

Content-aware Scalable Deep Compressed Sensing

Bin Chen, Jian Zhang, *Member, IEEE*

Abstract—To more efficiently address image compressed sensing (CS) problems, we present a novel content-aware scalable network dubbed CASNet which collectively achieves adaptive sampling rate allocation, fine granular scalability and high-quality reconstruction. We first adopt a data-driven saliency detector to evaluate the importances of different image regions and propose a saliency-based block ratio aggregation (BRA) strategy for sampling rate allocation. A unified learnable generating matrix is then developed to produce sampling matrix of any CS ratio with an ordered structure. Being equipped with the optimization-inspired recovery subnet guided by saliency information and a multi-block training scheme preventing blocking artifacts, CASNet jointly reconstructs the image blocks sampled at various sampling rates with one single model. To accelerate training convergence and improve network robustness, we propose an SVD-based initialization scheme and a random transformation enhancement (RTE) strategy, which are extensible without introducing extra parameters. All the CASNet components can be combined and learned end-to-end. We further provide a four-stage implementation for evaluation and practical deployments. Experiments demonstrate that CASNet outperforms other CS networks by a large margin, validating the collaboration and mutual supports among its components and strategies. Codes are available at <https://github.com/Guaishou74851/CASNet>.

Index Terms—Compressed sensing, image restoration, content-aware sampling, model scalability, deep unfolding network.

I. INTRODUCTION

COMPRESSED sensing (CS) is a novel paradigm that requires much fewer measurements than the Nyquist sampling for signal acquisition and restoration [1], [2]. For the signal $\mathbf{x} \in \mathbb{R}^N$, it conducts the sampling process $\mathbf{y} = \Phi \mathbf{x}$ to obtain the measurements $\mathbf{y} \in \mathbb{R}^M$, where $\Phi \in \mathbb{R}^{M \times N}$ with $M \ll N$ is a given sampling matrix, and the CS ratio (or sampling rate) is defined as $r = M/N$. Since it is hardware-friendly and has great potentials of improving sampling speed with high recovery accuracy, many applications have been developed including single-pixel imaging [3], [4], magnetic resonance imaging (MRI) [5], [6], sparse-view CT [7], etc. In this work, we focus on the typical block-based (or block-diagonal) image CS problem [8]–[10] that divides the high-dimensional natural image into non-overlapped $B \times B$ blocks and obtains measurements block-by-block with a small fixed sampling matrix for the subsequent reconstruction.

Recovering \mathbf{x} from the acquired \mathbf{y} is to solve an ill-posed under-determined inverse system. Many model-driven methods focus on exploiting structural prior with theoretical guarantees, such as sparse representation [11]–[15], low-rank [16]–[20], etc. However, they are suffering from high computational costs and rely on extensive fine-tuning and empirical results.

Recently, the development of deep convolutional neural network (CNN) greatly improves recovery accuracy and speed [21]. By adopting end-to-end pipelines, [22]–[24] can quickly perform recoveries but may leave undesired effects [25]. Motivated by the classic structure-texture decomposition paradigm, Sun *et al.* [26] propose a dual-path attention CS network. Deep unfolding methods [27]–[33] map restoration algorithms into network architectures and achieve balances between speed and interpretation. Based on traditional ISTA [34] and AMP [35], J. Zhang *et al.* [36] and Z. Zhang *et al.* [37] respectively adopt learnable sampling matrices and propose more powerful CNNs with inter-block relationship exploitation and deep structural insights. Chun *et al.* [33] further introduce momentum mechanism in each unfolded iteration for recovery acceleration.

In addition to the network structure and optimization algorithm design, the CS ratio allocation and model scalability have been concerned and studied. By considering the characteristics of human visual system, Yu *et al.* [38] adopt a strategy of allocating less sampling rates to non-salient blocks and more to salient ones. Zhou *et al.* [39] develop a multi-channel network to obtain high restoring accuracy with a similar allocating mechanism. To balance the CS system complexity and flexibility, Shi *et al.* [40] propose a hierarchical CNN to achieve scalable sampling and recovery. Recently, You *et al.* [41] propose a robust controllable network dubbed COAST to deal with the recoveries under arbitrary sampling matrices.

Although most state-of-the-art methods yield high performances, the problems they focus on are still not comprehensive enough, thus leading to many their restricted applications. In this paper, we propose a novel Content-Aware Scalable¹ deep Network dubbed CASNet to comprehensively solve natural image CS problems. Specifically, we explore the structural potentials of CASNet by developing the existing approaches and trying to keep and organically combine their merits. As illustrated in Fig. 1, CASNet is composed of a sampling subnet (SS), an initialization subnet (IS) and a recovery subnet (RS). We propose a unified learnable generating matrix to produce sampling matrices and a data-driven saliency detector so that our CASNet can perform efficient and adaptive CS ratio allocations with the fine granular model scalability. The multi-phase recovery subnet can explore the inter-block relationship under the saliency information guidance. With the well-defined structure and mutual supports among different components, CASNet can be learned in a completely end-to-end manner and enjoys the merits of accurate recovery and interpretability.

The main three contributions of this paper are as follows:

□ (1) A content-aware scalable network dubbed CASNet is proposed to achieve block-wise CS ratio allocation and handle image CS task under any sampling rate $r \in [0, 1]$ by a single network. To our knowledge, this is the first work integrating

Manuscript received August 9, 2021; revised April 12, 2022; accepted July 12, 2022. This work was supported by National Natural Science Foundation of China (61902009).

Bin Chen and Jian Zhang are with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China. J. Zhang is also with the Peng Cheng Laboratory, Shenzhen, China. (e-mail: chenbin74851@126.com; zhangjian.sz@pku.edu.cn).

¹Following [40], [41], we call a CS network scalable (or fine granular scalable) when it uses one set of parameters for a single network to handle multiple CS ratios (or all CS ratios in the common range of $[0, 0.5]$ or $[0, 1]$).

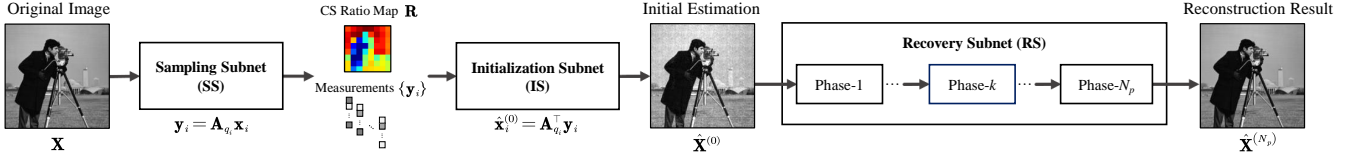


Fig. 1. Illustration of our CASNet framework, which consists of a sampling subnet (SS, Fig. 4), an initialization subnet (IS, Fig. 5), and a recovery subnet (RS, Fig. 6). In block-based CS scheme, the original image \mathbf{X} is divided into l non-overlapped $B \times B$ blocks $\{\mathbf{x}_i\}$. In SS, they are reshaped to be $N \times 1$ ($N = B^2$) and linearly sampled one-by-one to obtain block measurements $\{\mathbf{y}_i | \mathbf{y}_i \in \mathbb{R}^{q_i \times 1}, 0 \leq q_i \leq N\}$, which are then initialized by IS to get a joint recovered estimation $\hat{\mathbf{X}}$ by RS. Under this framework, CASNet can achieve content-aware sampling and fine granular scalability with only one single network.

CS ratio allocation, model scalability and unfolded recovery.

□ (2) We propose to adopt a lightweight CNN to adaptively detect the image saliency distribution, and design a block ratio aggregation (BRA) strategy to achieve block-wise CS ratio allocation instead of using a handcrafted detecting function adopted by previous saliency-based methods [38], [39].

□ (3) We further provide two boosting strategies and a four-stage implementation for CASNet evaluations. Experiments show that CASNet outperforms state-of-the-art CS methods with benefiting from the inherent strong compatibility and mutual supports among its different components and strategies.

II. RELATED WORKS

We group existing CS approaches into *optimization-* and *network-based* methods. In this section, we will retrospect both and focus on the specific methods most relevant to our own.

Optimization-based Methods: Traditional CS approaches usually recover \mathbf{x} from the acquired \mathbf{y} by solving the following optimization problem which is often assumed to be convex:

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

here $\lambda \mathcal{R}(\mathbf{x})$ is a prior term with regularization parameter λ .

Many of the classic domains [42], [43] and prior knowledge about transform coefficients [44], [45] have been exploited to reconstruct images by means of various iterative solvers (e.g. ISTA [34], ADMM [43] and AMP [35]). And there are lots of methods based on image nonlocal properties [16], [46]–[48] and denoiser-integrating techniques [49]–[52]. Furthermore, some data-driven methods are proved to be robust and effective including dictionary learning [53], tight-frame learning [54] and convolutional operator learning [55]. However, these methods give rise to high computational cost and are suffering from challenging prior or parameter settings and fine-tunings.

Network-based Methods: Recently, the network-based CS methods demonstrate their promising performances. Kulkarni *et al.* [23] propose to learn a CNN to regress an image patch from its corresponding measurement. Shi *et al.* [56] propose a framework called CSNet which avoids blocking artifacts by learning a mapping between block measurements and jointly recovered image. Sun *et al.* [26] propose a dual-path attention network dubbed DPA-Net, whose structural and textural paths are bridged by a texture attention module. Lately, unfolding networks are developed to combine the merits of optimization-based methods and network-based methods. Zhang *et al.* [32] develop a so-called ISTA-Net⁺ which works well for CS and CS-MRI tasks. J. Zhang *et al.* [36] and Z. Zhang *et al.* [37] further exploit the inter-block relationship and propose the networks dubbed OPINE-Net⁺ and AMP-Net, respectively.

However, most existing network-based methods regard the CS sampling-reconstruction under different sampling rates as

different tasks. They train a set of network parameters for only a specific CS ratio, and need to learn N networks to support all sampling rates in $\{q/N\}_{q=1}^N$. This causes the complex and huge CS system (with storing all parameters), which is expensive for hardware implementation. By considering CS system memory cost the importance of model scalability, Shi *et al.* [40] propose a scalable CNN named SCSNet which adopts a hierarchical structure and a heuristic greedy method performed on an auxiliary dataset to separately learn and sort measurement bases, but this brings its training difficulty and the defect of delicacy. Inspired by the block-wise sampling rate allocation mechanism in [38] and a block-based CS algorithm in [8], Zhou *et al.* [39] propose a multi-channel framework dubbed BCS-Net using a channel-specific sampling network to achieve adaptive CS ratio allocation. However, the handcrafted and fixed saliency detecting method based on DCT causes its weak adaptability, and the structural inadequacy of multi-channel framework brings its inflexibility and low efficiency. Recently, You *et al.* [41] solve the CS problems of arbitrary-sampling matrices by a controllable network named COAST with introducing a random projection augmentation (RPA) strategy to promote training diversity, but its sampling matrices are independently generated and lack adaptability with recovery network, and it needs hundreds of sampling matrices with several pre-defined CS ratios for training, thus leading to its expensive learning and restricted performance.

III. PROPOSED METHOD

In this section, we first give an overview of our main ideas, elaborate on the details of CASNet framework design, then describe the integrated model, all involved parameters of which can be jointly trained end-to-end, and finally provide an implementation scheme for evaluations and deployments.

A. Overview of Main Ideas

(1) **Saliency-based CS ratio allocation.** Block-based CS [8]–[10] is effective for processing high-dimensional images. In particular, we adopt an adaptive sampling rate allocation scheme that was preliminarily studied in [38], [39] with the human perception consideration. Since image information is not always evenly distributed, one way to get restored image quality improvements is to make better CS ratio allocations by using the saliency distribution. Here we use the definition of visual saliency in [57], that is, a location with low spatial correlation with its surroundings is salient. As Fig. 2 shows, for the given example image, the block in red box should be assigned a higher CS ratio compared to the one in blue box due to its more complex details with richer information.

