and meanwhile maintains high efficiency, and most recent works have extend KCS-based model into real-world CS systems (Qu, Wang, and Yuan 2024; Wang et al. 2025a). However, the highly structured distribution of sensing matrix caused by the KCS actually restricts its performance in DUN, which has not been discussed yet in previous works.

Proposed Methods

Asymmetric Kronecker CS

Recall the forward model of standard KCS, which can be formulated as:

$$\boldsymbol{Y} = \boldsymbol{\Phi} \boldsymbol{X} \boldsymbol{\Psi}^{\top}, \tag{2}$$

where $\boldsymbol{Y} \in \mathbb{R}^{m \times n}$ is the 2D compressed measurement, $\boldsymbol{X} \in \mathbb{R}^{H \times W}$ is the 2D image, $\boldsymbol{\Phi} \in \mathbb{R}^{m \times H}$, and $\boldsymbol{\Psi} \in \mathbb{R}^{n \times W}$ are two independent sub-matrices. The compression ratio is defined as $\frac{mn}{HW}$. Standard KCS introduces a measurement barrier at the sensing phase due to the strong correlations induced by the Kronecker product. AKCS introduces asymmetric sub-matrices to decouple the highly structured distribution introduced by the Kronecker product of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ in KCS, the forward model of AKCS can be expressed as:

$$\boldsymbol{Y} = \mathcal{C}_{i=1:m}[\boldsymbol{\Phi}_i \boldsymbol{X} \boldsymbol{\Psi}_i^{\top}]. \tag{3}$$

where $\mathcal{C}$ denotes *concatenation operation* along row dimension, $\boldsymbol{\Phi}_i \in \mathbb{R}^{1 \times H}$ and $\boldsymbol{\Psi}_i \in \mathbb{R}^{n \times W}$ are row-wise matched matrices. For image $\boldsymbol{X} \in \mathbb{R}^{H \times W}$, the compressed measurement is $\boldsymbol{Y} \in \mathbb{R}^{m \times n}$. Thus, the compression ratio is also $\frac{mn}{HW}$, which is the same as KCS. In this way, the coupling relation of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ in standard KCS is broken, which means that the coherence of the generated large sensing matrix would largely drop and further provides more freedom in DUN for optimization. In the following part, we theoretically analyze the coherence to demonstrate the superiority of AKCS.

Coherence Analysis of KCS and AKCS. In CS, mutual coherence is an important factor to measure the expressive power of the sensing matrix, which is defined as the maximum absolute inner product of the normalized column vectors of the matrix, i.e., the absolute value of the maximum off-diagonal element of the Gram matrix, i.e., $\mu_A = \max_{i \neq j} | < a_i, a_j > |$, where a_i denotes the i-th column of A. Theoretically, the sensing matrix with lower coherence guarantees a better reconstruction quality with less number of measurement. Then, we will analyze the coherence of generated sensing matrices $\boldsymbol{A}_{\mathrm{K}}$ using KCS and $\boldsymbol{A}_{\mathrm{AK}}$ using AKCS model.

(i) The mutual coherence of $\boldsymbol{A}_{\mathrm{K}}$.

The Gram matrix of Kronecker product has a closed-form function. Considering the Kronecker product of $\boldsymbol{A}_{\mathrm{K}} = \boldsymbol{O} \otimes \boldsymbol{P}$, where $\boldsymbol{O} \in \mathbb{R}^{m \times H}$, and $\boldsymbol{P} \in \mathbb{R}^{n \times W}$. Its Gram matrix can be expressed as:

$$\begin{aligned} \boldsymbol{G}_{\mathrm{K}} &= \boldsymbol{A}_{\mathrm{K}}^{\top} \boldsymbol{A}_{\mathrm{K}} = (\boldsymbol{O} \otimes \boldsymbol{P})^{\top} (\boldsymbol{O} \otimes \boldsymbol{P}) \\ &= (\boldsymbol{O}^{\top} \boldsymbol{O}) \otimes (\boldsymbol{P}^{\top} \boldsymbol{P}) = \boldsymbol{D} \otimes \boldsymbol{B}, \end{aligned} \tag{4}$$

The elements of $\boldsymbol{G}_{\mathrm{K}}$ can be represented as $\boldsymbol{G}_{\mathrm{K}}[(i,k),(j,l)] = \boldsymbol{B}_{i,j} \boldsymbol{D}_{k,l}$, where the indexes are the positions of the corresponding vectorized form. Assume that the columns of $\boldsymbol{P}$ and $\boldsymbol{O}$ have been normalized, then $\boldsymbol{B}_{i,i} = 1$ and $\boldsymbol{D}_{i,i} = 1$. The coherence of $\boldsymbol{A}_{\mathrm{K}}$ is:

$$\begin{aligned} \mu_{\mathrm{K}} &= \max_{(i,k) \neq (j,l)} |\boldsymbol{B}_{i,j}, \boldsymbol{D}_{k,l}| \\ &= \max(\max_{i \neq j} |\boldsymbol{B}_{i,j}|, \max_{k \neq l} |\boldsymbol{D}_{k,l}|, \max_{i \neq j, k \neq l} |\boldsymbol{B}_{i,j} \boldsymbol{D}_{k,l}|). \end{aligned} \tag{5}$$

