

# Deep-Learned Regularization and Proximal Operator for Image Compressive Sensing

Zan Chen, Wenlong Guo, Yuanjing Feng, Yongqiang Li, Changchen Zhao, Yi Ren, Ling shao, *Fellow, IEEE*,

**Abstract**—Deep learning has recently been intensively studied in the context of image compressive sensing (CS) to discover and represent complicated image structures. These approaches, however, either suffer from nonflexibility for an arbitrary sampling ratio or lack an explicit deep-learned regularization term. This paper aims to solve the CS reconstruction problem by combining the deep-learned regularization term and proximal operator. We first introduce a regularization term using a carefully designed residual-regressive net, which can measure the distance between a corrupted image and a clean image set and accurately identify to which subspace the corrupted image belongs. We then address a proximal operator with a tailored dilated residual channel attention net, which enables the learned proximal operator to map the distorted image into the clean image set. We adopt an adaptive proximal selection strategy to embed the network into the loop of the CS image reconstruction algorithm. Moreover, a self-ensemble strategy is presented to improve CS recovery performance. We further utilize state evolution to analyze the effectiveness of the designed networks. Extensive experiments also demonstrate that our method can yield superior accurate reconstruction (PSNR gain over 1 dB) compared to other competing approaches while achieving the current state-of-the-art image CS reconstruction performance. The test code is available at <https://github.com/zjut-gwl/CSDRCANet>.

## I. INTRODUCTION

The theory of compressive sensing (CS) has drawn considerable research interest as a joint sampling and compression approach [1]. CS indicates that a sparse or compressible high-dimensional signal can be reconstructed from a limited number of measurements by utilizing prior knowledge [2]. As CS can reduce the amount of information to be observed and processed while maintaining a reasonable reconstruction of the sparse or compressible signal, it has been widely used in applications such as medical imaging [3], image compression [4], single-pixel cameras [5], and snapshot compressive imaging [6].

Since CS reconstruction is an ill-posed problem, reliable prior information must be used to constrain the solution space [7]. Traditional CS focuses on hand-crafted regularization, in which prior information comes from years of experience. For instance, nonlocal prior models dominated

the field of CS reconstruction before deep-based models became popular. [8] proposes a hybrid structural nonlocal model that exploits nonlocal self-similarity priors from both the internal and external image corpora. [9] further applies a Gaussian mixture model to learn mutually complementary information from external and internal nonlocal similar priors. [10] integrates the nonlocal self-similarity priors along with self-supervised learning, which can reduce the group sparsity residual. These hand-crafted CS reconstruction methods are usually solved by forming an optimization problem that often has theoretical convergence guarantees. However, these hand-crafted prior approaches often lead to unsatisfactory results, especially in cases with low sampling ratios, as they suffer due to ignoring collected data information.

To address the above shortcomings in traditional hand-crafted approaches, a recent trend in CS reconstruction is to take advantage of deep neural networks to discover and represent complicated image structures. Generally, these deep-based methods fall into two categories: end-to-end approaches and plug-and-play approaches. Specifically, the basic idea of end-to-end approaches is to directly learn a network that maps CS measurements into the original signals. Some advanced end-to-end approaches are designed based on the unfolding of some optimization iterative algorithms onto deep neural networks. For instance, ISTA-Net+ [11] replaces the soft-thresholding step in the traditional iterative soft-thresholding algorithm with a learning-based threshold operator. ADMM-CSNet [12] casts the iterative ADMM algorithm for sparse regularization into a deep architecture. AMP-Net [13] solves the image CS problem by unfolding the iterative denoising process of the approximate message passing algorithm. Such end-to-end approaches have interpretability while still being able to reconstruct images quickly.

However, end-to-end deep-based CS reconstruction methods suffer from non-flexibility, as they need to train different models for different sampling ratios. A trained end-to-end model will not work well if some elements of measurements are lost directly, meaning it is inapplicable for adaptive sensing [14]. As shown in Fig. 1, a model-fixed end-to-end approach cannot reconstruct the image accurately once the measurement loss rate increases to some level. SCSNet [15] achieves a scalable CS end-to-end net by using a greedy strategy to search the most important measurement bases. However, SCSNet still needs to update the network parameters for different sampling ratios, and the complexity of greedy searching is no less than that of retraining the model at a high sampling ratio. Except for non-flexibility, to deal with different image resolutions and ease the computational burden, end-to-end methods sample

This research was sponsored in part by the National Natural Science Foundation of China (Grant Nos. 62002327, 61976190, 61903336), Natural Science Foundation of Zhejiang Province (Grant No. LQ21F020017), and the Key Research and Development Program of Zhejiang Province (Grant No. 2020C03070). (Corresponding author: Yuanjing Feng.)

Z. Chen, W. Guo, Y. Feng, Y. Li, and C. Zhao are with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China email: zanchen2@zjut.edu.cn.

Y. Ren is with the School of Computing Sciences, University of East Anglia, Norwich, UK.

L. Shao is with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates.

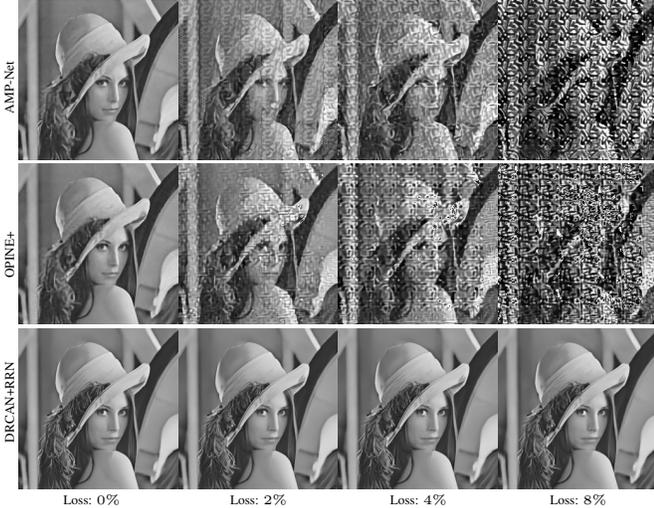


Fig. 1. Illustration of *Lena* reconstructed by AMP-Net, OPINE-Net+ and the proposed method if randomly losing some measurements at a 0.10 sampling ratio.

and reconstruct images block-by-block, requiring deblocking modules to eliminate the blocking artifacts, which may lead to poor visual quality.

In addition to deep end-to-end approaches, deep plug-and-play approaches have also been extensively studied, which replace some components in traditional hand-crafted prior methods with deep neural networks [16]. These deep-based plug-and-play approaches have the advantages of both the interpretability of hand-crafted prior methods and the powerful feature expression ability of deep neural networks, maintaining the measurement fidelity term while reducing the regularization term efficiently [17]. For example, LDAMP [18] utilizes the denoising convolutional neural network (DnCNN) [19] in place of the traditional image denoising step in D-AMP algorithms, and has achieved state-of-the-art CS reconstruction performance [20], while [21] presents a hybrid plug-and-play framework for CS reconstruction that combines a deep image Gaussian denoiser with nonlocal low-rank priors. The performance of deep plug-and-play approaches depends mainly on the regularization term and the corresponding proximal operators. As shown in Fig. 2, simply replacing the DnCNN in LDAMP with other denoisers can lead to different CS reconstructed PSNRs. Thus, pairing a well-designed deep-learned regularization and the corresponding proximal operator can improve the CS reconstruction performance.

This paper aims to design an effective CS reconstruction method using deep-learned regularization and a proximal operator. We exploit the elaborately designed residual-regressive net (RRN) and the dilated residual channel attention net (DRCAN) to simulate the regularization term and proximal operator, respectively. We then embed two designed neural networks into the loop of the proximal gradient descent algorithm: RRN as the regularization term for noise-level estimation and DRCAN as a proximal operator for image denoising, in which the self-ensemble strategy can further enhance the reconstruction performance. From the state evolution analysis, we see that our efficient noise-level estimator and denoiser can

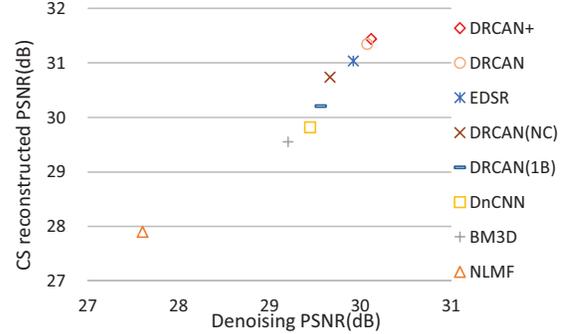


Fig. 2. Illustration of different denoising performances and the corresponding CS reconstructed performances. The X-coordinate refers to the average denoising PSNRs of images with white noise of variance 25, and the Y-coordinate refers to the average CS reconstructed PSNRs at a 0.10 sampling ratio. The results are evaluated based on the union of Set8 and Waterloo140.

improve the final CS reconstruction performance. Extensive experiments also show that the proposed method can yield much better CS reconstruction results than the current state-of-the-art methods in terms of both quantitative metrics and subjective visual quality. In summary, the contributions of this work are as follows:

- We introduce a tailored residual-regressive network for regularization, which can measure the noise level of the reconstructed image accurately.
- We design a dilated residual channel attention network as the proximal operator, which can efficiently denoise a corrupted image.
- We adopt an adaptive proximal operator selection strategy to embed the designed regularization and proximal operator into the proximal momentum-gradient descent algorithm.
- We utilize state evolution to analyze the effectiveness of the designed networks, and our experimental results demonstrate that the proposed method achieves promising performance.

The remainder of this paper is organized as follows. Section II introduces the background of the image CS optimization problem. Section III presents the architectures of our designed networks for the deep-learned regularization term and proximal operator. Property analysis of the denoiser and noise-level estimator is presented in Section IV, followed by performance evaluation in Section V. Section VI presents our conclusions.

## II. BACKGROUND

### A. Image CS Problem

Image CS is used to reconstruct a clean image  $x \in \mathbb{R}^n$  only from its  $m$  ( $m \ll n$ ) randomized linear observations  $y \in \mathbb{R}^m$ , i.e.,

$$y = \Phi x, \quad (1)$$

where  $\Phi$  is a short-fat sensing matrix satisfying the mutual coherence property. Since the sensing matrix  $\Phi$  is rank-deficient, there exist an infinite number of feasible solutions satisfying Eq. 1, which makes such underdetermined systems hard to solve [22]. Such a challenging ill-posed inverse problem requires some prior information about the image to

constrain the solution space, which can be represented as an optimization problem with the following form:

$$x = \arg \min_x F(x) + \lambda G(x), \quad (2)$$

where  $F(x)$  and  $G(x)$  are respectively the fidelity term and regularization term, and where  $\lambda$  is a regularization parameter balancing the contributions of the two terms. The fidelity term ensures that the possible solution is consistent with the CS measurement process, and Euclidean distance in the measurement domain is the usual choice, i.e.,

$$F(x) = \frac{1}{2} \|y - \Phi x\|_2^2. \quad (3)$$

The regularization term  $G(x)$  is used to guarantee that the possible solution satisfies the prior information. Designing and exploiting the regularization term are the main challenges when using the image CS reconstruction algorithm [23].