(2) **Fine granular scalable sampling based on a unified learnable generating matrix.** The design of sampling matrix Φ , which is composed of one or more measurement bases

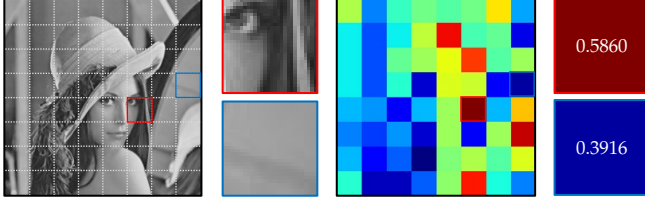


Fig. 2. Illustration of an image named ‘‘Lena’’ from Set11 [23] with non-uniform information distribution (left) and its content-aware adaptive sampling rate allocation result with keeping an average CS ratio $r = 50\%$ (right). We can assign higher sampling rates to the more salient blocks (e.g. the block in red box) compared to the less salient ones (e.g. the block in blue box).

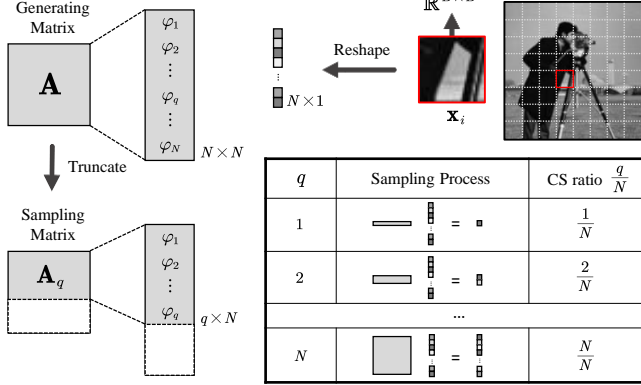


Fig. 3. Illustration of our learnable generating matrix \mathbf{A} with a descending base importance order for achieving fine granular model scalability.

$\{\varphi_i\}_{i=1}^M$, has been one of the main challenges in CS fields [40]. Motivated by the low-rank theory and [40], we propose to obtain all sampling matrices from a unified learnable generating matrix \mathbf{A} with a decreasing trend of base importance from the first row to the last, *i.e.* the generating matrix is designed to generate the sampling matrix for any CS ratio and can be learned from data. As Fig. 3 illustrates, by preserving the most important q measurement bases, we can obtain the sampling matrix $\mathbf{A}_q = \mathbf{A}[1 : q]$ for CS ratio $r = q/N$, where $\mathbf{A}[i : j] \in \mathbb{R}^{(j-i+1) \times N}$ denotes a truncated matrix that consists of the i -th to the j -th rows of generating matrix \mathbf{A} . With this idea, the model scalability can be achieved with better insights into the matrix structure. Compared with most methods that need to train N sampling matrices with a total memory cost of $\sum_{q=1}^N (qN) = [N^2(N+1)/2] \in \mathcal{O}(N^3)$ for all CS ratios $\{q/N\}_{q=1}^N$, our generating matrix takes a largely reduced storage complexity of $(N^2) \in \mathcal{O}(N^2)$.

(3) Deep unfolding reconstruction. With theoretical guarantees and favorable strong interpretability, deep unfolding technique integrates both optimization-based and network-based methods by fusing the data-fidelity constraints into the learning model. Inspired by the previous works [30], [32], [33], [36], [37], [41], our recovery network is built on unfolding framework, maps each optimization iteration into a learnable phase structure, and recovers the target image step-by-step.

B. Architecture Design of CASNet

In this subsection, we will illustrate the architecture design of CASNet and related techniques. As Fig. 1 shows, CASNet is composed of a sampling subnet (SS), an initialization subnet (IS), and a recovery subnet (RS). Note that our CASNet mainly focuses on one-channel natural images, and it could be easily

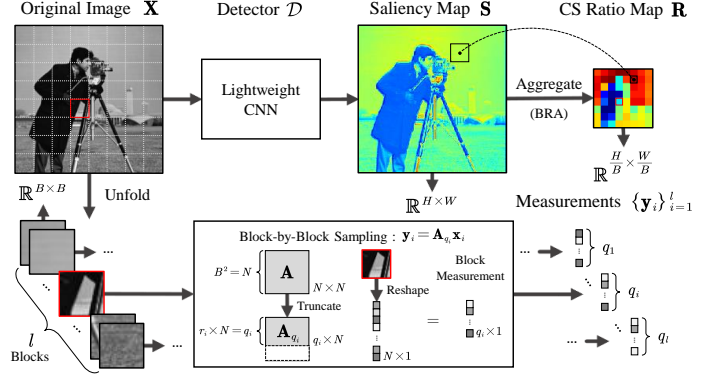


Fig. 4. Illustration of the sampling subnet (SS), which conducts a three-stage sampling process consisted of saliency detection, CS ratio allocation, and block-by-block sampling. Each $B \times B$ block is reshaped (or vectorized) to $N \times 1$ and sampled by its corresponding sampling matrix $\mathbf{A}_{q_i} \in \mathbb{R}^{q_i \times N}$.

extended² to colorful image or video CS tasks.

1) Sampling Subnet (SS): As Fig. 4 illustrates, the sampling process in SS can be divided into three stages: saliency detection, CS ratio allocation, and block-by-block sampling.

(1) Saliency detection. In the first stage, instead of using a manually set detecting method [38], [39], we adopt a CNN as the saliency detector \mathcal{D} to evaluate the saliency of each location and highlight the importance information of different regions. It consists of a convolution layer, three residual blocks and another convolution layer to estimate and give out a single-channel saliency map \mathbf{S} with the same size $H \times W$ as input.

Algorithm 1: Block ratio aggregation (BRA).

Input: Saliency map $\mathbf{S} \in \mathbb{R}^{H \times W}$, block size B , target average measurement size q (the expected CS ratio is q/B^2 or q/N , upper bound K).

Output: CS ratio map $\mathbf{R} \in \mathbb{R}^{(H/B) \times (W/B)}$.

```

1  $\mathbf{S} := \text{softmax}(\mathbf{S});$  // softmax normalization
2  $l := (H/B) \times (W/B);$  // total number of blocks
3  $\mathbf{Q} := q \times l \times \text{sumpool}_{B \times B}(\mathbf{S});$ 
4  $i := 0, T := 10;$ 
5 while true do
6    $i := i + 1;$ 
7    $\mathbf{Q} := \text{round}(\text{clip}_{0,K}(\mathbf{Q}));$ 
8    $\delta := \text{average}(\mathbf{Q}) - q;$ 
9   if  $\delta$  equals 0 then
10    break;
11   else if  $i \leq T$  then
12     $\mathbf{Q} := \mathbf{Q} - \delta;$  // method #1
13   else
14    generate a random matrix  $\Delta \in \mathbb{N}^{(H/B) \times (W/B)}$ 
      following the multinomial distribution with
      parameters  $\text{abs}(\delta l)$  and  $\mathbf{P} \in \mathbb{R}^{(H/B) \times (W/B)}$ ,
      where all elements of  $\mathbf{P}$  are set to  $(1/l);$ 
15     $\mathbf{Q} := \mathbf{Q} - \text{sign}(\delta) \times \Delta;$  // method #2
16   end
17 end
18  $\mathbf{R} := \mathbf{Q}/(B^2);$  // final normalization
19 return  $\mathbf{R};$  // end of BRA

```

²For colorful image or video data, the CASNet extension could be achieved by sampling and reconstructing channel-by-channel or frame-by-frame.

(2) CS ratio allocation. In the second stage, \mathbf{S} is logically divided into l blocks of size $B \times B$, where $B = \sqrt{N}$. Then the block aggregation is performed to get a CS ratio map \mathbf{R} which incorporates the allocated block sampling rates $\{r_i\}_{i=1}^l$. In fact, the available sampling rate of each block can be only selected from $\{q/N\}_{q=1}^N$ due to the limited sampling matrix size, where we denote q_i as the measurement size of the i -th block. We design a block ratio aggregation (BRA) strategy, which can be summarized as softmax normalization, sumpooling aggregation and error correction, to achieve accurate CS ratio allocation. Concretely, BRA applies softmax normalizer to \mathbf{S} , performs $B \times B$ sumpooling to get an aggregated weight map, and times it with the target sum of measurement size (ql) to get the measurement size map \mathbf{Q} , then \mathbf{Q} is checked and corrected iteratively. In each correction iteration, \mathbf{Q} is sheared and discretized to make all its elements be integers in $[0, N]$ with specifying the block measurement size upper bound as N , then its average error δ determines whether the correction needs to be performed. Alg. 1 exhibits the details of BRA, and numbered lines 11-15 show its two different correction methods: uniform descent and random error elimination based on the multinomial distribution. In our experiments, the BRA strategy is implemented by PyTorch [58] with the differential property which enables the backpropagations can reach the saliency detector \mathcal{D} and guide the update of its parameters. It distributes \mathbf{R} in no more than 16 correction iterations in all our evaluations. Appx. A and Appx. B provide a simple instance and our convergence analysis of BRA strategy, respectively.

(3) Block-by-block sampling. In the final stage, the original image \mathbf{X} is unfolded into $B \times B$ blocks $\{\mathbf{x}_i\}_{i=1}^l$, and each block \mathbf{x}_i is sampled by $\mathbf{y}_i = \mathbf{A}_{q_i} \mathbf{x}_i$ with its corresponding sampling matrix \mathbf{A}_{q_i} . Note that \mathbf{A}_{q_i} is obtained by truncating \mathbf{A} and preserving its first q_i rows, here $q_i = r_i \times N$, and r_i is the corresponding allocated CS ratio in \mathbf{R} .

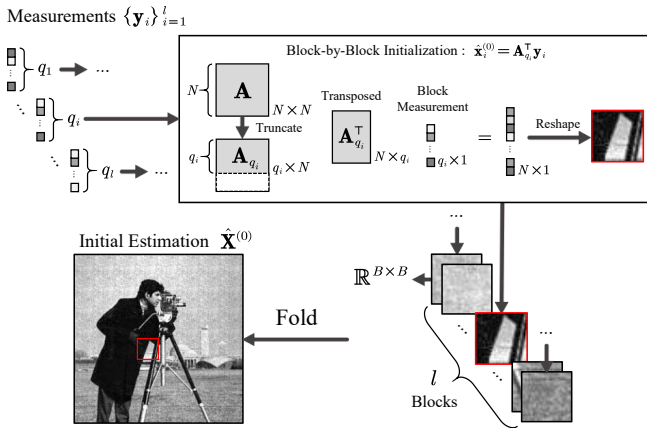


Fig. 5. Illustration of the initialization subnet (IS), which reuses the sampling matrices to handle the blocks-measurements dimensionality mismatch.

2) Initialization Subnet (IS): Fig. 5 illustrates the initialization process in IS. Instead of exploiting a fully-connected layer to handle the dimensionality mismatch between the image blocks and their measurements [23], [37], following [59], the IS directly uses $\hat{\mathbf{x}}_i^{(0)} = \mathbf{A}_{q_i}^\top \mathbf{y}_i$ for block initialization. Concretely, each block is initialized by its corresponding transposed sampling matrix $\mathbf{A}_{q_i}^\top$, then all results are reshaped and

folded to form the initial estimation $\hat{\mathbf{X}}^{(0)}$. Without introducing parameters, IS is a simple but fast and efficient implementation to bridge SS and the recovery subnet, and can be easily applied to multi-block processing tasks with different CS ratios.