Because $|\boldsymbol{B}_{i,j}| \leq \mu_{\boldsymbol{P}} = \max_{i \neq j} |\boldsymbol{B}_{i,j}|$, and $|\boldsymbol{D}_{k,l}| \leq \mu_{\boldsymbol{O}} = \max_{k \neq l} |\boldsymbol{D}_{k,l}|$, and $|\boldsymbol{B}_{i,j} \boldsymbol{D}_{k,l}| \leq \mu_{\boldsymbol{P}} \mu_{\boldsymbol{O}}$, while $\mu_{\boldsymbol{P}} \mu_{\boldsymbol{O}} \leq \max(\mu_{\boldsymbol{P}}, \mu_{\boldsymbol{O}})$, thus

$$\mu_{\mathrm{K}} = \max(\mu_{\boldsymbol{P}}, \mu_{\boldsymbol{O}}), \tag{6}$$

where $\mu_{\boldsymbol{P}}$ and $\mu_{\boldsymbol{O}}$ represent the coherence of $\boldsymbol{P}$ and $\boldsymbol{O}$, respectively.

Assume the elements in $\boldsymbol{O}$ and $\boldsymbol{P}$ are independent identically distributed (i.i.d.), and follow Gaussian distribution $\sim \mathcal{N}(0,1)$, the coherence is approximately presented as:

$$\mu_{\boldsymbol{O}} = \sqrt{\frac{2 \log H}{n}}, \quad \mu_{\boldsymbol{P}} = \sqrt{\frac{2 \log W}{m}}, \tag{7}$$

where H and W denote the number of columns of $\boldsymbol{O}$ and $\boldsymbol{P}$, m and n are the number of rows.

(ii) The mutual coherence of $\boldsymbol{A}_{\mathrm{AK}}$.

Unlike KCS, $\boldsymbol{A}_{\mathrm{AK}}$ is generated by the asymmetric row-adaptive sub-matrices, whose columns can be represented as:

$$c_{i,j} = [a_{1,j} \boldsymbol{b}_{1,i}, a_{2,j} \boldsymbol{b}_{2,i}, ...]^{\top}, \tag{8}$$

where $a_{m,j}$ denotes the elements in $\boldsymbol{\Phi}$, $\boldsymbol{b}_{m,i}$ is the i-th column of $\boldsymbol{\Psi}$. Thus, the Gram matrix can be expressed as:

$$\boldsymbol{G}_{\mathrm{AK}}[(i,j),(k,l)] = < c_{i,j}, c_{k,l} > = \sum_{r=1}^{m} a_{r,j} a_{r,l} < \boldsymbol{b}_{r,i}, \boldsymbol{b}_{r,k} > \tag{9}$$

The coherence of $\boldsymbol{A}_{\mathrm{AK}}$ can be obtained:

$$\mu_{\mathrm{AK}} = \max_{(i,j) \neq (k,l)} \frac{|\sum_{r=1}^{m} a_{r,j} a_{r,l} < \boldsymbol{b}_{r,i}, \boldsymbol{b}_{r,k} > |}{\sqrt{\sum_m a_{m,j}^2 ||\boldsymbol{b}_{m,j}||_2^2} \sqrt{\sum_m a_{m,l}^2 ||\boldsymbol{b}_{m,k}||_2^2}}. \tag{10}$$

The situation of μ_{AK} is more complicated than μ_A, we directly present conclusion here, summarized as **Theorem 1**.

Theorem 1. Assume a sensing matrix

$$A_{\mathrm{AK}} = [a_1 \otimes B_1, ..., a_m \otimes B_m]^{\top}, \tag{11}$$

where $a_i \in \mathbb{R}^{1 \times H}$ follows i.i.d., $\sim \mathcal{N}(0,1)$, $B_i \in \mathbb{R}^{n \times W}$ follows i.i.d., $\sim \mathcal{N}(0,1)$, there exists an absolute constant $C_o > 0$ ($C_o \approx 1$ for Gaussian distribution), such that with high probability:

$$\mu_{\mathrm{AK}} \leq C_o \sqrt{\frac{2 \log(HW)}{mn}}. \tag{12}$$

The theoretical demonstration of **Theorem 1** is presented in Supplementary material $\mathcal{S}1$. Thus, we get the upper bounds of μ_{K} and μ_{AK}, and obviously, $\mu_{\mathrm{AK}} < \mu_K$. This means the coherence of sensing matrix generated by AKCS is much less than that generated by KCS, which indicates that the AKCS model is potential to promote the CS reconstruction in DUN.

AKCS Model in DUN. Bearing in mind the theoretical analysis above, we adopt AKCS in DUN model, and consider the optimization problem of Eq. (3), which can be rewritten as:

$$\hat{\boldsymbol{X}} = \mathrm{argmin}_{\boldsymbol{X}} \ \frac{1}{2} \left\| \boldsymbol{Y} - \mathcal{C}_{i=1:m}[\boldsymbol{\Phi}_i \boldsymbol{X} \boldsymbol{\Psi}_i^{\top}] \right\|_F^2 + \lambda \boldsymbol{g}(\boldsymbol{X}), \tag{13}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Further developing ISTA algorithms to solve (13), then we have:

$$\boldsymbol{U}_k = \boldsymbol{X}_{k-1} + \rho \textstyle\sum_{i=1:m} [\boldsymbol{\Phi}_i^{\top} (\boldsymbol{Y}_i - \boldsymbol{\Phi}_i \boldsymbol{X}_{k-1} \boldsymbol{\Psi}_i^{\top}) \boldsymbol{\Psi}_i], \tag{14}$$

$$\boldsymbol{X}_k = \mathrm{argmin}_{\boldsymbol{X}} \ \tfrac{1}{2\sigma^2} ||\boldsymbol{U}_k - \boldsymbol{X}||_F^2 + \boldsymbol{g}(\boldsymbol{X}), \tag{15}$$