### B. Regularization Term for Image CS

The basic idea of existing image CS algorithms is to design robust image regularization terms, which can integrate information loss in undersampling measurements. In general, existing studies fall into two categories: hand-crafted regularization and deep-learned regularization.

Popular hand-crafted regularizations for image CS include sparsity-based  $l_1$ -norm ( $\|\cdot\|_1$ ), gradient-based TV-norm ( $\|\cdot\|_{TV}$ ), and nonlocal self-similarity-based rank-norm ( $\|\cdot\|_*$ ). Various optimization algorithms can be applied to solve the CS optimization problem with hand-crafted regularization, such as the iterative shrinkage algorithm [24], approximate message passing (AMP) method [25], and Douglas-Rachford splitting method [26]. BM3D-CS [27] and NLR-CS [28] are usually taken as benchmark methods due to their high performance for CS image reconstruction. However, these hand-crafted methods require heavy computation and usually involve some manually chosen parameters when they are hard to determine.

In addition to traditional hand-crafted regularization, recent work has shown that better empirical performance is achieved when deep neural networks are applied. Deep-learned regularization achieves better performance due to its ability to learn realistic image priors from a large amount of training data [29]. In general, the deep deep -learned priors are contained in the designed network architecture and trained weights [30]. One popular strategy is to maintain the fidelity term while replacing hand-crafted regularization components with neural networks that map from corrupted image space to clean image space [31]. In this way, image CS methods such as LDAMP [18] can efficiently exploit deep-learned prior information. A well-designed deep-learned regularization can help to train the relative proximal operator and enhance the explainability of the deep-learned image CS algorithms.

### C. Proximal Momentum-Gradient Descent for CS problem

1) *Proximal Gradient Descent*: Proximal gradient descent (PDG) offers a general framework for solving the optimization problem represented in Eq. 2, which decouples the fidelity

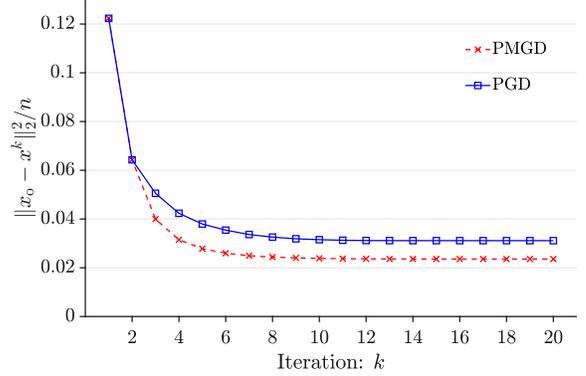


Fig. 3. Convergence with iteration  $k$  of PGD and PMGD. The results are averaged over 8 test images from Set8 at a 0.10 sampling ratio.

term and regularization term by alternating between the projecting step and gradient descent step. Starting from  $x^0$  and adopting a step size  $\alpha$ , the overall iterative procedure of PGD can be expressed as

$$v^k = \nabla F(x^k), \quad (4)$$

$$x^{k+1} = \text{Prox}(x^k - \alpha v^k), \quad (5)$$

where  $\text{Prox}(\tilde{x})$  finds  $\bar{x} \in C$  such that  $\|\tilde{x} - \bar{x}\|$  is minimized, and the gradient of  $F(\cdot)$  at point  $x^k$  can be expressed as

$$\nabla F(x^k) = \Phi^T(\Phi x^k - y). \quad (6)$$

This two-step splitting method decouples the proximal operator from the specifics of CS reconstruction iteration, which is a suitable plug-and-play framework combining deep networks, as it does not need to train different networks for different CS sampling ratios. Such a CS framework allows people to use different proximal operators in the iteration resolving process.

2) *Proximal Momentum-Gradient Descent*: Approximate message passing (AMP) [32] introduces an extra momentum item into Eq. 4 to accelerate PGD, i.e.

$$v^k = \gamma^{k-1} v^{k-1} + \nabla F(x^k), \quad (7)$$

$$x^{k+1} = \text{Prox}(x^k - \alpha v^k). \quad (8)$$

The momentum term  $\gamma^{k-1} v^{k-1}$  gives gradient descent a short-term memory by adding a fraction of the update vector of the past time step to the current update vector [33].

The parameter  $\gamma^{k-1}$  defines the amount of momentum, which balances the effect of local gradient  $\nabla F(x^k)$  on the iteration process and prevents the iteration from becoming trapped in a shallow local minimum. One can update  $\gamma^k$  using the divergence of  $\text{Prox}(\cdot)$  at point  $x^k - \alpha v^k$  [25], i.e.,

$$\gamma^k = \frac{1}{m} \nabla \cdot \text{Prox}(x^k - \alpha v^k). \quad (9)$$

Based on [34], the divergence term can be estimated with a fast Monte Carlo approximation method. The divergence of  $\text{Prox}$  at any point  $\tilde{x}$  can be calculated by

$$\nabla \cdot \text{Prox}(\tilde{x}) = \epsilon^T (\text{Prox}(\tilde{x} + \epsilon) - \text{Prox}(\tilde{x})), \quad (10)$$

where  $\epsilon \sim N(0, 1)$  is a standard normal random vector. As shown in Fig. 3, the proximal momentum-gradient descent (PMGD) framework achieves faster and more accurate

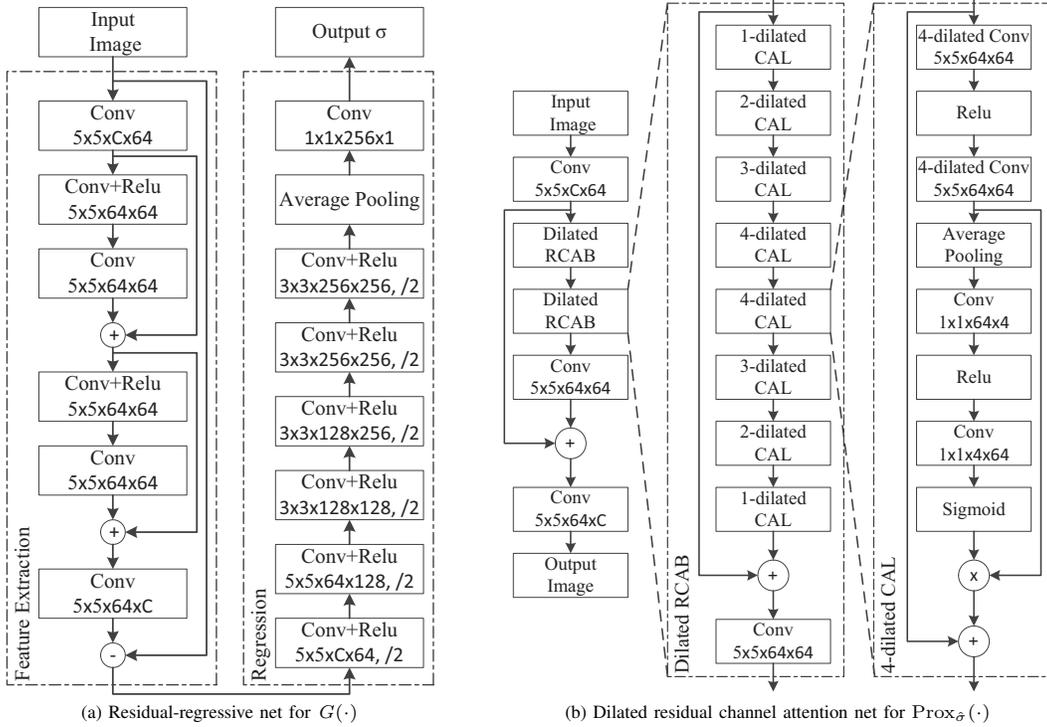


Fig. 4. The architecture of residual-regressive net as regularization and dilated residual channel attention net as proximal operator.

convergence results than the PGD framework. In this paper, we included two types of deep networks within the loop of the PMGD to simulate the regularization term and proximal operator.

### III. DEEP-LEARNED REGULARIZATION AND PROXIMAL OPERATOR FOR IMAGE CS RECONSTRUCTION

In this section, we first present an adaptive proximal operator selection strategy. Next, we describe the architectures of the residual-regressive net (RRN) and the dilated residual channel attention network (DRCAN), which are associated with the learned regularization and the corresponding proximal operator, respectively. Furthermore, we describe a self-ensemble strategy to enhance CS reconstruction performance. Finally, we detail the training strategy for the RRN and the DRCAN.

#### A. Adaptive Proximal Operator selection

We apply an adaptive proximal operator selection strategy to enhance flexibility in the applicable sampling ratios. Specifically, we divide the 2-dimensional image space into multiple subspaces according to the distortion distance estimated by  $G(\cdot)$ . In PMGD, we select the corresponding proximal operation  $\text{Prox}(\cdot)$  according to the subspace to which the intermediate image  $(x^k - \alpha v^k)$  belongs (see Fig. 5). That is, the distortion distance  $\hat{\sigma}$  obtained by  $G(\cdot)$  determines the weights of the proximal operator, i.e.,  $\text{Prox}(\cdot) = P_{\hat{\sigma}}(\cdot)$ .

Accurate distortion distance estimation is important for determining the weights of the deep network [35]. For the learned regularization term, we use a well-trained residual-regressive network that can accurately determine the distortion distance  $\sigma$ . Additionally, the proximal operator plays an important

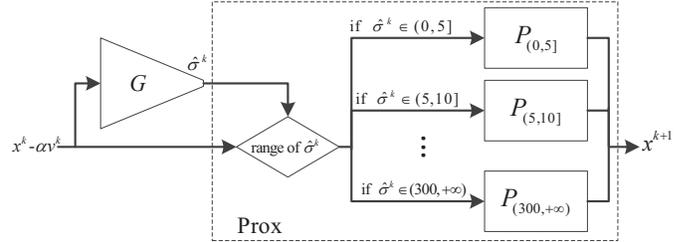


Fig. 5. Adaptive selection of proximal operator  $\text{Prox}(\cdot)$  based on deep-learned regularization  $G(\cdot)$  term.

role in the PMGD algorithm. The learned proximal operator resembles the projection on the clean image manifold; this can be interpreted as a denoiser that removes aliasing artifacts [36]. In this work, we train both the regularization term  $G(\cdot)$  and the corresponding multiple off-the-shelf proximal operators  $P_{\hat{\sigma}}(\cdot)$ .