3) Recovery Subnet (RS): Considering the simplicity and interpretability, we follow [60] and directly unfold the traditional proximal gradient descent (PGD) [61] which solves Eq. (1) by iterating between the following two update steps:

$$\mathbf{z}^{(k)} = \hat{\mathbf{x}}^{(k-1)} - \rho \Phi^\top (\Phi \hat{\mathbf{x}}^{(k-1)} - \mathbf{y}), \quad (2)$$

$$\hat{\mathbf{x}}^{(k)} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}^{(k)}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad (3)$$

where k denotes the PGD iteration index, and ρ is the step size. Here Eq. (2) is a trivial gradient descent step, while Eq. (3) is the so-called proximal mapping. It is worth noting that CASNet is not limited with PGD and may be extended to other optimization algorithms like the iterative shrinkage-thresholding algorithm (ISTA) [34] and half quadratic splitting (HQS) [62]. As Fig. 6 shows, RS is composed of N_p phases, each phase conducts a two-stage process consisted of block gradient descent and saliency information guided proximal mapping, which correspond to the two update PGD steps.

(1) Block gradient descent. In the first stage, we directly map the first step of the PGD iteration and define the block gradient descent according to Eq. (2). By introducing the learnable step size $\rho^{(k)}$, this stage can be expressed as:

$$\mathbf{z}_i^{(k)} = \hat{\mathbf{x}}_i^{(k-1)} - \rho^{(k)} \mathbf{A}_{q_i}^\top (\mathbf{A}_{q_i} \hat{\mathbf{x}}_i^{(k-1)} - \mathbf{y}_i). \quad (4)$$

In this stage, the image data $\hat{\mathbf{X}}^{(k-1)}$ is unfolded into l blocks and each block is processed individually. Then the intermediate results of all blocks are folded to form $\mathbf{Z}^{(k)}$ and sent to the next stage for joint recovery with perceiving the inter-block relationships and addressing the blocking artifacts.

(2) Saliency information guided proximal mapping. Following [63] that proposes a powerful denoiser exhibiting flexibility and effectivity in image restoration tasks, we propose to conduct a CNN-based proximal mapping to solve Eq. (3) with the exploitation of CS ratio information. As Fig. 6 illustrates, the CS ratio map $\mathbf{R} \in \mathbb{R}^{(H/B) \times (W/B)}$ is repeated to obtain an expanded map $\mathbf{R}' \in \mathbb{R}^{H \times W}$ by filling the elements of \mathbf{R}' of different image blocks with their corresponding CS ratios in \mathbf{R} . Then a data-driven extractor that consists of a convolution layer, three residual blocks and another convolution layer with 1×1 kernels are used to embed \mathbf{R}' into a three-dimensional feature space. By concatenating $\mathbf{Z}^{(k)}$ with the saliency feature, we propose a powerful proximal mapping network providing a flexible way of handling different CS ratios by taking the concatenated feature as input and giving out the recovered residual content. Here we denote the feature extractor and the proximal mapping network as $\mathcal{E}^{(k)}$ and $\mathcal{P}^{(k)}$ respectively, then this process can be formulated as:

$$\hat{\mathbf{X}}^{(k)} = \mathbf{Z}^{(k)} + \mathcal{P}^{(k)} \left(\left[\mathbf{Z}^{(k)} \mid \mathcal{E}^{(k)}(\mathbf{R}') \right] \right), \quad (5)$$

where $[\mathbf{F}_1 | \mathbf{F}_2]$ is the concatenation of two feature maps \mathbf{F}_1 and \mathbf{F}_2 with the spatial size $H \times W$ in the channel dimension.

Each $\mathcal{P}^{(k)}$ adopts a U-shaped structure like [63], [64] with four scales. It consists of three encoder blocks and three decoder blocks. Each encoder block is stacked by a convolution

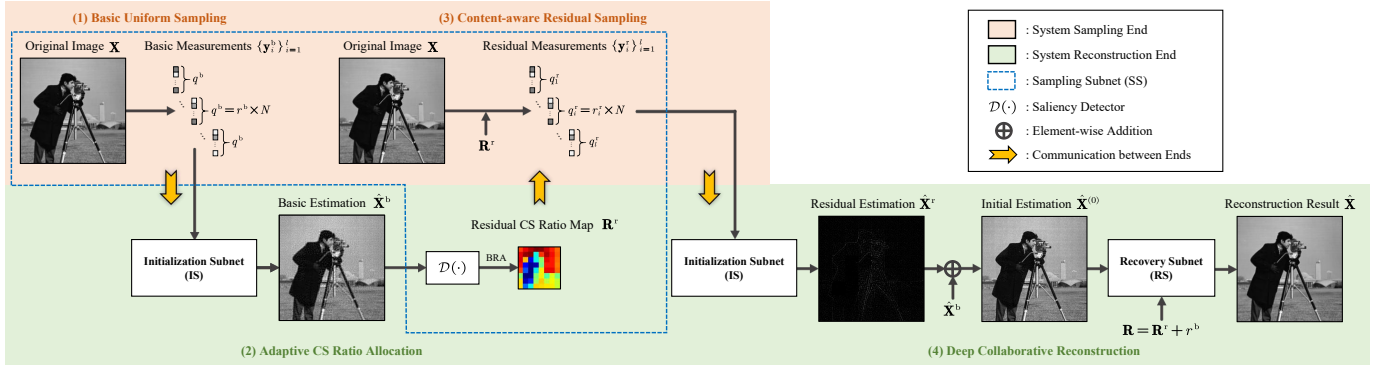


Fig. 7. Illustration of our proposed CASNet implementation, which is designed to be an interface between CASNet and the physical CS system and driven by a four-stage pipeline consisted of basic uniform sampling, adaptive CS ratio allocation, content-aware residual sampling and deep collaborative reconstruction.

uniform sampling, adaptive CS ratio allocation, content-aware residual sampling and deep collaborative reconstruction.

(1) Basic uniform sampling. Considering that the CASNet may not directly access the complete signal information for saliency detection and CS ratio allocation before its sampling in practical evaluations. Here we propose to first perform a trivial basic uniform sampling which uses the first few rows of the learned \mathbf{A} to obtain some measurements to drive the following saliency-based allocation process and ensure the content-aware property. In the first stage, with presetting the expected CS ratio r and a basic sampling proportion denoted as $\gamma \in [0, 1]$, the system sampling end uniformly samples the original image signal with the basic CS ratio $r^b = \gamma \times r$ by $\mathbf{A}_{q^b}(q^b = r^b \times N)$ generated from \mathbf{A} and sends the basic measurements $\{\mathbf{y}_i^b | \mathbf{y}_i^b = \mathbf{A}_{q^b} \mathbf{x}_i\}_{i=1}^l$ to the reconstruction end.

(2) Adaptive CS ratio allocation. In the second stage, the system reconstruction end adopts the fast and efficient IS to bring the received measurements back to the image domain by $\hat{\mathbf{x}}_i^b = \mathbf{A}_{q^b}^\top \mathbf{y}_i^b$, then folds them to form the basic estimation $\hat{\mathbf{X}}^b \in \mathbb{R}^{H \times W}$ and perform the residual CS ratio allocation based on the saliency map $\mathbf{S} = \mathcal{D}(\hat{\mathbf{X}}^b)$ and BRA with the target average measurement size of $(r \times N - q^b)$ and the corrected upper bound $(N - q^b)$ of block measurement size to obtain the residual CS ratio map \mathbf{R}^r and send it to the sampling end as a key part of the secondary sampling request.

(3) Content-aware residual sampling. In the third stage, the system sampling end achieves the network content-aware property by the non-uniform residual sampling based on the received \mathbf{R}^r . Each block \mathbf{x}_i is sampled by the residual sampling matrix slice $\mathbf{A}_{q^b+1,q_i} = \mathbf{A}[q^b+1 : q_i]$ ($q_i = q^b + r_i^r \times N$), and the residual measurements $\{\mathbf{y}_i^r | \mathbf{y}_i^r = \mathbf{A}_{q^b+1,q_i} \mathbf{x}_i\}_{i=1}^l$ are then sent to the reconstruction end.

(4) Deep collaborative reconstruction. In the last stage, the reconstruction end employs IS again to obtain the block residual initialization $\hat{\mathbf{x}}_i^r = \mathbf{A}_{q^b+1,q_i}^\top \mathbf{y}_i^r$, folds them to form the residual estimation $\hat{\mathbf{X}}^r$ and gives out the initial estimation by $\hat{\mathbf{X}}^{(0)} = \hat{\mathbf{X}}^b + \hat{\mathbf{X}}^r$. Note that each block \mathbf{x}_i is equivalent to be sampled and initialized by its corresponding complete sampling matrix \mathbf{A}_{q_i} . Then $\hat{\mathbf{X}}^{(0)}$ and the corrected CS ratio map \mathbf{R} obtained by adding r^b to each entry of \mathbf{R}^r are sent into RS to get the collaborative reconstruction $\hat{\mathbf{X}} = \hat{\mathbf{X}}^{(N_p)}$.

As we can see, the above four-stage implementation focuses on the sampling-initialization process and eliminates the gap

between CASNet and the physical CS system by acquiring a proportion of measurements in a uniform sampling manner to achieve the saliency-based CS ratio allocation and content-aware residual sampling. Under this scheme, SS is distributed at both the system sampling end and reconstruction end (see Fig. 7), while IS and RS are deployed at the reconstruction end. An appropriate value of the basic sampling proportion γ needs to be determined in a specific real-world scenario as it directly controls the trade-off between the signal pre-knowledge sufficiency for saliency evaluation and the CS ratio allocating space. Furthermore, we note that our CASNet is not limited to the proposed scheme, it may also support many other implementations such as the schemes mentioned in [38], [39] which adopt a low-resolution complementary sensor to acquire a sample image to generate a saliency map of the scene under view. For keeping the training simplicity and fair evaluations in our experiments, we train CASNet based on the simple pipeline connected by the three subnets in series with a unidirectional information flow (see Fig. 1) and the learning and boosting strategies in Sec. III-C with using the complete image for saliency detections and samplings, and only test it under our system implementation described in this subsection.

IV. EXPERIMENTAL RESULTS

A. Implementation Details

Following [24], [39], [40], we set the block size $B = 32$ and $N = 1024$, choose the basic sampling proportion $\gamma = 0.2822$, and extract luminance components of 25600 randomly cropped 128×128 image patches from T91 [67] and Train400 [68], i.e., $N_b = 25600$ and $l = 16$. All residual blocks adopt the classic Conv-ReLU-Conv structure with an identity connection [69]. The convolution layers in \mathcal{D} and $\mathcal{P}^{(k)}$ use 3×3 kernels, and the intermediate feature channel numbers of \mathcal{D} and $\mathcal{E}^{(k)}$ are set to 32 and 8, respectively. Except for $\mathcal{E}^{(k)}$, all learnable components are empirically set to be bias-free, i.e., there are no bias used in \mathbf{A} and all 3×3 convolution layers.