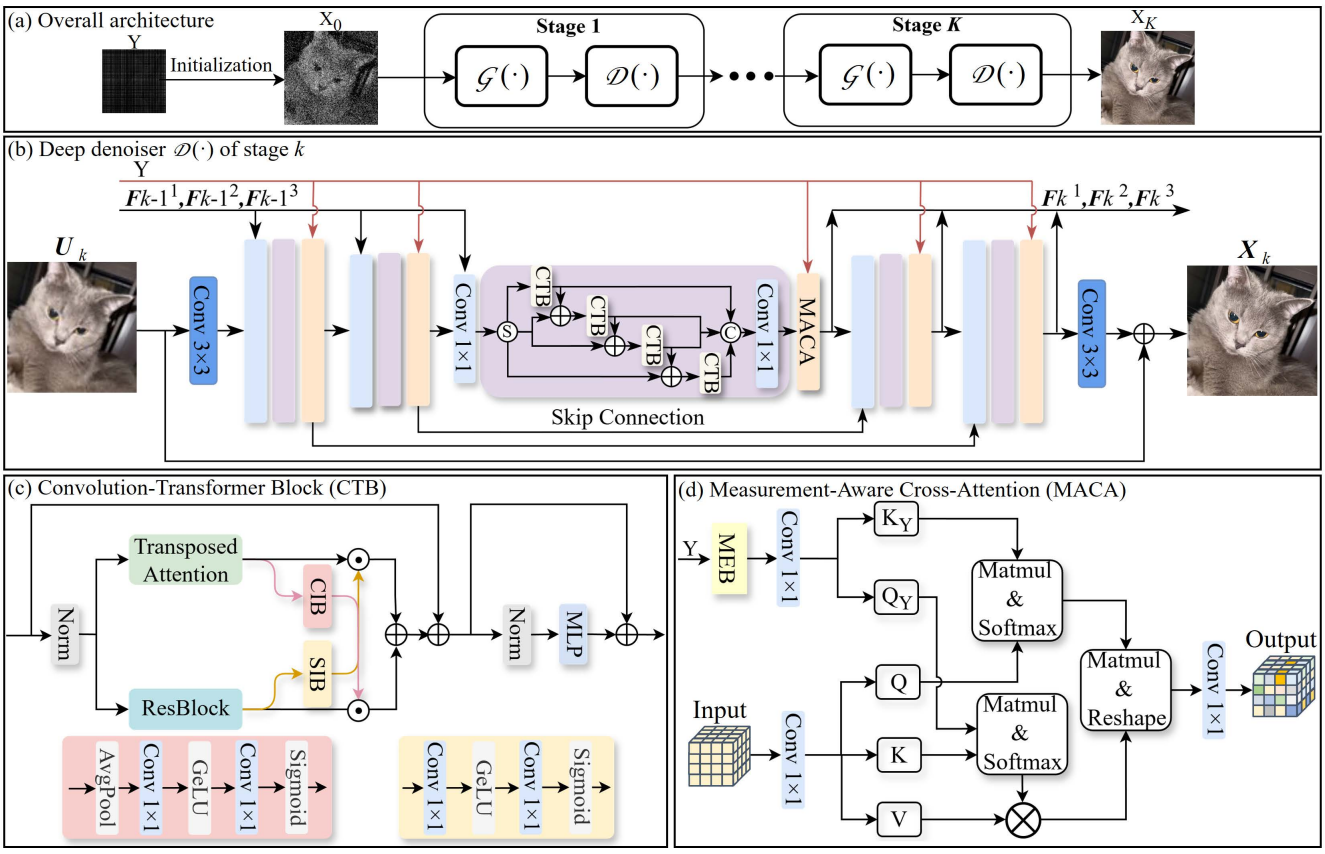

Figure 2: The pipeline of the proposed method. (a) The overall architecture of the proposed DUN model, containing K stages. Each stage includes proximal gradient descent projection of Eq. (16) and trainable deep deniser of Eq. (17). (b) The design of deep denoiser in k-th stage. The denoisers in all stages are sharing the same parameters. (c) The convolution-Transformer block (CTB) in deep denoiser. (d) The design of measurement-aware cross-attention (MACA) block.

where $\boldsymbol{Y}_i$ denotes the i-th row of 2D measurement $\boldsymbol{Y}$. Eq. (14) denotes the gradient descent from the current stage input $\boldsymbol{X}_{k-1}$ with a step size ρ, $\boldsymbol{U}_k$ is an auxiliary variable. Eq. (15) denotes proximal mapping process, which can be seen as a denoising problem under the noise level σ. In DUN, this denoising problem is conducted by a trainable deep neural network, which has shown its superiority compared to those hand-designed denoisers. When integrating the above ISTA algorithm with a specific denoising network, a AKCS-DUN model can be established:

$$\boldsymbol{U}_k = \boldsymbol{X}_{k-1} + \rho_{k-1} \sum\nolimits_{i=1:m} [\boldsymbol{\Phi}_i^\top (\boldsymbol{Y}_i - \boldsymbol{\Phi}_i \boldsymbol{X}_{k-1} \boldsymbol{\Psi}_i^\top) \boldsymbol{\Psi}_i], \tag{16}$$

$$\boldsymbol{X}_k = \mathcal{D}_{(k,\theta)}(\boldsymbol{U}_k), \tag{17}$$

where ρ_{k-1} is a learnable parameter that controls the step size in each stage, and $\mathcal{D}_{(k,\theta)}$ represents the learnable deep denoiser.