#### B. RRN for Learned Regularization

We set the noise level as the regularization term and design a residual-regressive net to measure the noise level  $\sigma^k$  of the corrupted image. Specifically, we assume that the original image  $x_o$  lies in a set  $C$ , which can be intuitively thought of as the manifold of the pristine images. Considering any corrupted image  $\tilde{x}$  outside the set  $C$ , we use a deep net  $G(\cdot)$  measuring a certain kind of distortion distance between  $\tilde{x}$  and pristine image set  $C$ .

Our noise-level estimation network is composed of two major parts: the residual operation and the regressive operation. Supposing  $w$ ,  $h$  and  $c$  to be the width, height, and channel

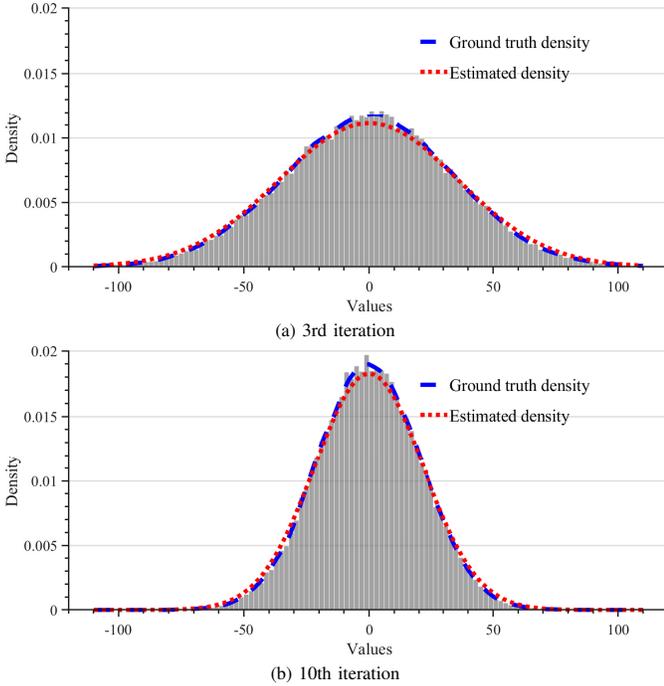


Fig. 6. Histograms of residual coefficients  $x_o - (x^k - \alpha v^k)$  for the “Boat” image at a 0.10 sampling ratio at the 3rd and 10th iterations ( $k = 3$  and 10) of PMGD.

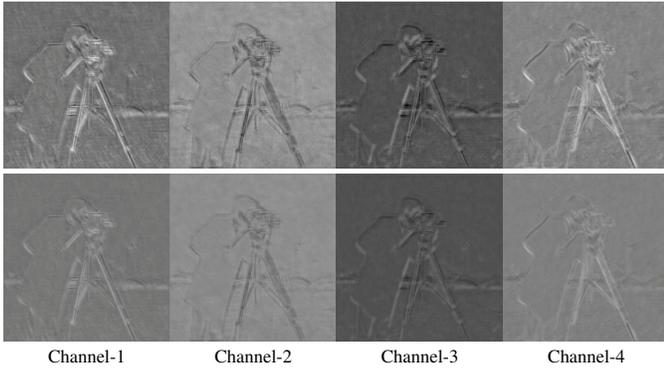


Fig. 7. Intermediate features before (top) and after (below) channel attention operation. Features are selected from the 2nd CAL of the 1st RCAB in DRCAN.

of the input image, respectively, our noise-level estimation network can be formulated as

$$\hat{\sigma}^k = R_1(R_2((x^k - \alpha v^k)) - (x^k - \alpha v^k)). \quad (11)$$

where  $R_2(\cdot)$  refers to two stacked residual blocks that extract a  $w \times h \times 64 \times 64$  size feature and  $R_1(\cdot)$  refers to the mapping of the extracted feature to the regularization value through 6 convolutions with a stride of 2, one average pooling, and a final fully connected layer (see Fig. 4 (a)).

Fig. 6 illustrates the accuracy of our noise-level estimator, where the ground truth residual coefficients are calculated by  $x_o - (x^k - \alpha v^k)$ , the ground-truth density is calculated by  $\sqrt{\|x_o - (x^k - \alpha v^k)\|/n}$ , and the estimated density is calculated by  $G(x^k - \alpha v^k)$ . As seen by comparing histograms, our estimated density is a good fit for the ground-truth residual density.

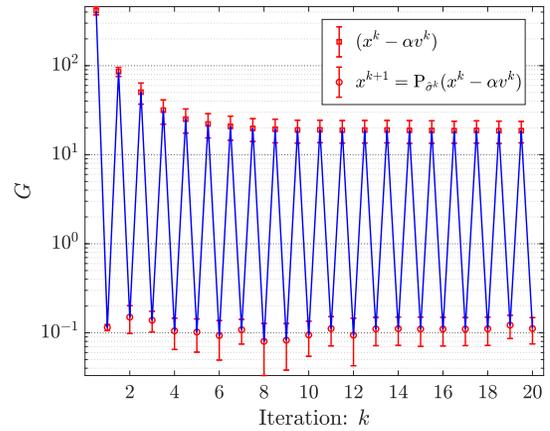


Fig. 8. Regularization term value  $G$  before and after the proximal operation with iteration  $k$  of PMGD. The results are averaged over 8 test images from Set8 at a 0.10 sampling ratio.

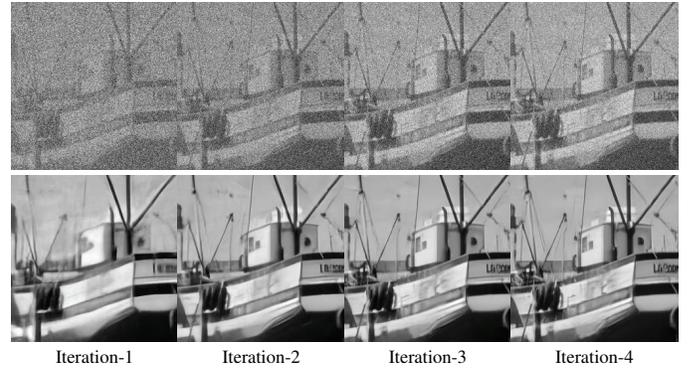


Fig. 9. Intermediate results before (top) and after (below) the  $P_{\hat{\sigma}}(\cdot)$  operation, i.e., top:  $x^k - v^k$ ; bottom:  $x^{k+1} = P_{\hat{\sigma}}(x^k - v^k)$ .

### C. DRCAN for Learned Proximal Operator

The core function of PMGD is to map the distorted image to the pristine image set  $C$ . Based on the learned noise-level estimator  $G$ , we can learn a projection function  $P$  that maps a corrupted image to the pristine image set  $C$ .

In our deep network, we take advantage of some recent progress in the field. Specifically, we adopt into the design a channel attention mechanism, dilated convolution, and multiple skip connections. The architecture of the proposed network for the proximal operator is illustrated in Fig. 4(b), and is primarily composed of sixteen stacked dilated channel attention layers (dilated CAL) and multiple skip connections.

In convolution layers, each channel-wise feature represents a different component of the signal extracted by the corresponding filter. Some channels focus on the complanate regions, while some focus more on the texture or edge regions. An intuitive strategy is to adjust the weights of channel-wise features adaptively instead of treating them equally. The channel attention mechanism allows the net to effectively exploit the interdependencies among feature channels. Denoting  $f_i$  as the input feature of the  $i$ -th CAL, the CAL operation can be formulated as

$$f_{i+1} = f_i + (R_i \circ C_i)(f_i) \cdot C_i(f_i), \quad (12)$$

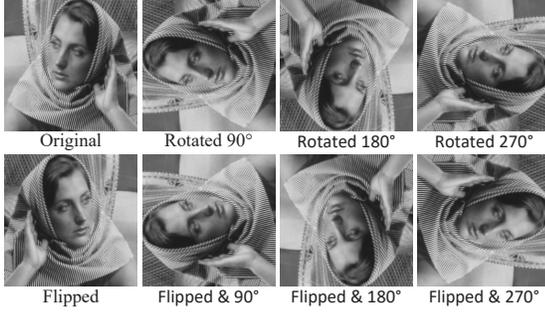


Fig. 10. Self-ensemble by rotation and flip.

where  $C_i$  represents stacked dilated convolution layers to extract intermediate features, and  $R_i$  represents a series of downsampling operations to extract the channel-wise rescaling factor. Fig. 7 shows the intermediate features before and after the channel attention operation, from which we can see that CAL can suppress some noise-like structures. Thus, we employ the channel attention mechanism to enhance the feature extraction ability of the network.

Furthermore, eight dilated CALs group into a dilated residual channel attention block (DRCAB), in which the relative dilation factors of dilated CALs are set to 1, 2, 3, 4, 4, 3, 2, and 1. Dilated convolution can expand the capacity of the receptive field while not increasing the number of filter weights. There exists a short skip connection in each dilated CAL, a medium skip connection in each dilated RCAB, and a long skip connection from the initial to the end of the whole net. Such a recursive residual connection design allows multiple pathways through which information can flow.

Fig. 8 shows the effectiveness of the designed DRCAN. It can be seen that the value of the regularization term  $G$  falls to a small value after using  $P_{\hat{\sigma}^k}(\cdot)$ , which conforms to the function of the proximal operation. Fig. 9 also shows the reconstructed images before and after DRCAN for the first four intermediate iterations, from which we can observe that the top ones have significant noise, while those below are increasingly closer to the real image as iteration increases.

#### D. Self-Ensemble Strategy

To enhance the potential CS reconstruction performance of our model, we further adopt the self-ensemble strategy, which is widely used in single image superresolution [37]. We apply rotations and flips on the image to generate an additional seven augmented inputs  $T_i(x_t)$ , where  $T_i$  represents geometric transformations, as shown in Fig. 10. We then apply the deep networks  $P_{\hat{\sigma}^k}(\cdot)$  on each  $T_i(x_t)$  to obtain eight corresponding outputs. After that, we apply an inverting transformation  $T^{-1}(\cdot)$  on the eight denoised outputs and average the inverting transformed outputs together for the final self-ensemble result. Thus, if adopting the self-ensemble strategy in PGMD, the relevant Eq. 4 should be revised to become

$$x^{k+1} = \frac{1}{8} \sum_{i=1}^8 (T_i^{-1} \circ P_{\hat{\sigma}^k} \circ T_i)(x^k - \alpha v^k). \quad (13)$$

TABLE I  
RANGE OF DISTORTION DISTANCE FOR DIFFERENT SUBSPACE NUMBER

No.	Range of $\sigma$				
2	(0,100]	(100, + $\infty$ )			
10	(0,10]	(10,20]	(20,40]	(40,60]	(60,80]
	(80,100]	(100,150]	(150,300]	(300,500]	(500, + $\infty$ )
17	(0,5]	(5,10]	(10,15]	(15,20]	(20,30]
	(30,40]	(40,50]	(50,60]	(60,70]	(70,80]
	(80,90]	(90,100]	(100,125]	(125,150]	(150,300]
	(300,500]	(500, + $\infty$ )			

#### E. Training Strategy

In our model, image space is partitioned into multi-subspaces, where each subspace has its own learned proximal operator. We first train the deep-learned regularization  $G(\cdot)$  estimating noise level, after which we train the learned proximal operator for each subspace. We corrupt the original clean images  $x_o$  in the training set  $C$  by adding different noise level ranges  $\sigma$  and adopt the mean square error (MSE) losses

$$G = \arg \min_G E_{(x_o, \epsilon, \tilde{\sigma})} \|\tilde{\sigma} - G(x_o + \tilde{\sigma}\epsilon)\|^2, \quad (14)$$

and

$$P_\sigma = \arg \min_{P_\sigma} E_{(x_o, \epsilon)} \|x_o - P_\sigma(x_o + \sigma\epsilon)\|^2, \quad (15)$$

to train  $G(\cdot)$  and  $P_\sigma(\cdot)$ , respectively, where  $x_o \in C$ ,  $\epsilon \sim N(0, 1)$ , and  $\tilde{\sigma} \sim U(0, 600)$ . Table I shows the noise range  $\sigma$  designed for a different number of subspaces, where the noise range is gradually refined as the subspace number increases. In the process of training the proximal operator, we initialize the network weights for the refined noise range using the well-trained network weights for the relatively coarse noise range.