We implement CASNet with PyTorch [58] on a Tesla V100 GPU, employ Adam [70] optimizer with a momentum of 0.9 and a weight decay of 0.999, and adopt a batch size of 64. It takes about five days to train a 13-phase CASNet for 300 epochs with a learning rate of 1×10^{-4} and 20 fine-tuning epochs with a learning rate of 1×10^{-5} . Two widely used benchmarks: Set11 [23] and CBSD68 [71] are utilized for test,

TABLE I

THE EXPERIMENTS OF ABLATION STUDIES CONDUCTED WITH FIVE CS RATIOS ON SET11 [23]. THE BEST RESULTS ARE LABELED IN **BOLD**.

| Setting | | CS Ratio | | | | | #Param. |
|---------|---|--------------|--------------|--------------|--------------|--------------|---------------|
| | | 1% | 4% | 10% | 25% | 50% | |
| (1) | Uniform Sampling | 21.68 | 26.14 | 30.10 | 35.42 | 40.69 | 16.839M |
| (2) | w/o Saliency Information | 21.74 | 26.25 | 30.23 | 35.49 | 40.79 | 16.883M |
| (3) | Random A Initialization | 21.85 | 26.36 | 30.33 | 35.56 | 40.84 | 16.895M |
| (4) | w/o RTE | 21.81 | 26.25 | 30.15 | 35.44 | 40.79 | 16.895M |
| (5) | $\rho^{(k)}$ -Shared | 21.93 | 26.40 | 30.34 | 35.63 | 40.93 | 16.895M |
| (6) | $\mathcal{E}^{(k)}$ -Shared | 21.90 | 26.35 | 30.31 | 35.58 | 40.89 | 16.889M |
| (7) | $\mathcal{P}^{(k)}$ -Shared | 21.85 | 26.24 | 30.29 | 35.50 | 40.85 | 2.325M |
| (8) | Phase-Shared | 21.85 | 26.22 | 30.29 | 35.48 | 40.84 | 2.319M |
| (9) | Replace $\mathcal{E}^{(k)}$ with $\mathcal{E}_1^{(k)}$ | 21.83 | 26.28 | 30.27 | 35.56 | 40.87 | 16.885M |
| (10) | Replace $\mathcal{E}^{(k)}$ with $\mathcal{E}_8^{(k)}$ | 21.97 | 26.37 | 30.35 | 35.67 | 40.91 | 16.985M |
| (11) | Replace $\mathcal{E}^{(k)}$ with $\mathcal{E}_{\text{bias-free}}^{(k)}$ | 21.80 | 26.15 | 30.25 | 35.36 | 40.56 | 16.894M |
| (12) | Trained with ℓ_2 -loss | 21.92 | 26.41 | 30.52 | 35.70 | 41.01 | 16.895M |
| (13) | Ours | 21.97 | 26.41 | 30.36 | 35.67 | 40.93 | 16.895M |

TABLE II

DETAILED STRUCTURAL CONFIGURATIONS OF FOUR DIFFERENT DATA-DRIVEN SALIENCY FEATURE EXTRACTORS FOR EACH PHASE.

| Setting | Extractor | Structure | Bias-free | #Param. |
|---------|--|--|-----------|----------|
| (9) | $\mathcal{E}_1^{(k)}$ | $\text{Conv}_{1,1}^{1 \times 1}$ | × | 2 |
| (10) | $\mathcal{E}_8^{(k)}$ | $\text{Conv}_{1,32}^{1 \times 1} + 3 \times \mathcal{T}_{32}^{1 \times 1} + \text{Conv}_{32,8}^{1 \times 1}$ | × | 6664 |
| (11) | $\mathcal{E}_{\text{bias-free}}^{(k)}$ | $\text{Conv}_{1,8}^{1 \times 1} + 3 \times \mathcal{T}_8^{1 \times 1} + \text{Conv}_{8,3}^{1 \times 1}$ | ✓ | 416 |
| (13) | $\mathcal{E}^{(k)}$ (Ours) | $\text{Conv}_{1,8}^{1 \times 1} + 3 \times \mathcal{T}_8^{1 \times 1} + \text{Conv}_{8,3}^{1 \times 1}$ | × | 475 |

and all the recovered results are evaluated with the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [72] on the Y channel.

B. Ablation Studies and Discussions

(1) Effect of Block-wise CS Ratio Allocation and Saliency Information Guidance: As one of the CASNet main ideas, the saliency-based allocating scheme in SS can perform adaptive CS ratio allocations for different blocks. Fig. 8 shows that \mathcal{D} can learn to identify the locations with rich details, and the proposed BRA strategy guarantees the allocation precision by cooperating with **A**. Tab. I (1) corresponds to the CASNet variant trained and evaluated under a uniform sampling scheme without content-aware property, and exhibits our more efficient allocations with an average PSNR gain of about 0.26dB. As an efficient approach with low cost to exploit CS ratio information by making $\mathcal{P}^{(k)}$ can perceive the sampling rate distribution, the introduction of \mathbf{R}' and $\mathcal{E}^{(k)}$ results in an average PSNR gain of 0.18dB by comparing with giving only $\mathbf{Z}^{(k)}$ to $\mathcal{P}^{(k)}$, as we can see in Tab. I (2).

(2) Effect of SVD-based Initialization and RTE: Fig. 9 demonstrates the training processes of CASNets with different initialization schemes. Compared with the random initialization [36], our SVD-based scheme gives a better **A** starting point and leads to faster and more stable convergence. As a simple but generalizable enhancement scheme, the RTE strategy is to improve the network robustness by making full use of training data and the inter-phase bottlenecks. These boosting methods are parameter-free but bring 0.08dB and 0.18dB average PSNR gains as Tab. I (3)-(4) exhibit.

(3) Study of Parameter-sharing Strategies: We train four CASNet variants with different parameter-sharing strategies as reported by Tab. I (5)-(8). The most compressed and inflexible variant sharing $\rho^{(k)}$, $\mathcal{E}^{(k)}$ and $\mathcal{P}^{(k)}$ among all phases brings the largest average PSNR loss of about 0.13dB. This demonstrates

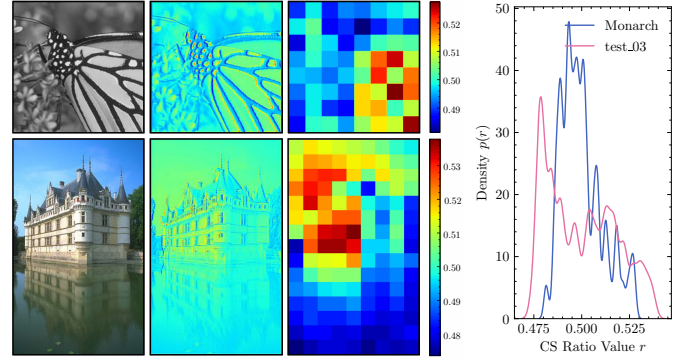


Fig. 8. Two saliency-based CS ratio allocation instances on “Monarch” from Set11 [23] (top left) and “test_03” from CBSD68 [71] (bottom left) with $r = 50\%$. Images (left) are scanned by \mathcal{D} to obtain saliency maps (middle left) and CS ratio maps \mathbf{R} (middle right). The corresponding CS ratio distribution curves (right) exhibit the instance-wise adaptability of our BRA strategy.

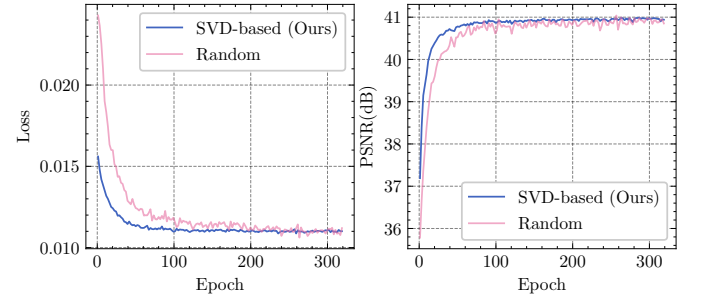


Fig. 9. Loss and average PSNR curves with CS ratio $r = 50\%$ on Set11 [23] achieved by CASNet variants with different initialization methods.

the structural effectiveness and further potential of CASNet in greatly reducing its memory complexity (with about 86.3% parameter number reduction) and being deployed on some lightweight devices with acceptable accuracy drops.

(4) Study of Saliency Feature Extractor Structure: As mentioned above, the saliency feature brings the information to guide the proximal mapping network in each phase to recover blocks with non-uniform sampling rates. Instead of manually setting a fixed sampling rate embedding operator, we adopt a data-driven approach and conduct the experiments reported in Tab. I (9)-(11). Here we denote the convolution layer with n_o kernels of size $n_i \times 1 \times 1$ as $\text{Conv}_{n_i, n_o}^{1 \times 1}$, and the residual block with a classic structure of $\text{Conv}_{n_f, n_f}^{1 \times 1} + \text{ReLU} + \text{Conv}_{n_f, n_f}^{1 \times 1}$ and an identity skip connection as $\mathcal{T}_{n_f}^{1 \times 1}$. The results in Tab. I corresponding to the extractor structural details provided in Tab. II lead to our default setting which only uses a lightweight CNN with 475 parameters including kernel biases to embed the CS ratio into a three-dimensional feature space. Three subgraphs in the first row of Fig. 10 show the mappings from CS ratio to saliency feature channel values done by (9), (13) and (10) in the first CASNet phases. One can observe that the larger extractor structure of $\mathcal{E}_8^{(k)}$ even brings a weaker performance compared with ours, and we notice that the values in five channels of the output feature given by $\mathcal{E}_8^{(1)}$ have little change as the input varies, this indicates that there are only three channels take the prominent feature information of CS ratio. Three subgraphs in the second row of Fig. 10 illustrate the diversity of mappings done by our default extractors in the 5-th, 9-th, and 13-th phases. These results validate the effectiveness of the proposed extractor $\mathcal{E}^{(k)}$ design and the

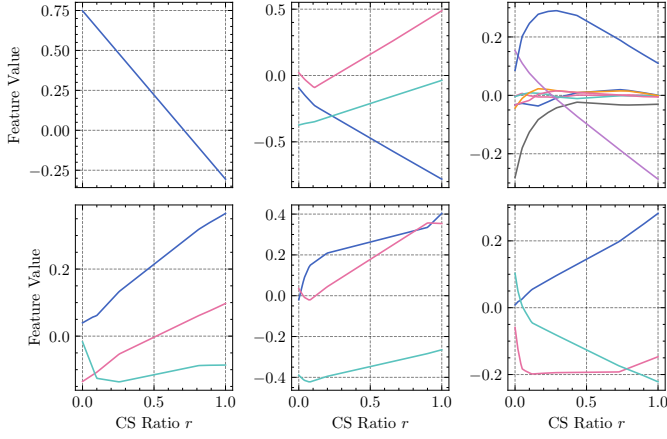


Fig. 10. Mappings from CS ratio to saliency feature space done by $\mathcal{E}_1^{(1)}$ (top left), $\mathcal{E}_1^{(1)}$ (top middle), $\mathcal{E}_8^{(1)}$ (top right), $\mathcal{E}_5^{(5)}$ (bottom left), $\mathcal{E}_9^{(9)}$ (bottom middle) and $\mathcal{E}_{13}^{(13)}$ (bottom right) in different CASNet variants. The curves in each subgraph correspond to different output feature channels, and the order of which is not concerned in our experiments.

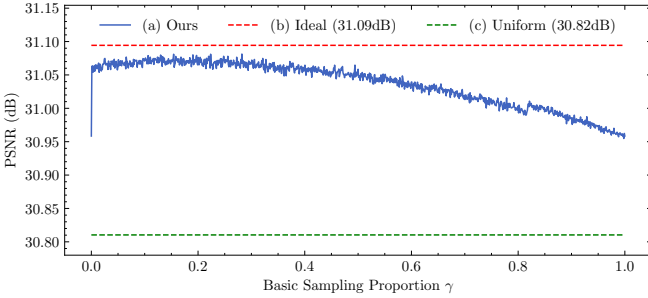


Fig. 11. Average PSNR curves on Set11 [23] and CBSD68 [71] with CS ratio $r \in \{1\%, 4\%, 10\%, 25\%, 30\%, 40\%, 50\%\}$ achieved by (a) the default CASNet version, (b) the ideal version (uses complete image for CS ratio allocation in tests) and (c) the uniform sampling version (corresponds to Tab. I (1)). Our default version (a) peaks highest accuracy 31.08dB at $\gamma = 0.2822$.

flexibility of multi-phase framework.