DUN with Measurement-Aware Cross-Attention

In the deep denoising phase, we aim to learn informative implicit representations from previous output $\boldsymbol{U}_{\mathrm{K}}$ and the measurement $\boldsymbol{Y}$. Next, the overall architecture of the proposed deep denoiser is first given and then two core modules, hybrid Convolution-Transformer block (CTB) and measurement-aware cross-attention (MACA) block are introduced in detail.

Overall Architecture. As shown in Fig. 2 (b), our deep denoiser $\mathcal{D}$ is a symmetric UNet architecture for multi-scale representation learning, composed of two encoder layer, a bottleneck layer, and two decoder layer. The input of deep denoiser is the result of the gradient descent projection $\boldsymbol{U}_k$, which then passes through encoders to generate the deep features. In each layer of encoder, the features of current layer are concatenated with features from previous stage and fused by channel concatenation and 1×1 convolution. To reduce the computational complexity, we separate the input features into four groups along channel dimension and adopt the residual connection for interactions among them. The encoder in each layer mainly contains CTB, MACA, and down-sampling layer. The goal of encoders is to progressively reduce the spatial resolution by half and double the channel dimensions, yielding the multi-scale features transferred to the decoder by skip connections. In the decoder branch, the same design of encoder is adopted, with simple 1×1 convolution and pixel shuffle operations for inner-stage

feature fusion and up-sampling. The output of each decoder $\boldsymbol{F}_{\mathrm{K}}$ is the feature map of the current stage and are then transferred to the next stage. Finally, the feature map from the last decoder is converted to the output $\boldsymbol{X}_{\mathrm{K}} \in \mathbb{R}^{N_H \times N_W}$ of current stage with a 3×3 convolution.

Convolution-Transformer Block. The mixture modules of convolution and Transformer is a common design to maintain the local and non-local modeling capability of deep neural network. Considering the efficiency and the following design of cross-attention in spatial dimension, we adopt the parallel blocks of convolution and transposed attention with cross interactions. As depicted in Fig. 2(c), the CTB mainly contains Residual Block (ResBlock), transposed attention (Zamir et al. 2022), and interaction blocks. ResBlock contains only two 3×3 convolution with LeakyReLU and residual connections inside, which captures local interactions effectively. The output of Resblock $\boldsymbol{F}_{\mathrm{R_{out}}} \in \mathbb{R}^{N_H \times N_W \times N_C}$ is just the same size with input $\boldsymbol{F}_{\mathrm{in}}$. As for transposed attention, considering the same input $\boldsymbol{F}_{\mathrm{in}} \in \mathbb{R}^{N_H \times N_W \times N_C}$, it is first reshaped to $\boldsymbol{F}_{\mathrm{TA_{in}}} \in \mathbb{R}^{N_H N_W \times N_C}$. Then the Query ($\boldsymbol{Q}$), Key ($\boldsymbol{K}$), and Value ($\boldsymbol{V}$) can be obtained by linear transform:

$$\boldsymbol{Q} = \boldsymbol{F}_{\mathrm{TA_{in}}}\boldsymbol{W}^q, \boldsymbol{K} = \boldsymbol{F}_{\mathrm{TA_{in}}}\boldsymbol{W}^k, \boldsymbol{V} = \boldsymbol{F}_{\mathrm{TA_{in}}}\boldsymbol{W}^v, \quad (18)$$

where $\boldsymbol{W}^{(\cdot)} \in \mathbb{R}^{N_C \times N_C}$ denotes the learnable linear projection. Then the multi-head attention mechanism is performed:

$$\boldsymbol{F}_i = \boldsymbol{A}_i * \boldsymbol{V}_i = \mathrm{softmax}(\frac{\boldsymbol{Q}_i \boldsymbol{K}_i^\top}{\sqrt{d}}) * \boldsymbol{V}_i, \quad (19)$$

$i = 1, \ldots, N$ is the number of heads, d is a learnable scaling parameter that controls the magnitude of the product.

CTB further introduces a residual high-frequency enhancement branch. The core intuition is using 'squeeze' operation to aggregate the feature map along channel dimension, which means the global channel average and represents the low-frequency information (dominant component of each feature channel). The difference between the original feature and its low-frequency component constitutes the high-frequency residual. By explicitly isolating and then adaptively amplifying this residual, our module forces the network to pay specific attention to and preserve these high-frequency details, which are often lost in standard attention mechanisms:

$$\boldsymbol{F}_{\mathrm{TA_{out}}} = \boldsymbol{F}_i + \gamma(\boldsymbol{F}_i - \phi(S(\boldsymbol{F}_i))), \quad (20)$$

where ϕ represents GELU activation function, S denotes 'squeeze' operation to squeeze channel dimension from C to 1, γ is a learnable scaling factor. After a linear transform, we obtain the final output of transposed attention block $\boldsymbol{F}_{\mathrm{TA_{out}}} \in \mathbb{R}^{N_H \times N_W \times N_C}$. Then, spatial and channel-wise interactions are also introduced to fuse the deep features among spatial and channel dimensions, which can be expressed as:

$$\begin{aligned} \boldsymbol{F}_1 &= \boldsymbol{F}_{\mathrm{S_{out}}} \cdot \mathrm{Sigmoid}(\boldsymbol{F}_{\mathrm{TA_{out}}}), \\ \boldsymbol{F}_2 &= \boldsymbol{F}_{\mathrm{TA_{out}}} \cdot \mathrm{Sigmoid}(\boldsymbol{F}_{\mathrm{S_{out}}}), \end{aligned} \quad (21)$$