#### IV. PROPERTY ANALYSIS

Taking the designed DRCAN as the proximal operator and utilizing the designed RRN to adaptively select the proximal operator, the PGMD recursion process represented in Eq. 7 and Eq. 8 can be formulated as

$$v^k = \gamma^{k-1} v^{k-1} + \nabla F(x^k) \quad (16)$$

$$\hat{\sigma}^k = G(x^k - \alpha v^k) \quad (17)$$

$$x^{k+1} = P_{\hat{\sigma}^k}(x^k - \alpha v^k). \quad (18)$$

From Eq. 16 to Eq. 18, one can see that the final reconstruction performance is affected mainly by two operations, the noise-level estimator  $G(\cdot)$  and proximal operator  $P_\sigma(\cdot)$ , which are also the components we simulate with deep networks.

#### A. State Evolution

We utilize state evolution (SE) to describe the dynamical behavior of the recursion process described in Eq. 16 to Eq. 18.

Taking  $x_o$  as the original image and defining the errors of the  $k$ -th iterative results as

$$q^k = x^k - x_o \quad (19)$$

$$h^k = x^k + v^k - x_o, \quad (20)$$

we can obtain the error recursion formulated as

$$h^k = (I - \Phi\Phi^T)q^k + \gamma^{k-1}(h^{k-1} - q^{k-1}), \quad (21)$$

$$\hat{\sigma}^k = G(x_o + h^k), \quad (22)$$

$$q^{k+1} = P_{\hat{\sigma}^k}(x_o + h^k) - x_o, \quad (23)$$

which provides a convenient means for analysis of the intermediate MSE. Let  $\theta^k$  and  $\sigma^k$  be the standard deviations of  $q^k$  and  $h^k$ , respectively. Then if  $\Phi$  has i.i.d. entries and  $m, n \rightarrow \infty$ , SE refers to the following recursion process [25], [38],

$$(\sigma^k)^2 = \frac{n}{m}(\theta^{k-1})^2, \quad (24)$$

$$\hat{\sigma}^k = G(x_o + \sigma^k \epsilon) \quad (25)$$

$$(\theta^k)^2 = \frac{1}{n} E_\epsilon \{ \|P_{\hat{\sigma}^k}(x_o + \sigma^k \epsilon) - x_o\|^2 \}, \quad (26)$$

where  $\epsilon \sim N(0, 1)$  is independent of  $x_o$ . In other words, the empirical intermediate MSE can be estimated by SE [39], i.e.,

$$\frac{1}{n} \|P_{\hat{\sigma}^k}(x_o + h^k) - x_o\|^2 \approx \frac{1}{n} E_\epsilon \{ \|P_{\hat{\sigma}^k}(x_o + \sigma^k \epsilon) - x_o\|^2 \}. \quad (27)$$

### B. SE with ideal noise level

We first define the denoising level,  $\mu(P_{\hat{\sigma}}, \sigma)$ , of denoiser  $P_{\hat{\sigma}}(\cdot)$  for a certain noise level  $\sigma$ ,

$$\sup_{x_o \in C} \frac{E_\epsilon \|P_{\hat{\sigma}}(x_o + \sigma \epsilon) - x_o\|^2}{n\sigma^2} = \mu(P_{\hat{\sigma}}, \sigma), \quad (28)$$

where  $C$  is a set of natural images.

Assume that we can obtain the ground-truth value  $\sigma^k$  of each intermediate result  $(x_o + h^k)$  through an ideal noise-level estimator, and that we can tune or choose the parameters of the denoiser based on  $\sigma^k$ . Then, according to Eq. 26 and Eq. 28, we have

$$(\theta^k)^2 = \frac{1}{n} E_\epsilon \{ \|P_{\sigma^k}(x_o + \sigma^k \epsilon) - x_o\|^2 \} \leq \mu^k \cdot (\sigma^k)^2, \quad (29)$$

where  $\mu^k = \mu(P_{\sigma^k}, \sigma^k)$  is the denoising level of denoiser  $P_{\sigma^k}(\cdot)$  for noise level  $\sigma^k$ . Further, substituting Eq. 24 into Eq. 29, we have

$$(\theta^k)^2 \leq \frac{n}{m} \mu^k \cdot (\theta^{k-1})^2 \leq \left(\frac{n}{m}\right)^2 (\mu^k \cdot \mu^{k-1}) \cdot (\theta^{k-2})^2. \quad (30)$$

Then, it is clear that

$$(\theta^k)^2 \leq \left(\frac{n}{m}\right)^k \left(\prod_{i=0}^{k-1} \mu^i\right) (\sigma^0)^2 = \left(\frac{n}{m}\right)^k \left(\prod_{i=0}^{k-1} \mu^i\right) \|\Phi^T y\|^2, \quad (31)$$

if initializing  $x^0$  with zero vector. From Eq. 31, we can see that a smaller denoising level  $\mu^i$  can lead to a smaller upper bound on the reconstructed MSE, guaranteeing the accuracy of the final reconstructed results.

### C. SE with estimated noise level

For the real implementation of the iterative process, we cannot obtain the ground-truth value of the noise level. However, the following two properties still hold for the iterative process using an estimated noise level:

- 1) Better noise-level estimation leads to better CS recovery;
- 2) Better denoiser leads to better CS recovery.

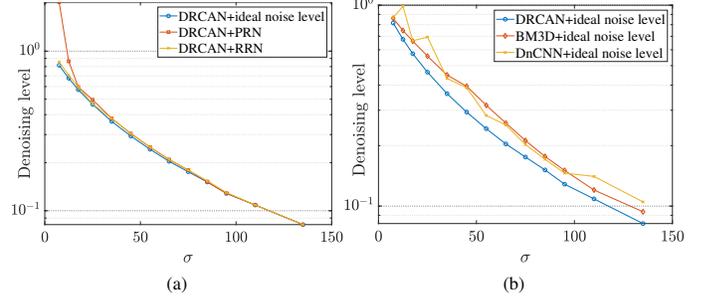


Fig. 11. Denoising level curve comparisons. The X-coordinate represents noise level  $\sigma$ , and the Y-coordinate represents the denoising level calculated by  $\max_{x_o \in C} \frac{E_\epsilon \|P_{\hat{\sigma}}(x_o + \sigma \epsilon) - x_o\|^2}{n\sigma^2}$ , where  $C$  is the union of Set8 and Waterloo140. (a) Denoising level curve of DRCAN with different noise-level estimators; (b) Denoising level curve of different denoisers with the ideal noise-level estimator.

For the first property, since Eq. 15 optimizes the parameters of  $P_\sigma$  for the noise level  $\sigma$ ,  $P_{\hat{\sigma}}$  exhibits a degraded performance compared with  $P_\sigma$  if dealing with  $x_o + \sigma \epsilon$  and  $\sigma \neq \hat{\sigma}$ , i.e.,

$$E_\epsilon (\|P_\sigma(x_o + \sigma \epsilon) - x_o\|^2) \leq E_\epsilon (\|P_{\hat{\sigma}}(x_o + \sigma \epsilon) - x_o\|^2), \quad (32)$$

Considering Eq. 28 and Eq. 32, we have  $\mu(P_\sigma, \sigma) \leq \mu(P_{\hat{\sigma}}, \sigma)$ . Further, considering Eq. 31, we have

$$\left(\frac{n}{m}\right)^k \left(\prod_{i=0}^{k-1} \mu(P_{\sigma^i}, \sigma^i)\right) \|\Phi^T y\|^2 \leq \left(\frac{n}{m}\right)^k \left(\prod_{i=0}^{k-1} \mu(P_{\hat{\sigma}^i}, \sigma^i)\right) \|\Phi^T y\|^2, \quad (33)$$

which means that an accurate noise-level estimator can lead to a smaller upper bound of reconstructed MSE. For the second property, let  $P_{\hat{\sigma}}^1$  be a denoiser which is better than  $P_{\hat{\sigma}}^2$  in the following sense:

$$E_\epsilon (\|P_{\hat{\sigma}}^1(x_o + \sigma \epsilon) - x_o\|^2) \leq E_\epsilon (\|P_{\hat{\sigma}}^2(x_o + \sigma \epsilon) - x_o\|^2), \quad (34)$$

from which we have  $\mu(P_{\hat{\sigma}}^1, \sigma) \leq \mu(P_{\hat{\sigma}}^2, \sigma)$ . Similar to the first property, the second property also holds. Thus, one can improve CS reconstruction performance based on these two properties.

Our method achieves an accurate noise-level estimator and efficient image denoiser by exploiting the elaborately designed residual-regressive net and the dilated residual channel attention net. Fig. 11 (a) illustrates the denoising level curves of the proposed DRCAN with different noise-level estimators, and Fig. 11 (b) illustrates the denoising level curves of different denoisers with the ideal noise-level estimator (i.e. the ground-truth noise-level value). PRN involves using a plain regressive net to estimate noise level and has a structure similar to RRN but without the residual operation, which revises Eq. 11 into  $\hat{\sigma}^k = R_1(R_2((x^k - \alpha v^k)))$ . Fig. 11 shows that the designed RRN and DRCAN can achieve a promising denoising level curve. The denoising level curve when using RRN is almost coincident with that when using the ideal noise level, while the curve when using PRN is only coincident with the ideal curve at high noise levels and is above the ideal curve at low noise levels, meaning that the plain structure has a relatively poor ability to estimate faint noise levels.