(5) Study of Basic Sampling Proportion γ : As a trade-off controller between the image information completeness and the sampling rate allocating space, the basic sampling proportion γ is expected to make CASNet degenerate into a uniform sampling version in evaluation when its value tends to be 0 (lack of signal pre-knowledge) or 1 (lack of residual CS ratio allocating space). Fig. 11 exhibits the average PSNR performances on Set11 [23] and CBSD68 [71] under seven different CS ratios achieved by (a) our default CASNet version (trained in content-aware sampling manner), (b) the ideal version which employs the complete image for saliency detection and sampling in evaluations (trained in content-aware sampling manner), and (c) the trivial uniform sampling version (trained in uniform sampling manner). The default version (a) meets its worst performance of 30.95dB when γ equals to 0 or 1. And there is a sharp increase (about 0.12dB) at the beginning end of its corresponding curve, which means that only a small proportion of the most important measurements can lead to relatively accurate saliency predictions and allocations. This curve then tends to slowly increase in the range of $[0, 0.2822]$, peaks to 31.08dB at $\gamma = 0.2822$, and tends to decrease in $[0.2822, 1]$. We observe that our CASNet implementation (a) can even bring a recovery accuracy close to the ideal version (b) with a PSNR gap of only about 0.01dB and remain a

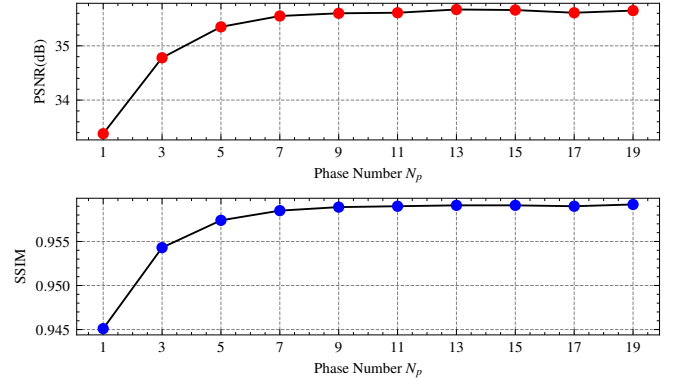


Fig. 12. Average PSNR/SSIM curves on Set11 [23] achieved by CASNet variants with various phase numbers in the case of CS ratio $r = 25\%$.

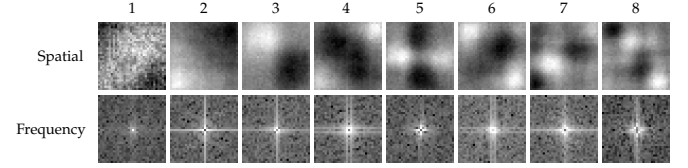


Fig. 13. Visualization of the first eight rows of the learned \mathbf{A} . The former ones exhibit narrower frequency distributions, which indicates that they pay more attention to low-frequency information.

distance of about 0.26dB compared with the uniform sampling version (c), which is also exceeded by the default degenerated uniform sampling one (a) with a PSNR gap of 0.13dB. These results fully verify the effectiveness of our four-stage CASNet implementation and show the importance of keeping the CS ratio diversity in the training process.

(6) Study of Phase Number N_p : Under the PGD-unfolding framework, our CASNet is expected to get higher performance with RS phase number N_p increases. Fig. 12 gives the comparison among CASNet variants with different phase numbers with CS ratio $r = 25\%$. We can see that both the curves increase as N_p increases, but they are becoming almost flat when $N_p \geq 13$. Therefore, we choose the CASNet with 13 phases as the default setting by considering the trade-off between model complexity and recovery accuracy.

(7) Analysis of the Learned \mathbf{A} : To get the insight of sampling matrix structure, we give some findings of the learned generating matrix \mathbf{A} . First, although there is no constraints on \mathbf{A} in loss function, the orthogonal form $\mathbf{A}\mathbf{A}^T = \eta\mathbf{I}$ is approximately satisfied during training, where η shows a logarithmic-like growth trend from 1 to about 4.865 as illustrated in the left subgraph of Fig. 14. The middle left subgraph visualizes $[(\tilde{\mathbf{A}})(\tilde{\mathbf{A}})^T]$ with the normalized $\tilde{\mathbf{A}} = \mathbf{A}/\sqrt{\eta}$, in which the position (u, v) corresponds to the inner product of the u -th and the v -th normalized bases. We observe that all diagonal elements are about 1.0 and others are near to 0. Second, we plot the histograms and the proportion of base elements close to zero of $\tilde{\mathbf{A}}$ and the fixed random one as shown in the right two subgraphs. We observe that the learned one exhibits wider and sparser distribution, and the base sparsity gets higher as the index increases. Third, we reshape the first eight learned rows into $B \times B$ and visualize them in Fig. 13 with frequency. They show structured and anisotropic spatial results different from traditional manually defined filters. And frequencies of the former bases are narrower, which means that they pay more

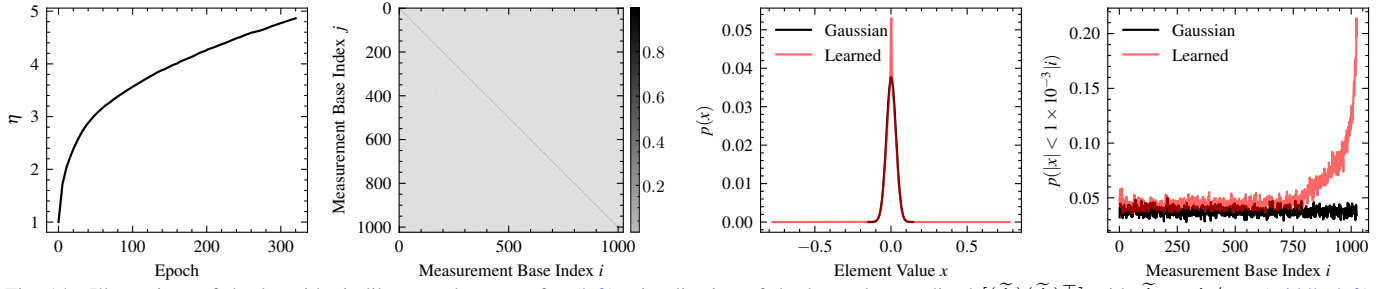


Fig. 14. Illustrations of the logarithmic-like growth curve of η (left), visualization of the learned normalized $[(\tilde{\mathbf{A}})(\tilde{\mathbf{A}})^T]$ with $\tilde{\mathbf{A}} = \mathbf{A}/\sqrt{\eta}$ (middle left), histograms of $\tilde{\mathbf{A}}$ and fixed random Gaussian matrix (middle right), and the proportion curves of measurement base elements close to zero (right).

TABLE III

COMPARISONS OF THE FUNCTIONAL FEATURES, TOTAL PARAMETER NUMBERS FOR SEVEN CS RATIOS AND AVERAGE INFERENCE TIME OF SAMPLING AND RECOVERING A 256×256 IMAGE ON AN 1080Ti GPU WITH CS RATIO $r = 10\%$ AMONG VARIOUS NETWORK-BASED CS METHODS.

| Method | Sampling Matrix Learnability | Adaptive CS Ratio Allocation | Fine Granular Scalability | Deblocking Ability | CS Ratio Information Exploitation | #Param. (M) /Time (ms) |
|--|------------------------------|------------------------------|---------------------------|--------------------|-----------------------------------|------------------------|
| ReconNet [23] (CVPR 2016) | × | × | × | × | × | 0.98/2.69 |
| ISTA-Net ⁺ [32] (CVPR 2018) | × | × | × | × | × | 2.38/5.65 |
| DPA-Net [26] (TIP 2020) | × | × | × | ✓ | × | 65.17/36.49 |
| ConvMMNet [73] (TCI 2020) | ✓ | × | × | × | × | 7.64/19.32 |
| CSNet ⁺ [24] (TIP 2019) | ✓ | × | × | ✓ | × | 4.35/16.77 |
| OPINE-Net ⁺ [36] (JSTSP 2020) | ✓ | × | × | ✓ | × | 4.35/17.31 |
| AMP-Net [37] (TIP 2021) | ✓ | × | × | ✓ | × | 6.08/27.38 |
| SCSNet [40] (CVPR 2019) | ✓ | × | ✓ | ✓ | × | 0.80/30.91 |
| BCS-Net [39] (TMM 2020) | ✓ | ✓ | × | ✓ | × | 1.64/83.86 |
| COAST [41] (TIP 2021) | ✓ | × | ✓ | ✓ | ✓ | 1.12/45.54 |
| CASNet (Ours) | ✓ | ✓ | ✓ | ✓ | ✓ | 16.90/97.37 |

TABLE IV

AVERAGE PSNR(DB)/SSIM PERFORMANCE COMPARISONS AMONG VARIOUS CS METHODS ON Set11 [23] AND CBSD68 [71] WITH SEVEN DIFFERENT CS RATIOS. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN RED AND BLUE COLORS, RESPECTIVELY.

| Dataset | Method | CS Ratio r | | | | | | |
|-------------|-----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | 1% | 4% | 10% | 25% | 30% | 40% | 50% |
| Set11 [23] | ReconNet [23] | 17.43/0.4017 | 20.93/0.5897 | 24.38/0.7301 | 28.44/0.8531 | 29.09/0.8693 | 30.60/0.9020 | 32.25/0.9177 |
| | ISTA-Net ⁺ [32] | 17.48/0.4479 | 21.32/0.6037 | 26.64/0.8087 | 32.59/0.9254 | 33.68/0.9352 | 35.97/0.9544 | 38.11/0.9707 |
| | DPA-Net [26] | 18.05/0.5011 | 23.50/0.7205 | 26.99/0.8354 | 31.74/0.9238 | 33.35/0.9425 | 35.21/0.9580 | 36.80/0.9685 |
| | ConvMMNet [73] | 19.53/0.4902 | 23.92/0.7384 | 27.63/0.8594 | 32.05/0.9300 | 33.25/0.9442 | 35.04/0.9579 | 36.72/0.9687 |
| | CSNet ⁺ [24] | 20.67/0.5411 | 24.83/0.7480 | 28.34/0.8580 | 33.34/0.9387 | 34.27/0.9492 | 36.44/0.9690 | 38.47/0.9796 |
| | OPINE-Net ⁺ [36] | 20.15/0.5340 | 25.69/0.7920 | 29.81/0.8904 | 34.86/0.9509 | 35.79/0.9541 | 37.96/0.9633 | 40.19/0.9800 |
| | AMP-Net [37] | 20.55/0.5638 | 25.14/0.7701 | 29.42/0.8782 | 34.60/0.9469 | 35.91/0.9576 | 38.25/0.9714 | 40.26/0.9786 |
| | SCSNet [40] | 21.04/0.5562 | 24.29/0.7589 | 28.52/0.8616 | 33.43/0.9373 | 34.64/0.9511 | 36.92/0.9666 | 39.01/0.9769 |
| | BCS-Net [39] | 20.86/0.5510 | 24.90/0.7531 | 29.42/0.8673 | 34.20/0.9408 | 35.63/0.9495 | 36.68/0.9667 | 39.58/0.9734 |
| | COAST [41] | —/— | —/— | 30.03/0.8946 | —/— | 36.35/0.9618 | —/— | 40.32/0.9804 |
| | CASNet (Ours) | 21.97/0.6140 | 26.41/0.8153 | 30.36/0.9014 | 35.67/0.9591 | 36.92/0.9662 | 39.04/0.9760 | 40.93/0.9826 |
| CBSD68 [71] | ReconNet [23] | 18.27/0.4007 | 21.66/0.5210 | 24.15/0.6715 | 26.04/0.7833 | 27.53/0.8045 | 29.08/0.8658 | 29.86/0.8951 |
| | ISTA-Net ⁺ [32] | 19.14/0.4158 | 22.17/0.5486 | 25.32/0.7022 | 29.36/0.8525 | 30.25/0.8781 | 32.30/0.9195 | 34.04/0.9424 |
| | DPA-Net [26] | 20.25/0.4267 | 23.50/0.7205 | 25.47/0.7372 | 29.01/0.8595 | 29.73/0.8827 | 31.17/0.9156 | 32.55/0.9386 |
| | ConvMMNet [73] | 21.27/0.4805 | 24.21/0.6508 | 26.75/0.7831 | 30.16/0.8935 | 31.02/0.9134 | 32.81/0.9402 | 34.30/0.9570 |
| | CSNet ⁺ [24] | 22.21/0.5100 | 25.43/0.6706 | 27.91/0.7938 | 31.12/0.9060 | 32.20/0.9220 | 35.01/0.9258 | 36.76/0.9638 |
| | OPINE-Net ⁺ [36] | 22.11/0.5140 | 25.20/0.6825 | 27.82/0.8045 | 31.51/0.9061 | 32.35/0.9215 | 34.95/0.9261 | 36.35/0.9660 |
| | AMP-Net [37] | 22.18/0.5207 | 25.47/0.6534 | 27.79/0.7853 | 31.37/0.8749 | 32.68/0.9291 | 35.06/0.9395 | 36.59/0.9620 |
| | SCSNet [40] | 22.03/0.5126 | 25.37/0.6623 | 28.02/0.8042 | 31.15/0.9058 | 32.64/0.9237 | 35.03/0.9214 | 36.27/0.9593 |
| | BCS-Net [39] | 21.95/0.5119 | 25.44/0.6597 | 27.98/0.8015 | 31.29/0.8846 | 32.70/0.9301 | 35.14/0.9397 | 36.85/0.9682 |
| | COAST [41] | —/— | —/— | 27.92/0.8061 | —/— | 32.66/0.9256 | —/— | 36.43/0.9663 |
| | CASNet (Ours) | 22.49/0.5520 | 25.73/0.7079 | 28.41/0.8231 | 32.31/0.9196 | 33.40/0.9359 | 35.43/0.9581 | 37.48/0.9728 |

attention to low-frequency information. These facts verify our \mathbf{A} design with a descending base importance order and the feasibility of its data-driven learning scheme.