The features after interactions are then concatenated for feature fusion along with a residual input to get the output:

$$\boldsymbol{F}_{\mathrm{out}} = \mathrm{Concat}(\boldsymbol{F}_1, \boldsymbol{F}_2) + \boldsymbol{F}_{in}, \quad (22)$$

Measurement-Aware Cross-Attention Block. The compressive measurement $\boldsymbol{Y}$ in image CS task is actually a global feature extracted from the ground truth with the sensing matrix, and the masks actually serve as global convolution kernels. Bearing this perspective in mind, we specifically design a measurement-aware cross-attention (MACA) block to further explore the implicit prior embedded in measurement and adopt it to guide the reconstruction in DUN. In MACA, we treat the high-level information encoded from the measurements $\boldsymbol{Y}$ as a set of Queries and keys. They are then used to establish an information bridge from the measurement domain to the image feature domain. To strike a balance between prohibitive computational cost and the need for precise information flow, MACA employs a highly efficient 'Aggregate-in-Low, Propagate-in-High' strategy. This ensures that global guidance is both potent and computationally tractable, even when operating on high-resolution feature maps. The proposed MACA block is illustrated in Fig. 2 (d). The computational flow is organized into two stages: (1) Forward Attention for Information Aggregation, and (2) Backward Attention for Information Propagation. Considering the input feature map $\boldsymbol{F}_{in} \in \mathbb{R}^{N_H \times N_W \times N_C}$ and the compressed measurement $\boldsymbol{Y} \in \mathbb{R}^{N_h \times N_w}$, the primary goal of stage (1) is to efficiently summarize the rich spatial information in $\boldsymbol{F}_{in}$ under the guidance of the measurement-derived queries. The measurement $\boldsymbol{Y}$ is first processed by a lightweight network, Measurement Encoder Block (MEB), to produce a sequence of query vectors $\boldsymbol{Q}_Y \in \mathbb{R}^{N_Y \times N_d}$ and key vectors $\boldsymbol{K}_Y \in \mathbb{R}^{N_Y \times N_d}$. This step transforms the raw, low-level measurements into a set of high-level, semantic 'questions' and 'keys' about the image content. To circumvent the quadratic complexity of attention on high-resolution feature maps, we then down-sample the input feature map $\boldsymbol{F}_{\mathrm{in}} \in \mathbb{R}^{N_H \times N_W \times N_C}$ by a factor of D using average pooling, resulting in $\boldsymbol{F}_{\mathrm{D}} \in \mathbb{R}^{\frac{N_H}{D} \times \frac{N_W}{D} \times N_d}$. This drastically reduces the sequence length of the keys and values from $N_H N_W$ to $\frac{N_H N_W}{D^2}$, directly tackling the primary computational and memory bottleneck. Then we compute the low-resolution attention scores between the measurement queries $\boldsymbol{Q}_{\mathrm{Y}}$ and the keys $\boldsymbol{K}_{\mathrm{D}}$ derived from the down-sampled features, which is then used to weigh the values $\boldsymbol{V}_{\mathrm{D}}$:

$$\boldsymbol{V}_{\mathrm{YD}} = \mathrm{softmax}(\frac{\boldsymbol{Q}_{\mathrm{Y}} \boldsymbol{K}_{\mathrm{D}}^\top}{\sqrt{d_{\mathrm{K}}}}) * \boldsymbol{V}_{\mathrm{D}}. \quad (23)$$

Here, $\boldsymbol{K}_{\mathrm{D}}$, $\boldsymbol{V}_{\mathrm{D}}$ are linearly projected from $\boldsymbol{F}_{\mathrm{D}}$. The output, $\boldsymbol{V}_{\mathrm{YD}} \in \mathbb{R}^{N_{\mathrm{Y}} \times N_{\mathrm{d}}}$, represents a compact summary of the image features, where each element is an 'answer' to the corresponding query from $\boldsymbol{Y}$.

Having aggregated the essential image information into $\boldsymbol{V}_{\mathrm{YD}}$, the stage (2) precisely propagates this guidance back to each pixel location in the original full-resolution feature map. First, We project the original, full-resolution feature map $\boldsymbol{F}_{\mathrm{in}} \in \mathbb{R}^{N_H \times N_W \times N_C}$ to generate a new set of Queries vectors $\boldsymbol{Q}_{\mathrm{H}} \in \mathbb{R}^{N_H N_W \times N_d}$. Then, a backward attention map is computed by high-resolution queries and keys map from measurement:

$$\boldsymbol{A}_{\mathrm{HY}} = \mathrm{softmax}(\frac{\boldsymbol{Q}_{\mathrm{H}} \boldsymbol{K}_Y^\top}{\sqrt{d_{\mathrm{K}}}}). \quad (24)$$