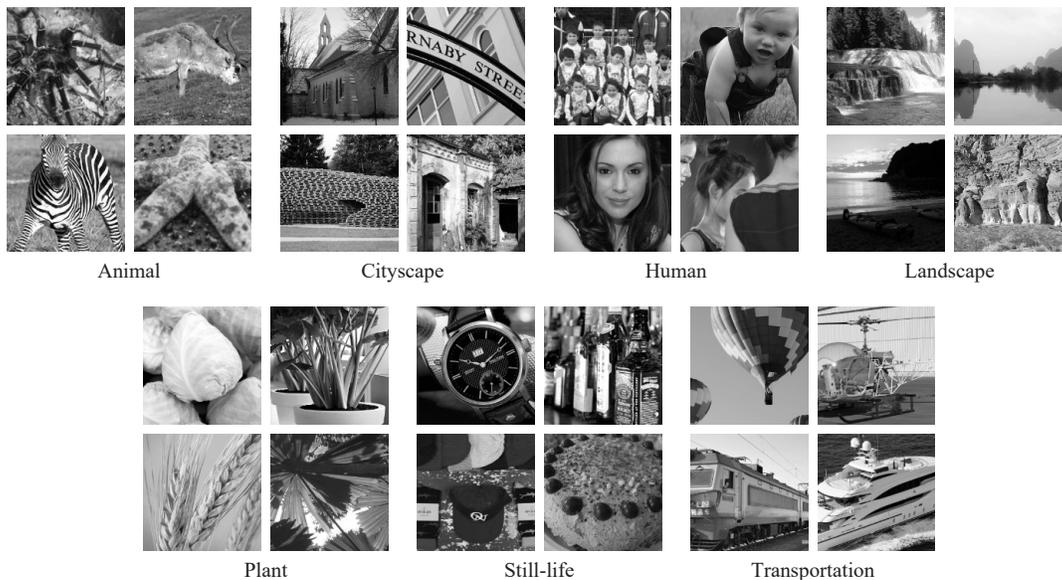


Fig. 12. Samples of test images from the Waterloo140 dataset [40].



Fig. 13. Eight standard test images (Set8).

TABLE II  
COMPARISON OF AVERAGE PSNR RESULTS WITH HAND-CRAFTED  
METHODS ON SET8 AND WATERLOO140 DATASETS.

Datasets	Methods	Sampling Ratio					
		0.05	0.1	0.15	0.2	0.25	0.3
Set8	BCS-SPL	23.01	25.07	26.51	27.67	28.57	29.46
	TV-CS	24.04	27.69	29.60	31.01	32.23	33.51
	NLR-CS	26.86	31.04	33.46	35.69	37.30	38.06
	BM3D-CS	26.52	31.38	33.54	35.22	36.68	37.99
	DRCAN+RRN	29.56	33.03	35.20	36.74	37.90	39.21
	DRCAN+RRN+	<b>29.69</b>	<b>33.12</b>	<b>35.32</b>	<b>36.84</b>	<b>38.03</b>	<b>39.36</b>
Waterloo-140	BCS-SPL	22.11	23.74	24.82	25.70	26.49	27.25
	TV-CS	22.47	25.37	27.03	28.35	29.53	30.67
	NLR-CS	24.57	27.38	29.41	31.19	32.68	33.52
	BM3D-CS	24.22	27.72	30.00	31.78	33.32	34.72
	DRCAN+RRN	26.21	29.67	31.77	33.55	35.03	36.47
	DRCAN+RRN+	<b>26.45</b>	<b>29.76</b>	<b>31.84</b>	<b>33.82</b>	<b>35.11</b>	<b>36.71</b>

## V. EXPERIMENTAL RESULTS

In this section, we provide a performance evaluation of the proposed image CS reconstruction method. We first describe the datasets for experiments and the parameter settings for the network training. We then compare the proposed method with state-of-the-art CS reconstruction methods. After that, we analyze the contribution of each component in the proposed method through the use of ablation experiments. Next, we provide the computational time of the proposed method. Finally, we test our method on noisy data.

### A. Experimental Setting

1) *Dataset*: To train the proposed models, we collect a large dataset including 500 images from Berkeley’s BSD-

500 datasets [41], 900 images from the DIV2K dataset [42], and 20,000 randomly selected images from the ImageNet database [43]. For all models, we crop training patches as  $192 \times 192$ , adding image rotation and flipping operations. To evaluate all competing methods, we use 2 datasets. One test dataset contains 140 images of size  $256 \times 256$  from the Waterloo Exploration Database [40], as shown in Fig. 12, which are broadly grouped into 7 categories (20 images in each class): human, animal, plant, landscape, cityscape, still-life, and transportation. Another test dataset contains 8 widely used standard images of size  $256 \times 256$ , as shown in Fig. 13. Note that none of those test images are included in the training dataset, and all images are converted to grayscale.

2) *Training and Testing*: All the training and testing processes were carried out on a PC with an Intel i5 CPU and Nvidia RTX 2070 GPU. We use PyTorch to train the designed networks, including 60 epochs. The ADAM optimizer was used to train the network with settings  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$ . The learning rate was initialized as  $10^{-4}$  and halved every 10 epochs. It takes approximately one day to train the noise-level estimation net  $G(\cdot)$  and one learned proximal operator  $P_\sigma(\cdot)$  for a certain range of noise levels  $\sigma$ . For the testing process, we restrict the model to 20 iterations, take  $\alpha$  as 1, and set the number of image subspaces to 17.

### B. Comparison with State-of-the-Art Methods

In this subsection, we first compare the performance of the proposed method with some hand-crafted CS image re-

TABLE III  
COMPARISON OF PSNRs WITH DEEP-BASED METHODS ON SET8 DATASET.

Images	Methods	Sampling Ratio						
		0.01	0.05	0.10	0.20	0.30	0.40	0.50
Barbara	ISTA-Net	18.38	21.20	23.51	24.97	29.87	32.85	35.49
	ISTA-Net+	18.53	21.61	23.52	26.72	30.13	33.98	36.65
	CSNet+	21.77	23.76	24.41	26.69	31.22	34.72	38.26
	SCSNet	<b>21.80</b>	23.73	24.43	26.84	31.43	35.16	38.58
	OPINE-Net+	20.98	23.53	24.73	27.52	32.17	35.45	39.47
	AMP-Net	21.05	23.79	24.76	28.79	33.53	37.12	39.87
	LDAMP	19.72	22.89	25.99	32.79	36.03	37.97	39.15
	DRCAN+RRN	20.34	25.75	29.91	34.65	36.76	38.81	40.61
	DRCAN+RRN+	16.46	<b>26.05</b>	<b>30.10</b>	<b>34.92</b>	<b>36.86</b>	<b>38.95</b>	<b>40.91</b>
	Boat	ISTA-Net	18.47	23.32	27.27	29.09	34.60	37.05
ISTA-Net+		18.51	23.73	27.41	31.28	35.22	37.84	39.91
CSNet+		21.98	27.26	29.98	33.40	35.84	38.03	40.38
SCSNet		<b>22.02</b>	27.23	30.11	33.57	36.30	38.62	40.97
OPINE-Net+		21.14	27.24	30.94	33.90	36.82	39.23	42.23
AMP-Net		21.23	27.64	30.69	34.66	37.60	40.17	42.15
LDAMP		19.33	25.60	30.17	34.81	37.92	39.89	41.28
DRCAN+RRN		21.14	27.96	32.00	36.09	38.70	40.66	41.83
DRCAN+RRN+		18.15	<b>28.03</b>	<b>32.09</b>	<b>36.15</b>	<b>38.89</b>	<b>40.87</b>	<b>42.40</b>
Camera-man		ISTA-Net	17.26	20.52	23.46	25.15	30.04	32.09
	ISTA-Net+	17.32	20.99	23.76	27.25	30.35	32.36	34.32
	CSNet+	20.32	23.46	25.57	28.63	30.71	32.47	34.03
	SCSNet	20.38	23.62	25.71	28.53	30.65	32.38	34.34
	OPINE-Net+	19.99	23.96	26.88	29.56	31.58	33.60	36.18
	AMP-Net	19.99	24.15	26.84	29.45	32.47	34.40	36.32
	LDAMP	17.96	24.92	29.10	32.12	33.36	37.57	38.85
	DRCAN+RRN	<b>20.73</b>	<b>27.23</b>	<b>30.09</b>	<b>34.03</b>	<b>36.72</b>	<b>39.61</b>	<b>41.66</b>
	DRCAN+RRN+	18.81	<b>27.47</b>	<b>30.37</b>	<b>34.12</b>	<b>37.02</b>	<b>39.64</b>	<b>41.80</b>
	Foreman	ISTA-Net	20.21	26.52	32.78	32.98	39.17	41.38
ISTA-Net+		20.34	27.46	33.49	36.95	40.22	42.30	44.18
CSNet+		26.77	32.37	35.04	38.66	40.59	42.18	43.74
SCSNet		26.74	32.31	35.17	38.46	40.82	42.66	44.33
OPINE-Net+		23.89	31.19	36.64	38.92	41.11	43.14	45.98
AMP-Net		23.72	32.53	35.32	39.25	41.45	43.47	45.32
LDAMP		23.74	34.18	36.80	40.00	42.26	43.67	44.66
DRCAN+RRN		<b>28.46</b>	34.91	37.93	40.53	42.87	44.64	46.02
DRCAN+RRN+		12.34	<b>35.05</b>	<b>37.93</b>	<b>40.67</b>	<b>42.98</b>	<b>44.85</b>	<b>46.20</b>
House		ISTA-Net	19.80	24.90	30.13	31.54	36.41	38.04
	ISTA-Net+	20.00	25.84	30.49	34.27	37.07	38.64	40.33
	CSNet+	24.14	30.07	32.60	35.39	37.90	39.54	41.89
	SCSNet	24.18	30.15	32.69	35.55	37.92	39.81	42.20
	OPINE-Net+	22.10	29.90	34.03	36.35	38.20	39.70	42.07
	AMP-Net	23.07	30.51	34.09	37.13	38.77	40.41	43.03
	LDAMP	22.37	32.71	34.73	37.23	39.19	40.51	42.04
	DRCAN+RRN	<b>26.65</b>	<b>33.79</b>	<b>36.00</b>	<b>38.12</b>	<b>40.11</b>	<b>41.90</b>	<b>43.47</b>
	DRCAN+RRN+	14.51	<b>27.04</b>	<b>33.89</b>	<b>36.03</b>	<b>38.16</b>	<b>40.16</b>	<b>43.66</b>
	Lena	ISTA-Net	18.29	23.28	27.44	28.72	33.19	35.61
ISTA-Net+		18.54	23.99	27.50	30.58	33.74	36.13	38.46
CSNet+		<b>22.43</b>	26.87	29.19	32.26	34.86	37.17	39.19
SCSNet		22.41	26.86	29.29	32.36	35.22	37.66	39.90
OPINE-Net+		21.34	26.95	30.09	33.06	35.81	38.05	40.82
AMP-Net		21.36	27.18	29.86	33.38	36.35	38.77	41.02
LDAMP		19.60	27.61	31.44	36.39	39.18	41.10	42.27
DRCAN+RRN		22.25	28.67	32.87	36.95	39.72	41.95	43.84
DRCAN+RRN+		21.14	<b>28.89</b>	<b>32.90</b>	<b>37.02</b>	<b>39.86</b>	<b>42.10</b>	<b>43.94</b>
Monarch		ISTA-Net	14.99	20.26	25.58	27.17	34.04	36.96
	ISTA-Net+	15.01	20.52	25.72	30.29	34.80	37.69	40.22
	CSNet+	18.07	25.55	28.58	32.76	35.06	37.06	38.84
	SCSNet	18.05	25.52	28.88	32.86	35.58	37.91	40.01
	OPINE-Net+	17.63	25.43	29.94	33.25	36.23	38.46	41.62
	AMP-Net	17.62	25.94	29.71	34.08	37.10	39.53	41.74
	LDAMP	15.50	24.67	29.88	34.75	38.38	40.66	42.25
	DRCAN+RRN	<b>18.14</b>	26.95	31.04	35.86	38.74	41.06	42.65
	DRCAN+RRN+	16.75	<b>27.04</b>	<b>31.05</b>	<b>35.77</b>	<b>38.89</b>	<b>41.31</b>	<b>43.17</b>
	Parrot	ISTA-Net	17.90	22.25	26.21	28.12	32.60	35.02
ISTA-Net+		18.06	22.97	26.37	30.09	32.91	35.31	37.26
CSNet+		22.23	25.61	28.11	31.34	33.84	36.09	38.20
SCSNet		22.30	25.46	28.10	31.29	34.13	36.41	38.26
OPINE-Net+		21.02	25.92	29.34	32.50	35.21	37.45	39.83
AMP-Net		21.23	22.87	29.20	30.06	35.85	38.22	40.30
LDAMP		20.40	30.29	33.27	37.23	39.62	41.39	42.73
DRCAN+RRN		<b>23.02</b>	30.98	34.40	37.73	40.10	41.52	43.21
DRCAN+RRN+		21.48	<b>31.12</b>	<b>34.46</b>	<b>37.88</b>	<b>40.19</b>	<b>42.05</b>	<b>43.65</b>
Average		ISTA-Net	18.16	22.78	27.05	28.47	33.74	36.12
	ISTA-Net+	18.29	23.39	27.28	30.93	34.31	36.78	38.92
	CSNet+	22.21	26.87	29.19	32.39	35.00	37.16	39.32
	SCSNet	22.23	26.86	29.30	32.43	35.26	37.58	39.82
	OPINE-Net+	21.01	26.76	30.32	33.13	35.89	38.13	41.02
	AMP-Net	21.16	26.83	30.06	33.35	36.64	39.01	41.22
	LDAMP	19.83	27.86	31.42	35.67	38.24	40.35	41.65
	DRCAN+RRN	<b>22.58</b>	29.56	33.03	36.74	39.21	41.19	42.91
	DRCAN+RRN+	17.46	<b>29.69</b>	<b>33.12</b>	<b>36.84</b>	<b>39.36</b>	<b>41.47</b>	<b>43.22</b>