C. Comparison with State-of-the-Arts

We compare CASNet with ten representative state-of-the-art CS networks: ReconNet [23], ISTA-Net⁺ [32], CSNet⁺ [24], DPA-Net [26], ConvMMNet [73], OPINE-Net⁺ [36], AMP-Net [37], SCSNet [40] BCS-Net [39] and COAST [41]. ReconNet, CSNet⁺, DPA-Net and ConvMMNet are traditional network-based methods; ISTA-Net⁺, OPINE-Net⁺ and AMP-Net are traditional unfolding methods; SCSNet is scalable

with a hierarchical structure; BCS-Net is saliency-based and achieves CS ratio allocation with a multi-channel architecture; COAST achieves scalability by generalizing to arbitrary sampling matrices. More details of high-level functional feature comparisons are given in Tab. III, where CASNet exhibits its organic integration of different features and merits.

(1) Average PSNR/SSIM Comparisons on Benchmarks:

The PSNR/SSIM results of different CS methods on Set11 [23] and CBSD68 [71] for seven CS ratios are provided in Tab. IV. Despite the fact that CSNet⁺, OPINE-Net⁺ and AMP-Net learn separate models and exceed the first three

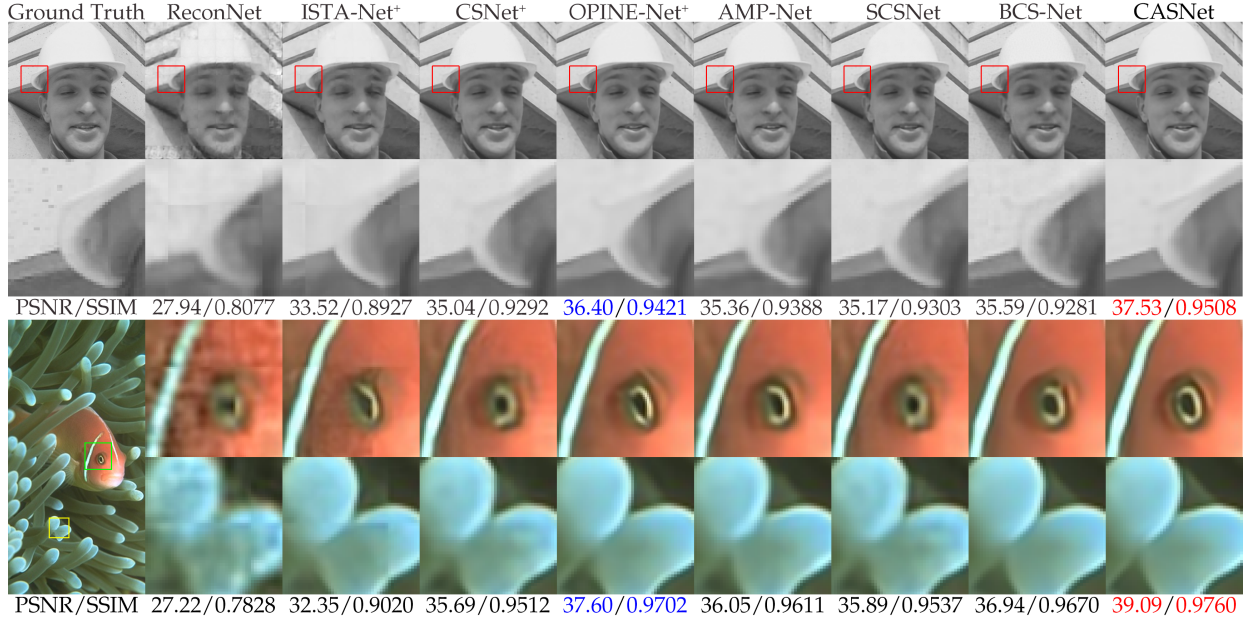


Fig. 15. Visual comparisons with PSNR(dB)/SSIM on recovering two images from Set11 [23] (top) and CBSD68 [71] (bottom) respectively in the case of CS ratio $r = 10\%$. Our CASNet shows its superiority by giving better visual results with more details and shaper edges.



Fig. 16. Visual comparisons with PSNR(dB)/SSIM on recovering an image from Set11 [23] (top) and two images from CBSD68 [71] (middle and bottom, respectively) in the case of CS ratio $r = 10\%$.

based on fixed random sampling matrices, they still fail to outperform CASNet and remain the average PSNR/SSIM distances of 1.38dB/0.0252 and 0.65dB/0.0253 on Set11 and CBSD68, respectively. Comparisons with SCSNet, BCS-Net and COAST (the default version with sampling matrices from learned OPINE-Nets) show the superiority of CASNet which combines our three main ideas with a well-defined structure

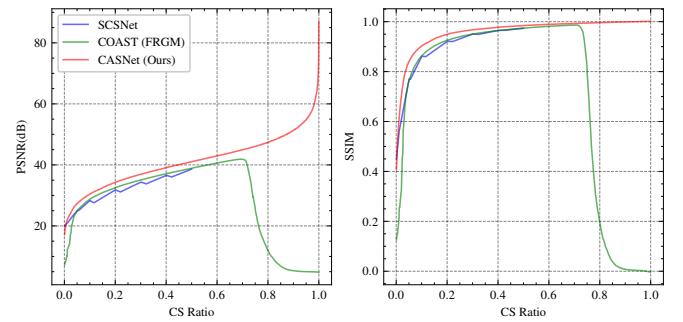


Fig. 17. Comparison of fine granular scalable CS reconstruction performance among SCSNet [40], the default COAST version with fixed random Gaussian matrices (FRGM) [41] and our CASNet on Set11 [23].

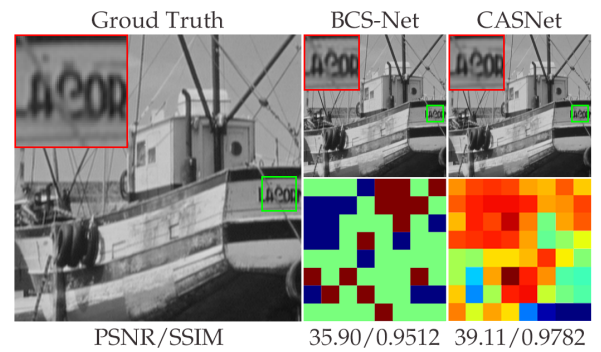


Fig. 18. Comparison between two saliency-based method on the image named "Boats" (left) from Set11 [23] with CS ratio $r = 30\%$. The recovered results are in top middle and top right. In addition, here we provide their CS ratio allocating results in bottom middle and bottom right.

and boosting schemes. Note that the uniform sampling CASNet version (see Tab. I (1)) is already able to show a large PSNR exceeding over 1dB on Set11 [23] compared with the existing best methods, and our saliency-based allocation can further breakthrough the accuracy saturation and bring a large step forward on this basic variant. The visual comparisons in Fig. 15 and Fig. 16 show that our CASNet is able to recover high-quality results with more details and sharper edges.

(2) Comparisons of Scalable Sampling and Recovery:

The recovery accuracy curves in Fig. 17 illustrate that CASNet outperforms the scalable SCSNet [40] and COAST [41] with a large margin in nearly all cases. Here we choose the default COAST version with fixed random Gaussian matrices (FRGM) as it is expensive to train thousands of OPINE-Nets with different CS ratios and utilize the learned matrices as described in [41]. Compared with using a hierarchical CNN and a greedy method to get the base importance order to support multiple CS ratios in $[0, 0.5]$ by SCSNet, and adopting a random projection augmentation strategy with CS ratios in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ for training by COAST which works normally in $[0.05, 0.7]$ but gets failed in other cases, our scalable scheme based on \mathbf{A} achieves the more robust fine granular scalability with the complete CS ratio range of $[0, 1]$, provides a much easier and intuitive joint learning method, and brings us a more clear framework with strong interpretability.

(3) Comparisons of Saliency-based CS Ratio Allocation:

The recovery and CS ratio allocation results by BCS-Net [39] and CASNet in Fig. 18 shows the flexibility of CASNet assigning a variety of CS ratio levels with smooth transition among blocks and its powerful recovery ability with a PSNR/SSIM exceeding of 3.21dB/0.0270 compared with BCS-Net, which adopts a seven-channel sampling method based on a handcrafted saliency detecting method and only assigns three ratio levels to the blocks.

V. CONCLUSION AND FUTURE WORK

A novel content-aware scalable network named CASNet is proposed to comprehensively address image CS problems, which tries to make full use of the merits of traditional methods by achieving adaptive CS ratio allocation, fine granular scalability, and high-quality reconstruction collectively. Different from the previous saliency-based methods, we use a data-driven saliency detector and a block ratio aggregation (BRA) strategy to achieve accurate sampling rate allocations. A unified learnable generating matrix is developed to produce sampling matrices with memory complexity reduction. The PGD-unfolding recovery subnet exploits the CS ratio information and the inter-block relationship to restore images step-by-step. We use an SVD-based initialization scheme to accelerate training, and a random transformation enhancement (RTE) strategy to improve the network robustness. All the CASNet parameters can be indiscriminately learned end-to-end with strong compatibility and mutual supports among its components and strategies. Furthermore, we consider the possible gap between the CASNet framework and physical CS systems, and provide a four-stage implementation for fair evaluations and practical deployments. Extensive experiments demonstrate that CASNet greatly improves upon the results of state-of-the-art CS methods with high structural efficiency and deep matrix insights. Our future work is to extend CASNet to video CS problems, and Fourier-based medical applications [9] like CS-MRI [5], [6] and sparse-view CT [7] tasks.