Dataset	Method	Sampling Ratio (SR)			
		4%	10%	25%	50%
Set11	DGUNet$^+$ (CVPR 2022)	26.82/0.8230	30.93/0.9088	36.18/0.9616	41.24/0.9837
	OCTUF$^+$ (CVPR 2023)	26.54/0.8150	30.73/0.9036	36.10/0.9607	41.35/0.9838
	SAUNet (ACM MM 2023))	27.80/0.8353	32.15/0.9147	37.11/0.9628	41.91/0.9838
	CPPNet (CVPR 2024))	27.23/0.8337	31.27/0.9135	36.35/0.9631	-
	HATNet (CVPR 2024))	27.98/0.8382	32.26/0.9182	37.24/0.9634	42.05/0.9838
	ProxUnroll (CVPR 2025))	28.30/0.8452	32.55/0.9226	37.35/0.9639	41.97/0.9838
	MEUNet (ours)	**28.49/0.8484**	**32.90/0.9270**	**37.64/0.9648**	**42.39/0.9842**
CBSD68	DGUNet$^+$ (CVPR 2022)	25.45/0.6986	28.13/0.8165	31.97/0.9158	37.04/0.9718
	OCTUF$^+$ (CVPR 2023)	25.65/0.6999	28.28/0.8177	32.24/0.9185	37.41/0.9729
	SAUNet (ACM MM 2023)	26.23/0.7050	29.25/0.8251	33.67/0.9243	39.28/0.9751
	HUNet (CVPR 2025))	25.90/0.7101	28.60/0.8265	32.67/0.9229	-
	ProxUnroll (CVPR 2025))	26.54/0.7200	29.43/0.8334	33.77/0.9284	39.23/0.9761
	MEUNet (ours)	**26.64/0.7218**	**29.56/0.8346**	**33.92/0.9296**	**39.65/0.9767**

Table 1: Average PSNR/SSIM of different methods on Set11 datasets with different SRs. The best and second best results are highlighted in **bold** and underlined, respectively.

Figure 3: The comparison of visualization results (Barbara in Set 11) at SR $= 10\%$ of different methods.

This step can be intuitively understood as each pixel location in the feature map is looking for the most relevant piece of information for restoration with given content. The backward attention map is then used to weight and aggregate the summarized features from $\boldsymbol{V}_{\mathrm{YD}}$, producing the refined feature sequence:

$$\boldsymbol{F}_{\mathrm{MACA}} = \boldsymbol{A}_{\mathrm{HY}} * \boldsymbol{V}_{\mathrm{YD}}. \tag{25}$$

The resulting sequence $\boldsymbol{F}_{\mathrm{MACA}} \in \mathbb{R}^{N_H N_W \times N_d}$ is then reshaped back to the input feature dimensions. This backward pass ensures that the global information contained in $\boldsymbol{V}_{\mathrm{YD}}$ is distributed precisely and adaptively across the entire spatial domain of the feature map, rather than being broadcast uniformly. It allows for a fine-grained, pixel-level modulation based on global measurement constraints.

Results and Analysis

Following previous works (Wang and Yuan 2023; Qu, Wang, and Yuan 2024), 400 images from BSD500 (Arbeláez et al. 2011) are employed as the training dataset. Data augmentation operations, including random horizontal flipping, scaling, and cropping, are performed to generate 20,000 images as the training dataset. All models are trained through 200 epochs with learning rate 1×10^{-3} and then fine-tuned through 20 epochs with learning rate 1×10^{-4} using Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The sensing matrices in AKCS model are set to be learnable for fair comparison in simulation. For testing on synthetic data, we evaluate the proposed method with different sampling ratios (SRs) $\{4\%, 10\%, 25\%, 50\%\}$ on commonly used Set11 and CBSD68 dataset. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are used to estimate the performance in our experiments. For real data, we build a real SPI system to verify the effectiveness of our method. The learned sensing matrices are used in real experiment.

Results of Simulation Data

We make a comprehensive comparison with previous methods to evaluate the performance of proposed MEUNet, including DGUNet$^+$ (Mou, Wang, and Zhang 2022), OCTUF$^+$ (Song et al. 2023), SAUNet (Wang and Yuan 2023), CPPNet (Guo and Gan 2024b), HATNet (Qu, Wang, and Yuan 2024), HUNet (Shen and Gan 2025), and ProxUnroll (Wang et al. 2025a). SAUNet, HATNet and ProxUnroll are the most recent works that adopt KCS model in DUN. Tab. 1 presents the average PSNR/SSIM of different methods (some unavailable data are not presented). Clearly, the proposed MEUNet outperforms previous methods at all SRs. Fig. (3) presents the visualization comparison on the reconstruction results of our MEUNet and previous competitive methods. The proposed method showcases the best results and also improvement in details and textures, as highlighted in the zoom-in regions.

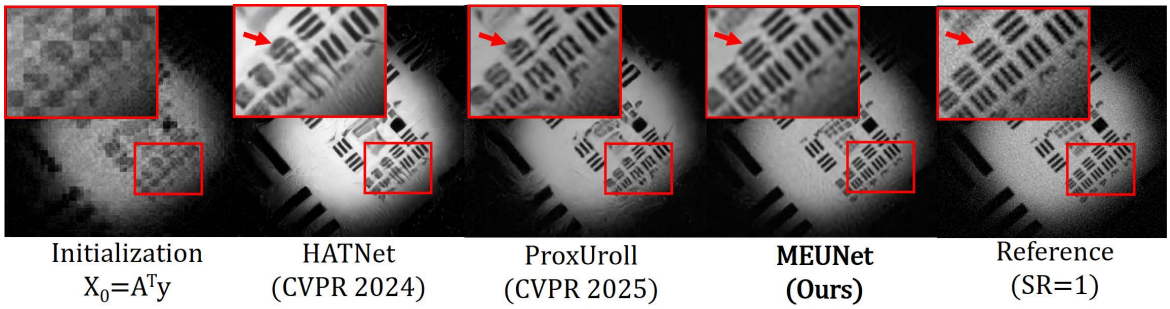

Figure 4: The comparison of real data at $\mathrm{SR} = 10\%$ of different methods.