TABLE IV  
COMPARISON OF AVERAGE PSNRs WITH DEEP-BASED METHODS ON WATERLOO140 DATASET.

Images	Methods	Sampling Ratio						
		0.01	0.05	0.10	0.20	0.30	0.40	0.50
Animal	ISTA-Net	18.56	23.09	26.41	28.08	32.19	34.44	36.59
	ISTA-Net+	18.66	23.50	26.51	29.70	32.62	34.84	37.04
	CSNet+	22.65	26.55	28.61	31.85	34.23	36.38	38.30
	SCSNet	<b>22.68</b>	26.55	28.90	31.97	34.34	36.49	38.55
	OPINE-Net+	21.64	26.46	29.57	32.31	34.85	36.97	39.76
	AMP-Net	21.75	25.90	29.49	31.27	35.48	37.80	39.99
	LDAMP	20.81	25.82	28.63	32.86	35.65	38.19	40.20
	DRCAN+RRN	21.95	26.75	29.85	33.57	36.60	38.83	40.91
	DRCAN+RRN+	20.39	<b>26.87</b>	<b>29.92</b>	<b>33.80</b>	<b>36.72</b>	<b>39.24</b>	<b>41.61</b>
	City-scape	ISTA-Net	16.78	20.43	23.69	25.22	30.15	32.69
ISTA-Net+		16.91	20.82	23.89	27.25	30.78	33.16	35.59
CSNet+		20.18	23.62	25.74	28.66	31.00	33.04	34.97
SCSNet		<b>20.19</b>	23.68	25.94	28.82	31.19	33.34	35.45
OPINE-Net+		19.42	23.74	26.92	29.50	32.14	34.55	37.72
AMP-Net		19.69	23.80	26.76	29.25	33.02	35.43	37.63
LDAMP		17.75	23.44	27.05	31.64	34.62	37.03	39.03
DRCAN+RRN		19.87	25.52	29.16	32.94	36.09	38.45	40.63
DRCAN+RRN+		18.94	<b>25.76</b>	<b>29.25</b>	<b>33.18</b>	<b>36.27</b>	<b>39.02</b>	<b>41.47</b>
Human		ISTA-Net	17.14	22.06	26.05	27.80	33.38	36.09
	ISTA-Net+	17.20	22.42	26.21	30.16	34.11	36.72	39.18
	CSNet+	20.96	25.35	27.69	31.35	33.95	36.32	38.20
	SCSNet	<b>20.99</b>	25.39	28.12	31.66	34.33	36.59	38.62
	OPINE-Net+	20.53	25.86	29.69	32.92	35.86	38.16	41.30
	AMP-Net	20.68	25.28	29.38	31.54	36.43	38.81	40.89
	LDAMP	18.41	25.30	29.69	34.19	37.39	39.91	41.54
	DRCAN+RRN	20.02	26.49	31.03	35.09	38.51	40.91	42.98
	DRCAN+RRN+	18.18	<b>27.05</b>	<b>31.14</b>	<b>35.71</b>	<b>38.74</b>	<b>41.42</b>	<b>43.83</b>
	Land-scape	ISTA-Net	19.48	22.95	25.63	26.74	30.04	31.91
ISTA-Net+		19.81	23.50	25.75	28.21	30.42	32.16	34.01
CSNet+		23.51	26.35	27.96	30.20	31.98	33.60	35.21
SCSNet		<b>23.51</b>	26.42	28.05	30.22	32.04	33.71	35.36
OPINE-Net+		22.20	25.94	28.31	30.33	32.23	33.94	36.18
AMP-Net		22.23	26.18	28.37	30.13	32.63	34.39	36.13
LDAMP		21.64	25.77	27.85	30.75	32.70	34.50	36.20
DRCAN+RRN		23.09	26.46	28.62	31.46	33.71	35.68	37.58
DRCAN+RRN+		22.81	<b>26.53</b>	<b>28.66</b>	<b>31.55</b>	<b>33.91</b>	<b>36.08</b>	<b>38.30</b>
Plant		ISTA-Net	17.91	21.87	25.38	26.87	31.43	33.80
	ISTA-Net+	18.02	22.33	25.51	28.76	31.98	34.34	36.64
	CSNet+	21.61	25.38	27.46	30.79	33.11	35.26	37.02
	SCSNet	<b>21.62</b>	25.38	27.70	30.93	33.36	35.46	37.29
	OPINE-Net+	20.64						

TABLE V

COMPARISON OF AVERAGE PSNR RESULTS ON THE TEST DATASET USING RANDOMLY PERMUTED CODED DIFFRACTION MEASUREMENTS.

Images	Methods	Sampling Ratio		
		0.05	0.10	0.20
Se8	BM3D-CS [27]	26.48	32.82	37.28
	LDAMP [18]	29.31	33.84	38.79
	ADMM-Net [12]	28.27	32.64	37.85
	DRCAN+RRN	<b>31.49</b>	<b>35.84</b>	<b>40.55</b>
Waterloo -140	BM3D-CS [27]	24.13	28.81	33.54
	LDAMP [18]	25.99	30.69	36.88
	ADMM-Net [12]	25.87	29.67	35.04
	DRCAN+RRN	<b>28.16</b>	<b>32.53</b>	<b>37.93</b>

Table II, one can see that our method significantly outperforms the compared hand-crafted methods with all sampling ratios. Specifically, our method outperforms NLR-CS and BM3D-CS by 2.29 dB and 1.95 dB on the Waterloo140 dataset at a sampling ratio of 0.10.

We further compare our method with some advanced deep-based CS image reconstruction methods (including ISTA-Net+ [11], CSNet+ [46], SCSNet [15], OPINE-Net+ [47], AMP-Net [13], and LDAMP [18]) on the Set8 and Waterloo140 datasets. Table III provides the PSNR values of the competing methods for every image in Set8, and Table IV lists the average PSNR results of seven classes of Waterloo140. From Table III and Table IV, one can see that the proposed method achieves the best performance in most sampling ratio cases, even without self-ensembles. For instance, the proposed method outperforms LDAMP by 1.61 dB and 1.45 dB at a sampling ratio of 0.10 for the Set8 and Waterloo140 datasets, respectively.

In addition to randomly scrambled block Bernoulli measurements, we also evaluated our method on the testing dataset with 0.05, 0.10, and 0.20 sampling ratios using randomly permuted coded diffraction measurements. Table V provides the average PSNR results for testing images, which indicate that our method outperforms the compared methods. Specifically, our method outperforms LDAMP and ADMM-Net by 1.76 dB and 2.70 dB on the Set8 dataset at a sampling ratio of 0.20.

To facilitate the evaluation of subjective qualities, Fig. 14 presents the parts of reconstructed images. The zoomed portions show that the reconstruction quality of the proposed method can restore sharper details with fewer artifacts. In particular, using the self-ensemble strategy can further enhance some edge details that are hard to restore. All of these testing results indicate that our method surpasses the existing state-of-the-art image CS reconstruction methods both in quantitative results and perceptual quality.

### C. Ablation Studies

#### 1) Ablation Studies with Different Convolution Kernels:

To verify the effects of the designed DRCAN, we modify the baseline designed network and compare their reconstruction performance. We mainly test the effects of dilatation convolution, convolution kernel size, and the number of subspaces on the reconstructed performance. Comparing the 4th column and the 7th column of Table VI, one can see that using dilatation convolution can achieve relatively high PSNR improvement at

TABLE VI

COMPARISON OF AVERAGE PSNR RESULTS ON THE TEST DATASET USING DIFFERENT CONVOLUTION KERNELS AND SUBSPACE NUMBERS.