APPENDIX A

A SIMPLE INSTANCE OF BRA STRATEGY

In order to provide a clear exhibition and a better understanding of our BRA strategy, here we provide an instance

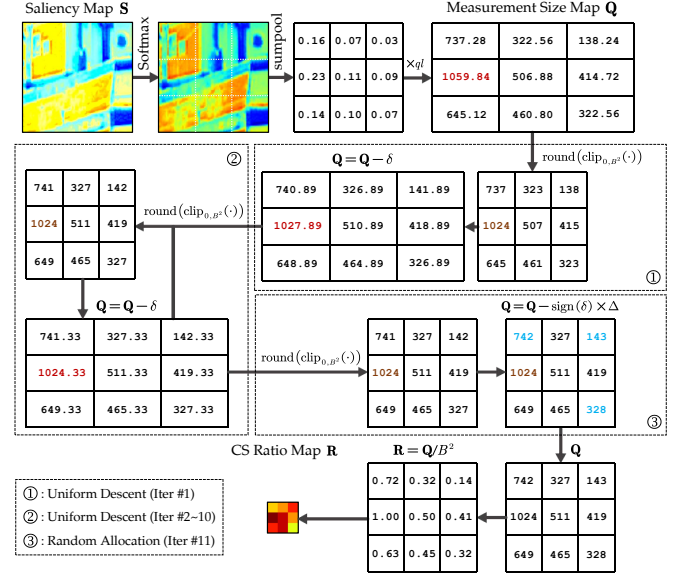


Fig. 19. Illustration of a simple but complete process of an accurate and fast CS ratio allocation performed by our proposed BRA strategy, which distributes the allocated CS ratio map \mathbf{R} with 11 iterations in the case of $B = 32$, $l = 9$, target CS ratio $r = 50\%$ ($q = 512$) and $K = N = 1024$. As we can see, the BRA strategy, which consists of softmax normalization, sumpooling aggregation, and error correction, takes $\{\mathbf{S}, B, q, K\}$ as inputs and gives out the accurately allocated CS ratio map \mathbf{R} to conduct the content-aware samplings of CASNet framework and its practical implementations.

of processing the saliency map \mathbf{S} of a 96×96 image patch and give out the allocated CS ratio map \mathbf{R} with a practical configuration of CS ratio $r = 50\%$, i.e., $B = 32$, $l = 9$, $q = 512$ and $K = N = 1024$. As illustrated in Fig. 19, \mathbf{S} is first normalized by the softmax operator in spatial to obtain the weight map, and initially aggregated by the 32×32 sumpooling to get the weights of all blocks. The aggregated weight map is then timed by the target sum of block measurement sizes 512×9 to obtain the measurement size map \mathbf{Q} . Since the measurement sizes of blocks can only be integers in the range of $[0, 1024]$ and their sum should be 512×9 for accuracy, the initial \mathbf{Q} is sent to acquire corrections iteratively. In each iteration, we first use $\text{round}(\text{clip}_{0,1024}(\cdot))$ to shear and discretize the allocated measurement size of each block independently, then the average error $\delta = \text{average}(\mathbf{Q}) - 512$ determines whether the correction should be performed. Once δ equals zero, then the error correction stage can be stopped and \mathbf{Q} is finally normalized by $\mathbf{R} = \mathbf{Q} / 1024$ to get the CS ratio map \mathbf{R} . Otherwise, \mathbf{Q} needs to be further corrected by our provided two correcting methods. The first method is uniform descent, which directly employs $\mathbf{Q} = \mathbf{Q} - \delta$ to make a fair adjustment. However, adopting the uniform descent only could result in a dead loop in some cases since the remained error may not be eliminated (quantitative analysis is provided in Appx. B), so we provide the second correcting method to randomly distribute the residual distances to blocks based on the multinomial distribution. In our default setting, the uniform descent has a maximum iteration number $T = 10$, which means that if the first method can not eliminate the error in the first ten iterations, the random approach will be utilized for further corrections. The instance in Fig. 19 gives an example of the BRA work principle, which exhibits the details of our two correcting methods for the initial measurement size map.

APPENDIX B

CONVERGENCE ANALYSIS OF BRA STRATEGY

Here we provide convergence analysis of BRA strategy in Alg. 1 to demonstrate its effectiveness. We first show that the first part of BRA: the uniform descent (marked by [method #1](#)) makes CS measurement size map \mathbf{Q} convergent to a fixed point \mathbf{Q}_∞ with a bounded average error $\delta_\infty \in (-1/2, 1/2]$.

We denote t as the decent step index, \mathbf{Q}_t and δ_t as the measurement size map and the average error in the t -th correction iteration (corresponding to the 7-th and 8-th numbered lines of Alg. 1), respectively, and q_t^i as the i -th element of \mathbf{Q}_t . Given any initial measurement size map \mathbf{Q}_0 (corresponding to the 3-rd numbered line), the parameter set $\{l, q, K\}$ (block number, target average measurement size and upper bound), we denote the discretization function $f: \mathbb{R} \mapsto \{0, 1, \dots, K\}$ with $f(\cdot) = \text{round}(\text{clip}_{0,K}(\cdot))$ and provide four lemmas for facilitating the convergence analysis behind them as follows:

Lemma 1 (“Shape convergence” of \mathbf{Q}). For series $\{\mathbf{Q}_t\}_{t=0}^T$, $\forall i \neq j$, it holds that: (1) the larger (or equal) elements in \mathbf{Q}_t remain larger (or equal) in \mathbf{Q}_{t+1} , i.e., $(q_t^i \geq q_t^j) \Rightarrow (q_{t+1}^i \geq q_{t+1}^j)$, and (2) absolute difference between elements is monotonically decreasing, i.e., $|q_t^i - q_t^j| \geq |q_{t+1}^i - q_{t+1}^j|$.

Proof. For (1), since $\forall 1 \leq i \leq l$, $q_{t+1}^i = f(q_t^i - \delta_t)$ and $\forall \delta_t \in \mathbb{R}$, $f(\cdot - \delta_t)$ is monotonically increasing, the order-preservation among map elements holds. For (2), it is obvious that $\forall s \in \mathbb{R}$, $f(s+1) - f(s) \leq 1$ (proof is omitted), we have:

$$[-f(s) \leq 1 - f(s+1)] \Leftrightarrow [s - f(s) \leq (s+1) - f(s+1)].$$

Let $s = x - \delta_t$, then it holds that:

$$\begin{aligned} [x - \delta_t - f(x - \delta_t) \leq (x - \delta_t + 1) - f(x - \delta_t + 1)] \\ \Leftrightarrow [x - f(x - \delta_t) \leq (x + 1) - f(x + 1 - \delta_t)], \end{aligned}$$

i.e., $[(\cdot) - f(\cdot - \delta_t)]$ is increasing on $\{0, 1, \dots, K\}$. Assume that $q_t^i \geq q_t^j$, we have $q_t^i - f(q_t^i - \delta_t) \geq q_t^j - f(q_t^j - \delta_t)$, i.e., $q_{t+1}^i \geq q_{t+1}^j$. Similarly, when $q_t^i \leq q_t^j$, it holds that $q_{t+1}^i \leq q_{t+1}^j$, i.e., $|q_t^i - q_t^j| \geq |q_{t+1}^i - q_{t+1}^j|$. ■

The Lemma 1 indicates that differences between each two \mathbf{Q} elements converge to fixed values, and the whole map may then shift up and down by constants across all elements. Next, we show that $\{\mathbf{Q}_t\}$ will converge and not oscillate like that.

Lemma 2 (Bound for absolute error). $\forall n \in \{0, 1, \dots, K\}$ and $\delta \in \mathbb{R}$, it holds that: $|n - f(n - \delta)| \leq \text{round}(\text{abs}(\delta))$.

Proof. Let $\text{floor}(x)$ and $\text{frac}(x)$ be the decimal and fractional parts of $x \in [0, +\infty)$, respectively. When $\delta \geq 0$, it holds that:

$$\begin{aligned} |n - f(n - \delta)| &= n - f(n - \text{floor}(\delta) - \text{frac}(\delta)) \\ &= \begin{cases} n, & n - \text{floor}(\delta) \leq 0 \\ \text{round}(\delta), & [0 < n - \text{floor}(\delta) \leq K] \wedge [\text{frac}(\delta) \neq 1/2] \\ \text{floor}(\delta), & [0 < n - \text{floor}(\delta) \leq K] \wedge [\text{frac}(\delta) = 1/2] \end{cases}, \end{aligned}$$

and when $\delta \leq 0$, it holds that:

$$\begin{aligned} |n - f(n - \delta)| &= f(n + \text{floor}(-\delta) + \text{frac}(-\delta)) - n \\ &= \begin{cases} \text{round}(-\delta), & 0 \leq n + \text{floor}(-\delta) < K \\ K - n, & n + \text{floor}(-\delta) \geq K \end{cases}. \end{aligned}$$

For all cases, it holds that $|n - f(n - \delta)| \leq \text{round}(\text{abs}(\delta))$. ■

Lemma 3 (Upper bound for $\text{round}(\cdot)$). $\forall \delta \geq 0$, it holds that: $\text{round}(\delta) \leq 2\delta$ and $[\text{round}(\delta) = 2\delta] \Leftrightarrow [\delta \in \{0, 1/2\}]$.

Proof. When $\delta = 0$, we have: $\text{round}(\delta) = 0 = 2\delta$. When $0 < \delta < 1/2$, we have: $\text{round}(\delta) = 0 < 2\delta$. When $\delta = 1/2$, we have: $\text{round}(\delta) = 1 = 2\delta$. When $1/2 < \delta \leq 1$, we have: $\text{round}(\delta) = 1 < 2\delta$. When $1 < \delta < 2$, we have: $\text{round}(\delta) \leq \text{floor}(\delta) + 1 < 2\text{floor}(\delta) + 2\text{frac}(\delta) = 2\delta$. And when $\delta \geq 2$, we have: $\text{round}(\delta) \leq \text{floor}(\delta) + 1 < 2\text{floor}(\delta) \leq 2\delta$. ■

Based on the above three lemmas, we present the following lemma and theorem to show convergency of uniform descent:

Lemma 4 (Convergence of the absolute average error). The absolute average error series $\{|\delta_t|\}_{t=0}^T$ converges as $T \rightarrow \infty$.

Proof. When $\delta_t = 0$, for a specific $t > 0$, $f(\cdot)$ will not change any element value of \mathbf{Q}_t and the latter δ_s will be all zeros (converged). And when $\delta_t \neq 0$, the ratio of $|\delta_{t+1}|$ to $|\delta_t|$ is:

$$\begin{aligned} \left| \frac{\delta_{t+1}}{\delta_t} \right| &= \left| \frac{\frac{1}{l} \sum_{i=1}^l [f(q_t^i - \delta_t) - q]}{\delta_t} \right| \\ &= \left| \frac{\frac{1}{l} \sum_{i=1}^l [f(q_t^i - \delta_t) - q_t^i + q_t^i - q]}{\frac{1}{l} \sum_{i=1}^l (q_t^i - q)} \right| \\ &= \left| 1 - \frac{d_t}{\delta_t} \right|, \text{ where } d_t = \frac{1}{l} \sum_{i=1}^l [q_t^i - f(q_t^i - \delta_t)]. \end{aligned}$$

Using Lemma 2, when $\delta_t > 0$, we have: $d_t \leq \text{round}(\delta_t)$, and when $\delta_t < 0$, we have: $d_t \geq -\text{round}(-\delta_t)$. Using Lemma 3, we have: $0 \leq d_t/\delta_t \leq 2$ and $|\delta_{t+1}/\delta_t| \leq 1$. ■

Theorem 1 (Convergence of measurement size map in the uniform descent). The series $\{\mathbf{Q}_t\}_{t=0}^T$ converges to a fixed point \mathbf{Q}_∞ with average error $\delta_\infty \in (-1/2, 1/2]$ as $T \rightarrow \infty$.