Results of Real Data

We further demonstrate the performance of the proposed method on real data. We use our SPI system to capture real data of a resolution target with 256×256 pixels, which are then reconstructed by HATNet, ProxUnroll and our proposed MEUNet, respectively. The results are shown in Fig. 4. The reference image is reconstructed with SR=1, which can be formulated as $\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{y}$ s.t. $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is an orthogonal Hadamard matrix. In theory, full-sampled image using Hadamard matrix is lossless because of its orthogonality, thus the full-sampled image can also be regarded as a reference for comparison. Compared to the other methods, our proposed MEUNet showcases the best performance on resolution target reconstruction, which demonstrates that the proposed method is also applicable in real-world CS system.

Model	AKCS	ResBlock	TA	MACA	PSNR (dB)	SSIM
(a)	✓	✓	✓	✓	**32.90**	**0.9270**
(b)		✓	✓	✓	32.35	0.9194
(c)	✓		✓	✓	32.63	0.9239
(d)	✓	✓		✓	32.42	0.9205
(e)	✓	✓	✓		32.79	0.9262

Table 2: Ablation study of the proposed MEUNet.

Ablation Study

Different Components of MEUNet. To quantitatively analyze the effect of different components, we perform ablation study on Set11 dataset at SR = 10%. The proposed MEUNet is mainly powered by the following designs: AKCS model, ResBlock, transposed attention (TA), MACA. The average PSNR and SSIM are shown in Tab.2. Baseline model (a) contains all the components and presents the best result of 32.90 dB/0.9270. Model (b) without AKCS, there is an average 0.55 dB /0.0076 degradation on PSNR and SSIM. The absence of AKCS model in MEUNet leads to the most severe degradation in reconstruction, which demonstrates the importance of incoherent sensing in CS and the effectiveness of the proposed AKCS model. Towards model (c) without ResBlock in convolution-Transformer block, there is an average 0.27 dB/0.0031 degradation on PSNR and SSIM. Towards model (d) without TA in convolution-Transformer block, there is an average 0.48 dB/0.0065 degradation on PSNR and SSIM. Towards model (e) without MACA, there is an average 0.11 dB/0.0008 degradation on PSNR and SSIM, meaning that the proposed MACA, as a plug-and-play part in DUN, is effective for CS reconstruction.

Comparison of Computational Complexity. Then we further make a comparison on the average PSNR, the number of parameters, FLOPs, and inference time on Set 11 dataset (image size is 256×256) of different methods, as presented in Tab. 3. The results also demonstrate the superiority of the proposed MEUNet compared to previous methods in terms of performance and efficiency.

Methods	SAUNet	HATNet	ProxUnroll	MEUNet
PSNR (dB)	32.15	32.26	<u>32.55</u>	**32.90**
FLOPs (G)	<u>143.05</u>	494.42	**107.73**	145.38
Params (M)	10.53	31.28	<u>3.90</u>	**2.60**
InferenceTime (s)	0.35	0.60	<u>0.27</u>	**0.21**

Table 3: Comparisons on average PSNR, parameters, FLOPs, and inference time on Set11 dataset of different methods at SR=10% (tested on NVIDIA 5090 GPU).

Conclusion

In this paper, we aim to break the measurement barriers in both sensing and reconstruction phases. Considering the drawback introduced by KCS, we propose AKCS model in DUN, which breaks the limitation of highly structured distribution introduced by the Kronecker product, thus leveraging the capability of expressiveness of DUN and significantly promoting the reconstruction quality with almost no increase of computational complexity. In addition, for the first time, we explore the implicit prior embedded in the measurement using a plug-and-play measurement-aware cross-attention module in DUN and demonstrate its effectiveness in promoting the reconstruction accuracy. The proposed AKCS model and MACA are combined with efficient Convolution-Transformer design in DUN leads to both SOTA performance and high efficiency compared to previous works. Furthermore, we also demonstrate the feasibility of the proposed method in real-world CS system, especially the AKCS model, would be a potential solution to further optimize the modulation patterns for CS detection and promote the application of deep-learning models in real imaging systems.

Acknowledgements

This work was supported by the National Key R&D Program of China under Grant 2024YFF0505603, National Natural Science Foundation of China under Grant 62271414, Zhejiang Provincial Science Fund for Distinguished Young Scholar Project under Grant LR23F010001, Zhejiang "Pioneer" and "Leading Goose" R&D Program under Grant 2024SDXHDX0006 and 2024C03182, the Key Project of Westlake Institute for Optoelectronics under Grant 2023GD007, and Ningbo Science and Technology Bureau, "Science and Technology Yongjiang 2035" Key Technology Breakthrough Program under Grant 2024Z126.

References

Arbeláez, P.; Maire, M.; Fowlkes, C.; and Malik, J. 2011. Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5): 898–916.

Candès, 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.

Dong, W.; Shi, G.; Li, X.; Ma, Y.; and Huang, F. 2014. Compressive sensing via nonlocal low-rank regularization. *IEEE Transactions on Image Processing*, 23(8): 3618–3632.

Donoho, D. L. 2006. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4): 1289–1306.

Duarte, M. F.; and Baraniuk, R. G. 2011. Kronecker compressive sensing. *IEEE Transactions on Image Processing*, 21(2): 494–504.