Images	w/o Dilatation	use	not use	use	use	use
	kernal size	3x3	5x5	5x5	5x5	5x5
	subspace no.	17	17	2	10	17
Set8	0.05	29.05	29.36	21.39	29.46	<b>29.56</b>
	0.10	32.36	32.85	22.22	32.96	<b>33.03</b>
	0.15	34.60	35.16	25.06	35.11	<b>35.20</b>
	0.20	36.33	36.71	27.54	36.66	<b>36.74</b>
	0.25	37.70	37.89	28.83	37.89	<b>37.90</b>
	0.30	38.87	39.16	29.59	39.15	<b>39.21</b>
Waterloo -140	0.05	25.90	26.14	20.40	26.15	<b>26.21</b>
	0.10	29.15	29.50	21.49	29.46	<b>29.67</b>
	0.15	31.37	31.76	24.02	31.71	<b>31.77</b>
	0.20	33.13	33.54	26.26	33.50	<b>33.55</b>
	0.25	34.62	35.01	27.48	35.01	<b>35.03</b>
	0.30	35.94	36.35	28.29	36.36	<b>36.47</b>

a low sampling ratio and insignificant improvement at a high sampling ratio, with an approximately 0.20 dB gain at a 0.05 sampling ratio and only a 0.05 dB gain at a 0.30 sampling ratio for Set8. Further comparing the 3rd column and the 7th column, one can see that the convolution kernel with  $5 \times 5$  size achieves a more stable improvement than the kernel with  $3 \times 3$  size both at low and high sampling ratios. Considering the trade-off between the computational burden and reconstruction performance, we choose the convolution kernel size as  $5 \times 5$ .

#### 2) Ablation Studies with different Subspace Numbers:

In our iterative CS reconstruction method, the intermediate corrupted images are attributed to some subspace based on the measured distortion distances. Considering that each subspace has its own individual-learned proximal operator mapping the noisy image to clean image sets, more subspaces can achieve better learned proximal operators and vice versa. To verify the effectiveness of the number of subspaces, the 5th, 6th, and 7th columns of Table VI provide the PSNR results of the proposed method with the same settings except for the subspace number. We can see that increasing the subspace number from 2 to 10 brings a notable gain of approximately 5.75 dB, while increasing the subspace number from 10 to 17 spaces brings a gain of approximately 0.06 dB at a 0.05 sampling ratio for Waterloo140. Considering that a subspace number larger than 17 would only bring a gain of less than 0.06 dB, we choose the number of subspaces as 17.

3) *Ablation Studies on Network Components:* We further remove some core components from the DRCAN and RRN baseline designs to verify their impact on CS reconstruction performance. The original DRCAN contains two residual channel attention blocks (RCABs), and each RCAB further contains 8 channel attention layers. For the ablation study of DRCAN, we can reduce the number of RCABs in DRCAN to one and obtain the DRCAN(1B) architecture. Additionally, we remove the channel attention operation in CAL and obtain the DRCAN with no channel attention, dubbed DRCAN(NC). For the ablation study of RRN, we construct a plain regressive net, which has the same structure as RRN except without residual operation. Table VII provides the average PSNRs of the different combinations of noise-level estimators and denoisers, from which we can see that the 2-RCAB structure using channel attention (i.e. DRCAN) achieves approximately

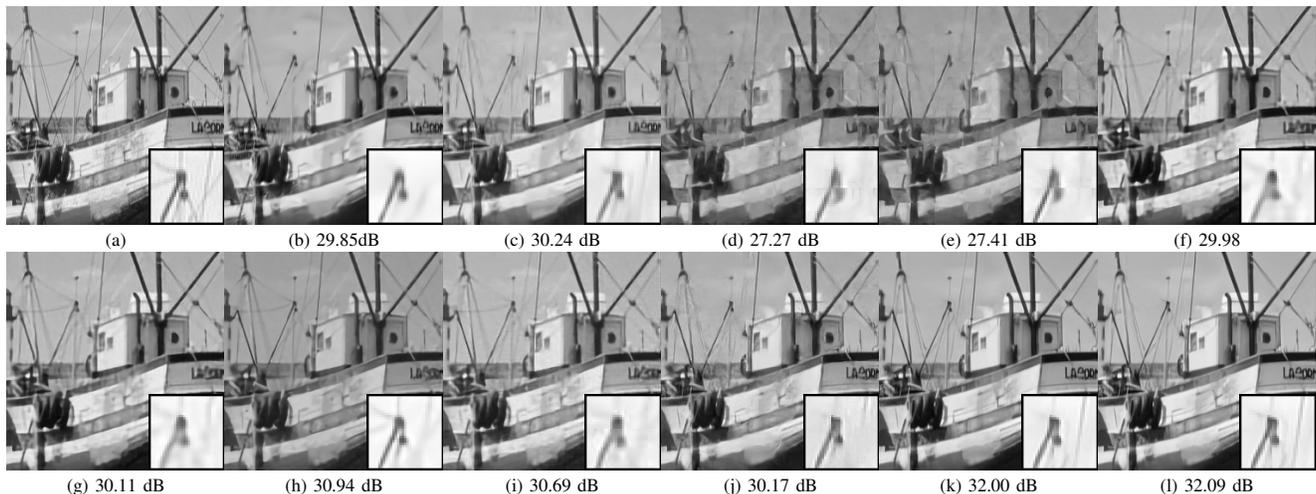


Fig. 14. Reconstructed *boat* at sampling ratio 0.1 using Bernoulli measurements. (a) Original image; (b) NLR-CS [28]; (c) BM3D-CS [27]; (d) ISTA-Net [11]; (e) ISTA-Net+ [11]; (f) CSNet+ [46]; (g) SCSNet [15]; (h) OPINE-Net+ [47]; (i) AMP-Net [13]; (j) LDAMP [18]; (k) DRCAN+RRN; (l) DRCAN+RRN+.

TABLE VII

AVERAGE PSNR COMPARISONS USING DIFFERENT COMBINATIONS OF NOISE-LEVEL ESTIMATORS AND DENOISERS ON SET8 AND WATERLOO140.

Images	Methods	Sampling Ratio						
		0.01	0.05	0.10	0.20	0.30	0.40	0.50
Set8	DRCAN(1B)+RRN	22.15	28.73	31.68	35.52	38.00	39.08	40.44
	DRCAN(NC)+RRN	22.54	29.41	32.22	36.25	38.39	38.93	39.52
	DRCAN+PRN	22.57	28.93	31.43	34.81	38.40	40.96	42.17
	DRCAN+RRN	<b>22.58</b>	<b>29.56</b>	<b>33.03</b>	<b>36.74</b>	<b>39.21</b>	<b>41.19</b>	<b>42.91</b>
Waterloo-140	DRCAN(1B)+RRN	20.78	25.83	28.74	32.45	34.92	36.75	38.31
	DRCAN(NC)+RRN	20.94	26.17	29.25	32.94	35.19	36.69	37.98
	DRCAN+PRN	20.92	26.14	28.73	32.25	35.43	37.82	38.90
	DRCAN+RRN	<b>20.95</b>	<b>26.21</b>	<b>29.67</b>	<b>33.55</b>	<b>36.47</b>	<b>38.74</b>	<b>40.80</b>

TABLE VIII

AVERAGE PSNR RESULTS ON THE TEST DATASETS USING DnCNN AND DRCAN COMBINED WITH DIFFERENT NOISE-LEVEL ESTIMATORS.

Datasets	Methods	Sampling Ratio					
		0.1		0.2		0.3	
		Bernoulli	Diffraction	Bernoulli	Diffraction	Bernoulli	Diffraction
Set8	DnCNN+ $\frac{\ v^k\ _2}{\sqrt{n}}$	31.42	33.84	35.67	38.79	38.24	42.36
	DnCNN+RRN	31.67	34.10	35.71	38.85	38.70	42.70
	DRCAN+ $\frac{\ v^k\ _2}{\sqrt{n}}$	32.76	35.62	36.57	40.15	39.03	43.33
	DRCAN+RRN	<b>33.03</b>	<b>35.84</b>	<b>36.74</b>	<b>40.55</b>	<b>39.21</b>	<b>43.59</b>
Waterloo-140	DnCNN+ $\frac{\ v^k\ _2}{\sqrt{n}}$	28.22	30.69	32.48	36.88	35.28	40.39
	DnCNN+RRN	28.63	31.10	32.73	36.49	35.44	40.56
	DRCAN+ $\frac{\ v^k\ _2}{\sqrt{n}}$	29.57	32.42	33.06	37.17	35.87	41.71
	DRCAN+RRN	<b>29.67</b>	<b>32.53</b>	<b>33.55</b>	<b>37.93</b>	<b>36.47</b>	<b>42.05</b>

TABLE IX

AVERAGE PSNR AND RUNNING TIME(SECONDS) COMPARISON OF DIFFERENT IMAGE SIZES.

Methods	PSNR(dB) / Times(s)				
	128x128	256x256	512x512	1024x1024	
Hand-crafted	BCS-SPL	20.94/3.403	21.87/6.101	22.65/26.613	23.46/86.311
	TV-CS	20.35/33.383	21.26/108.149	21.92/418.571	22.56/1644.189
	NLR-CS	20.00/37.701	20.66/145.909	21.15/597.557	21.64/2466.751
	BM3D-CS	26.78/2.184	30.38/8.895	32.31/37.754	34.63/154.466
Deep-based	ISTA-Net+	25.62/ <b>0.004</b>	28.01/ <b>0.005</b>	30.10/ <b>0.005</b>	32.34/ <b>0.006</b>
	CSNet+	27.59/0.021	29.29/0.29	31.08/0.073	33.36/0.235
	SCSNet	27.94/0.02	29.55/0.03	31.22/0.102	33.50/0.377
	OPINE-Net+	28.85/0.006	31.11/0.008	32.99/0.008	35.64/0.009
	AMP-Net	28.02/0.027	29.93/0.031	31.95/0.076	34.80/0.253
	LDAMP	26.81/0.388	30.94/0.512	32.70/1.035	35.12/3.352
	DRCAN+RRN	<b>28.95/0.652</b>	<b>31.70/0.921</b>	<b>33.44/2.232</b>	<b>35.67/6.185</b>