Proof. Following Lemma 1 and Lemma 4, $\exists T_0 \in \mathbb{N}$, s.t. when $t \geq T_0$, $|\delta_t|$ and the differences between elements converge, then we have: $\delta_t = 0$ or $(\delta_t \neq 0) \wedge (|\delta_{t+1}/\delta_t| = 1)$ (i.e., (1) $d_t = 0$ or (2) $d_t = 2\delta_t$). For (1), since $f(\cdot)$ is increasing and $q_t^i \in \{0, 1, \dots, K\}$, $\forall 1 \leq i \leq l$, it holds that: $q_t^i - f(q_t^i - \delta_t) = f(q_t^i) - f(q_t^i - \delta_t) = 0$, i.e., $\mathbf{Q}_t = \mathbf{Q}_{t+1}$. When $\delta_t > 0$, $\exists q_t^i > 0$ (otherwise $\delta_t \leq 0$), so $[\text{round}(\delta_t) = 0] \wedge [\text{frac}(\delta_t) \neq 1/2]$ or $[\text{floor}(\delta_t) = 0] \wedge [\text{frac}(\delta_t) = 1/2]$ (see our Proof for Lemma 2), i.e., $0 < \delta_t \leq 1/2$. Similarly, when $\delta_t < 0$, $\exists q_t^i < K$, so $\text{round}(-\delta_t) = 0$, i.e., $-1/2 < \delta_t < 0$. For (2), using Lemma 1, it holds that: $q_t^i - f(q_t^i - \delta_t) = q_t^i - q_{t+1}^i = 2\delta_t$. When $\delta_t > 0$, using Lemma 2 and Lemma 3, we have: $\text{round}(\delta_t) = 2\delta_t$, i.e., $\delta_t = 1/2$, but it leads $q_t^i - f(q_t^i - \delta_t) = 0$ and is excluded. Similarly, when $\delta_t < 0$, we have: $[-\text{round}(-\delta_t)] = 2\delta_t$, i.e., $\delta_t = -1/2$, but it can also not hold since $q_t^i - q_{t+1}^i = -1$ itself will hold for limited steps. In summary, \mathbf{Q} converges to a fixed \mathbf{Q}_∞ as (1) with a bounded $\delta_\infty \in (-1/2, 1/2]$. ■

Second, after the preset limited T uniform descent steps, the second part of BRA: random allocation (marked by [method #2](#)) will eliminate the remained error and converge since the sign of error is preserved and its absolute value monotonically decreases (proof is omitted). We'll get the accurately allocated measurement size map \mathbf{Q} and CS ratio map \mathbf{R} by our Alg. 1. Furthermore, Fig. 20 visualizes the curves of average error and mean squared error of \mathbf{Q}_{t+1} and \mathbf{Q}_t under two settings, and demonstrates that our BRA strategy converges in 16 steps.

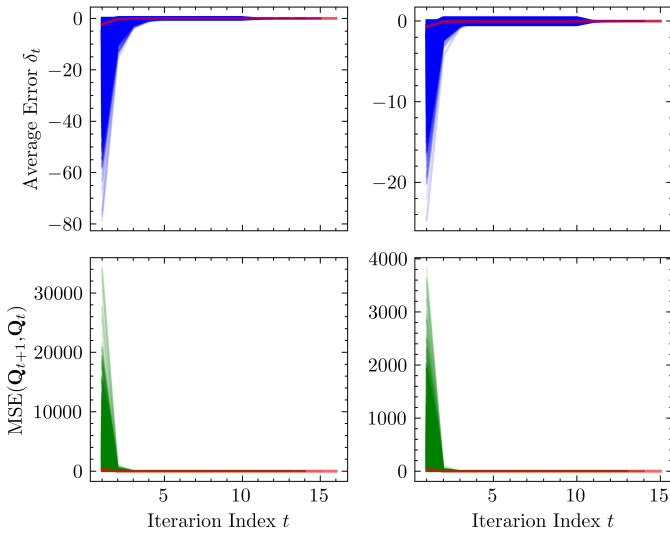


Fig. 20. Visualization of the curves of average error δ_t (top) and mean squared error of \mathbf{Q}_{t+1} and \mathbf{Q}_t (marked by $\text{MSE}(\mathbf{Q}_{t+1}, \mathbf{Q}_t)$, i.e., $(1/l)\|\mathbf{Q}_{t+1} - \mathbf{Q}_t\|_F^2$) (bottom) under two settings of block number: $l = 25$ (left) and $l = 100$ (right) with the practical configurations $B = 32$, $q = 512$ and $K = N = 1024$. The experiment consists of 1×10^5 rounds for each setting with randomly generated initial states. Each blue or green curve corresponds to one our simulated BRA test round, and the red curve corresponds to the mean data values in each subgraph.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candès, J. Emmanuel, and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [3] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [4] F. Rousset, N. Ducros, A. Farina, G. Valentini, C. DAndrea, and F. Peyrin, "Adaptive basis scan by wavelet prediction for single-pixel imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 36–46, 2017.
- [5] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [7] T. P. Szczytkutowicz and G.-H. Chen, "Dual energy CT using slow kVp switching acquisition and prior image constrained compressed sensing," *Physics in Medicine & Biology*, vol. 55, no. 21, p. 6411, 2010.
- [8] L. Gan, "Block compressed sensing of natural images," in *Proceedings of IEEE International Conference on Digital Signal Processing (ICDSP)*, 2007, pp. 403–406.
- [9] I. Y. Chun and B. Adcock, "Compressed sensing and parallel acquisition," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4860–4882, 2017.
- [10] I. Y. Chun and B. Adcock, "Uniform recovery from subgaussian multi-sensor measurements," *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 731–765, 2020.
- [11] J. Zhang, C. Zhao, D. Zhao, and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via L0 minimization," *Signal Processing*, vol. 103, pp. 114–126, 2014.
- [12] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1182–1195, 2016.
- [13] C. Zhao, J. Zhang, S. Ma, X. Fan, Y. Zhang, and W. Gao, "Reducing image compression artifacts by structural sparse representation and quantization constraint prior," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2057–2071, 2016.
- [14] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010, vol. 2, no. 1.
- [15] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparsity analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [16] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.
- [17] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [18] Z. Long, Y. Liu, L. Chen, and C. Zhu, "Low rank tensor completion for multiway visual data," *Signal Processing*, vol. 155, pp. 301–316, 2019.
- [19] Y. Liu, Z. Long, and C. Zhu, "Image completion using low tensor tree rank and total variation minimization," *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 338–350, 2018.
- [20] Y. Liu, Z. Long, H. Huang, and C. Zhu, "Low CP rank and Tucker rank tensor completion for estimating missing components in image data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 944–954, 2019.
- [21] S. Ravishanker, J. C. Ye, and J. A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2019.
- [22] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, 2015, pp. 1336–1343.
- [23] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 449–458.
- [24] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Image compressed sensing using convolutional neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 375–388, 2019.
- [25] Y. Huang, T. Würfl, K. Breininger, L. Liu, G. Lauritsch, and A. Maier, "Some investigations on robustness of deep learning in limited angle tomography," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2018, pp. 145–153.
- [26] Y. Sun, J. Chen, Q. Liu, B. Liu, and G. Guo, "Dual-path attention network for compressed sensing image reconstruction," *IEEE Transactions on Image Processing*, vol. 29, pp. 9482–9495, 2020.
- [27] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of International Conference on Machine Learning (ICML)*, 2010, pp. 399–406.
- [28] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2305–2318, 2018.
- [29] Z. Wang, Q. Ling, and T. S. Huang, "Learning deep L0 encoders," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016, pp. 2194–2200.
- [30] J. Sun, H. Li, Z. Xu et al., "Deep ADMM-Net for compressive sensing MRI," *Advances in Neural Information Processing Systems*, vol. 29, pp. 10–18, 2016.
- [31] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4293–4308, 2017.
- [32] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1828–1837.
- [33] I. Y. Chun, Z. Huang, H. Lim, and J. Fessler, "Momentum-net: Fast and convergent iterative neural network for inverse problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access)*, 2020.
- [34] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [35] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [36] J. Zhang, C. Zhao, and W. Gao, "Optimization-inspired compact deep compressive sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 765–774, 2020.
- [37] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "AMP-Net: Denoising-Based Deep Unfolding for Compressive Image Sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 1487–1500, 2021.

- [38] Y. Yu, B. Wang, and L. Zhang, "Saliency-based compressive sampling for image signals," *IEEE Signal Processing Letters*, vol. 17, no. 11, pp. 973–976, 2010.
- [39] S. Zhou, Y. He, Y. Liu, C. Li, and J. Zhang, "Multi-channel deep networks for block-based image compressive sensing," *IEEE Transactions on Multimedia*, pp. 2627–2640, 2020.
- [40] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Scalable convolutional neural network for image compressed sensing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 290–12 299.
- [41] D. You, J. Zhang, J. Xie, B. Chen, and S. Ma, "COAST: COntrollable Arbitrary-Sampling NeTwork for Compressive Sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 6066–6080, 2021.
- [42] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 3021–3024.
- [43] C. Li, W. Yin, H. Jiang, and Y. Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, 2013.
- [44] Y. Kim, M. S. Nadar, and A. Bilgin, "Compressed sensing using a Gaussian scale mixtures model in wavelet domain," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 3365–3368.
- [45] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.
- [46] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.
- [47] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Image compressive sensing recovery via collaborative sparsity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 380–391, 2012.
- [48] C. Zhao, J. Zhang, S. Ma, and W. Gao, "Nonconvex Lp nuclear norm based ADMM framework for compressed sensing," in *Proceedings of IEEE Data Compression Conference (DCC)*, 2016, pp. 161–170.
- [49] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.
- [50] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3929–3938.
- [51] J. H. R. Chang, C.-L. Li, B. Póczos, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "One Network to Solve Them All—Solving Linear Inverse Problems Using Deep Projection Models," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5888–5897.
- [52] C. Zhao, J. Zhang, R. Wang, and W. Gao, "CREAM: CNN-REGularized ADMM framework for compressive-sensed image reconstruction," *IEEE Access*, vol. 6, pp. 76 838–76 853, 2018.
- [53] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [54] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.
- [55] I. Y. Chun and J. A. Fessler, "Convolutional analysis operator learning: Acceleration and convergence," *IEEE Transactions on Image Processing*, vol. 29, pp. 2108–2122, 2019.
- [56] W. Shi, F. Jiang, S. Zhang, and D. Zhao, "Deep networks for compressed image sensing," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 877–882.
- [57] Y. Yu, B. Wang, and L. Zhang, "Hebbian-based neural networks for bottom-up visual attention systems," in *Proceedings of International Conference on Neural Information Processing (ICONIP)*, 2009, pp. 1–9.
- [58] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Proceedings of Neural Information Processing Systems (NeurIPS)*, 2019, pp. 8024–8035.
- [59] A. Mousavi and R. G. Baraniuk, "Learning to invert: Signal recovery via deep convolutional networks," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2272–2276.
- [60] Lefkimmiatis, Stamatios, "Universal denoising networks: a novel CNN architecture for image denoising," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3204–3213.
- [61] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [62] Geman, Donald and Yang, Chengda, "Nonlinear Image Recovery with Half-quadratic Regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.
- [63] K. Zhang, Y. Li, W. Zuo, L. Zhang, V. G. Luc, and T. Radu, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access)*, 2021.
- [64] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [65] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
- [66] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided CNN for image denoising," *Neural Networks*, vol. 124, pp. 117–129, 2020.
- [67] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2014, pp. 184–199.
- [68] K. Zhang and W. Zuo and Y. Chen and D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [70] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2015.
- [71] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 2001, pp. 416–423.
- [72] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [73] R. Mdrafi and A. C. Gurbuz, "Joint learning of measurement matrix and signal reconstruction via deep learning," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 818–829, 2020.



Bin Chen received the B.E. degree in the School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China, in 2021. He is currently working toward the master's degree in computer applications technology at Peking University Shenzhen Graduate School, Shenzhen, China. His research interests include compressive sensing, image restoration and computer vision.



Jian Zhang (M'14) received the B.S. degree from the Department of Mathematics, Harbin Institute of Technology (HIT), Harbin, China, in 2007, and received his M.Eng. and Ph.D. degrees from the School of Computer Science and Technology, HIT, in 2009 and 2014, respectively. From 2014 to 2018, he worked as a postdoctoral researcher at Peking University (PKU), Hong Kong University of Science and Technology (HKUST), and King Abdullah University of Science and Technology (KAUST).

Currently, he is an Assistant Professor with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen, China. His research interests include intelligent multimedia processing, deep learning and optimization. He has published over 90 technical articles in refereed international journals and proceedings. He received the Best Paper Award at the 2011 IEEE Visual Communications and Image Processing (VCIP) and was a co-recipient of the Best Paper Award of 2018 IEEE MultiMedia.