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.

Guo, Z.; and Gan, H. 2024a. 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*, 25086–25095.

Guo, Z.; and Gan, H. 2024b. CPP-Net: Embracing Multi-Scale Feature Fusion into Deep Unfolding CP-PPA Network for Compressive Sensing. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25086–25095.

Kim, Y.; Nadar, M. S.; and Bilgin, A. 2010. Compressed sensing using a Gaussian scale mixtures model in wavelet domain. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*.

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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 449–458.

Li, Z.; Qu, G.; Suo, J.; and Yuan, X. 2023. Deep-learning enables single-pixel spectral imaging. In *Optoelectronic Imaging and Multimedia Technology IX*, volume 12317, 75–85. SPIE.

Metzler, C.; Mousavi, A.; and Baraniuk, R. 2017. Learned D-AMP: Principled Neural Network based Compressive Image Recovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30.

Metzler, C. A.; Maleki, A.; and Baraniuk, R. G. 2016. From denoising to compressed sensing. *IEEE Transactions on Information Theory*, 62(9): 5117–5144.

Mou, C.; Wang, Q.; and Zhang, J. 2022. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17399–17410.

Qin, M.; Feng, Y.; Wu, Z.; Zhang, Y.; and Yuan, X. 2025. Detail Matters: Mamba-Inspired Joint Unfolding Network for Snapshot Spectral Compressive Imaging. *arXiv preprint arXiv:2501.01262*.

Qu, G.; Meng, X.; Yang, X.; Wu, H.; Wang, P.; He, W.; and Chen, H. 2021a. Optical color watermarking based on single-pixel imaging and singular value decomposition in invariant wavelet domain. *Optics and Lasers in Engineering*, 137: 106376.

Qu, G.; Meng, X.; Yin, Y.; Wu, H.; Yang, X.; Peng, X.; and He, W. 2021b. Optical color image encryption based on Hadamard single-pixel imaging and Arnold transformation. *Optics and Lasers in Engineering*, 137: 106392.

Qu, G.; Meng, X.; Yin, Y.; and Yang, X. 2022. A demosaicing method for compressive color single-pixel imaging based on a generative adversarial network. *Optics and Lasers in Engineering*, 155: 107053.

Qu, G.; Wang, P.; and Yuan, X. 2024. Dual-Scale Transformer for Large-Scale Single-Pixel Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 25327–25337.

Qu, G.; Wang, X.; and Yuan, X. 2023. Plug-and-Play Deep Image Prior for Snapshot Optical Coherence Tomography. In *Propagation Through and Characterization of Atmospheric and Oceanic Phenomena*, JW2A–10. Optica Publishing Group.

Qu, G.; Zheng, S.; Qin, M.; and Yuan, X. 2025. BMVC+: An Enhanced Block Modulation Video Compression Codec for Large-scale Image Compression. *IEEE Journal of Selected Topics in Signal Processing*, 1–12.

Shen, F.; and Gan, H. 2025. HUNet: Homotopy Unfolding Network for Image Compressive Sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12799–12808.

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.

Song, J.; Chen, B.; and Zhang, J. 2021. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the 29th ACM international conference on multimedia*, 4249–4258.

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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6174–6184.

Wang, P.; Wang, L.; Qiao, M.; and Yuan, X. 2023. Full-resolution and full-dynamic-range coded aperture compressive temporal imaging. *Optics Letters*, 48(18): 4813–4816.

Wang, P.; Wang, L.; Qu, G.; Wang, X.; Zhang, Y.; and Yuan, X. 2025a. Proximal Algorithm Unrolling: Flexible and Efficient Reconstruction Networks for Single-Pixel Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 411–421.

Wang, P.; Wang, L.; and Yuan, X. 2023. Deep Optics for Video Snapshot Compressive Imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10646–10656.

Wang, P.; and Yuan, X. 2023. SAUNet: Spatial-Attention Unfolding Network for Image Compressive Sensing. In *Proceedings of the 31st ACM International Conference on Multimedia*, 5099–5108.

Wang, P.; Zhang, Y.; Wang, L.; and Yuan, X. 2024. Hierarchical separable video transformer for snapshot compressive imaging. In *European Conference on Computer Vision*, 104–122. Springer.

Wang, X.; He, Z.; Wang, P.; Wang, L.; Hu, Y.; and Yuan, X. 2025b. Spectral Compressive Imaging via Chromaticity-Intensity Decomposition. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Yuan, X.; Brady, D. J.; and Katsaggelos, A. K. 2021. Snapshot Compressive Imaging: Theory, Algorithms, and Applications. *IEEE Signal Processing Magazine*, 38(2): 65–88.

Yuan, X.; Jiang, H.; Huang, G.; and Wilford, P. A. 2016. SLOPE: Shrinkage of Local Overlapping Patches Estimator for Lensless Compressive Imaging. *IEEE Sensors Journal*, 16(22): 8091–8102.

Yuan, X.; and Pu, Y. 2018. Parallel lensless compressive imaging via deep convolutional neural networks. *Optics Express*, 26(2): 1962–1977.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

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

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, S.; Xue, Y.; Tahir, W.; Wang, Z.; Zhang, H.; Meng, Z.; Qu, G.; Ma, S.; and Yuan, X. 2024. Block-modulating video compression: an ultralow complexity image compression encoder for resource-limited platforms. *Advanced Imaging*, 1(2): 021002.

Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; and Dai, J. 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.