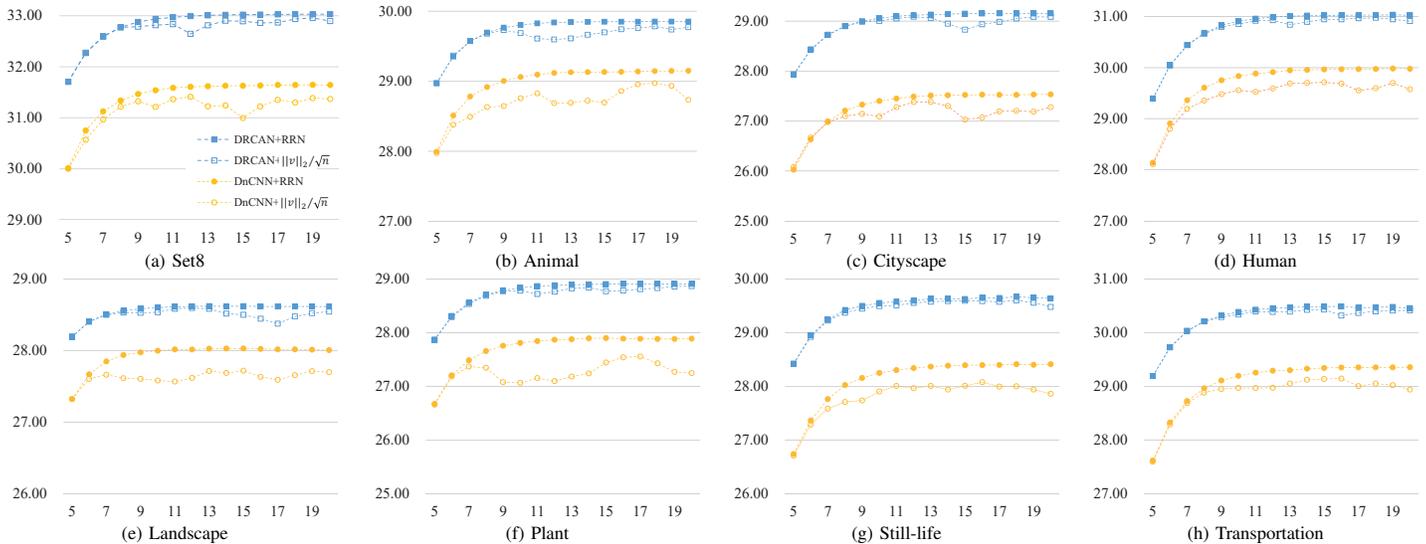


Fig. 15. Iteration PSNRs (dB) of the reconstructed images at a sampling ratio of 0.1. The Y-coordinate represents the average PSNRs of the reconstructed images, and the X-coordinate represents the iteration number.

0.93 dB and 0.42 dB boosts versus the 1-RCAB structure (i.e. DRCAN(1B)) and no-channel attention structure (i.e. DRCAN(NC)) at 0.10 sampling ratio. Replacing PRN with RRN achieves approximately 0.94 dB boosts at a 0.10 sampling ratio for the Waterloo140 dataset. All of these results show the effectiveness of the designed networks.

#### 4) Ablation Studies using different Noise-Level estimators:

In our iterative CS reconstruction method, noise level estimation is a crucial step that determines the parameters of the adopted denoiser. Under the assumption of  $x^k$  and  $v^k$  in Eq. 18 being the pure clean image and noise [18], [39], traditional methods utilize the intermediate feature  $\frac{\|v^k\|_2}{\sqrt{n}}$  to estimate the noise level, ignoring the noise component in the intermediate result  $x^k$ . Table VIII presents the average CS reconstructed PSNR of methods that combine DnCNN and DRCAN with different noise-level estimators. Note that  $\text{DnCNN} + \frac{\|v^k\|_2}{\sqrt{n}}$  is similar to the LDAMP method described in [18]. Comparing  $\text{DnCNN} + \text{RRN}$  and  $\text{DnCNN} + \frac{\|v^k\|_2}{\sqrt{n}}$ , one can see that simply replacing the noise estimator of LDAMP from  $\frac{\|v^k\|_2}{\sqrt{n}}$  with RRN brings about 0.25 dB and 0.41 dB improvement for the Set8 and Waterloo140 datasets at a 0.10 sampling ratio. Fig. 15 shows the iterative reconstructed PSNR results of different combinations, from which one can also see that DRCAN+RRN includes a better and more stable reconstructed iterative process.

#### D. Computational Time with Different Image Resolutions

Here, we compare the computational time of CS reconstruction algorithms using different image resolutions. We randomly selected five  $1024 \times 1024$ -sized images from the ImageNet dataset which are not contained in the training set. We then obtained their corresponding images of sizes  $512 \times 512$ ,  $256 \times 256$ , and  $128 \times 128$  by bicubic downsampling. Table IX presents the average computational time and reconstructed PSNR of test images. The proposed method can achieve the best PSNR results for all image resolutions. Additionally,



Fig. 16. Reconstructed *Lena* from noisy measurements at a 0.3 sampling ratio. (a) Original image; (b) Noisy image; (c) ISTA-Net+; (d) CSNet+; (e) SCSNet; (f) OPINE-Net+; (g) AMP-Net; (h) LDAMP; (i) DRCAN+RRN.

one can see that end-to-end deep-based CS reconstruction algorithms (e.g., ISTA-Net+, CSNet+, SCSNet, and OPINE-Net+) run faster than plug-and-play algorithms (e.g., LDAMP and Proposed), and the running time of all methods rises as the image size increases.

#### E. Performance on Noisy Data

Here, we conduct image CS reconstruction from noisy CS measurements to demonstrate the robustness of the proposed

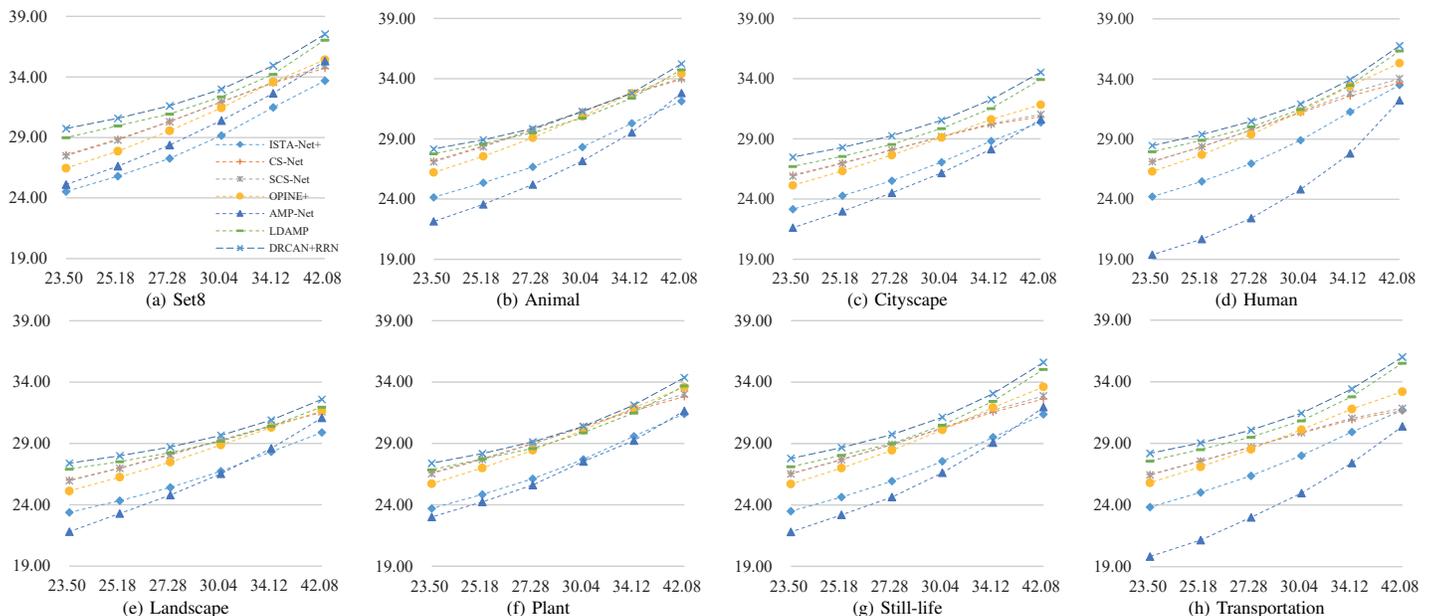


Fig. 17. PSNRs (dB) of the reconstructed images from noisy measurements at sampling ratio 0.3. The Y-coordinate represents the PSNRs of the reconstructed images, and the X-coordinate represents the PSNRs of the measured noisy images.

method with respect to noise. In this case, we distort the CS measurements by corrupting the original images with different levels of Gaussian noise. The measurement ratio is fixed at 0.30, and measured images with PSNRs between 23.50 dB and 42.08 dB are generated by varying the standard deviation of the Gaussian noise added to the original image. Fig. 17 presents the average PSNR results for various noise levels for different algorithms. This demonstrates that the reconstruction performance degrades for all competing methods as the noise level increases, while our method is less affected by noise. Fig. 16 shows the reconstructed *Lena* from noisy measurements. Both PSNR and subjective quality comparison results show the efficiency and robustness of the proposed method in the presence of noise.

## VI. CONCLUSION

In this paper, we have developed a novel framework for image CS reconstruction based on learned regularization and proximal operators. The proposed framework leverages the PMGD algorithm to solve the CS optimization problem and utilizes the elaborately designed residual-regressive net and the dilated residual channel attention net to simulate the regularization term and proximal operator, respectively. Meanwhile, we partition the image into multi-subspaces, where each subspace has its own proximal operator mapping images contained in the subspace into the clean subspace. Furthermore, we introduce the self-ensemble strategy to improve CS reconstruction performance. State evolution analysis indicates the effectiveness of the designed networks. Experimental results also demonstrated that the proposed method outperforms existing state-of-the-art NLR-CS and LDAMP algorithms in terms of PSNR and visual perception with both noiseless and noisy settings.

Regarding our future work, one interesting topic is to design a more powerful deep architecture by using U-net structure,

dense connection, self-attention strategy, and other deep network techniques. This will address the research gap pertaining to the performance of the developed strategy proportions with respect to the representation ability of the deep net for the proximal operator. Another promising direction is to apply the proposed framework in CS-based applications, such as image encryption, fast MRI, snapshot imaging, etc.

## REFERENCES

- [1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] O. V. Michailovich, Y. Rathi, and S. Dolui, "Spatially regularized compressed sensing for high angular resolution diffusion imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1100–1115, 2011.
- [4] Z. Chen, X. Hou, L. Shao, C. Gong, X. Qian, Y. Huang, and S. Wang, "Compressive sensing multi-layer residual coefficients for image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1109–1120, 2020.
- [5] M. F. Duarte, M. A. Davenport, D. Takhar, 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.
- [6] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 65–88, 2021.
- [7] Z. Chen, X. Hou, X. Qian, and C. Gong, "Efficient and robust image coding and transmission based on scrambled block compressive sensing," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1610–1621, 2018.
- [8] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "A hybrid structural sparsification error model for image restoration," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [9] Z. Zha, X. Yuan, B. Wen, J. Zhou, and C. Zhu, "Group sparsity residual constraint with non-local priors for image restoration," *IEEE Transactions on Image Processing*, vol. 29, pp. 8960–8975, 2020.
- [10] Z. Zha, X. Yuan, J. Zhou, C. Zhu, and B. Wen, "Image restoration via simultaneous nonlocal self-similarity priors," *IEEE Transactions on Image Processing*, vol. 29, pp. 8561–8576, 2020.

- [11] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1828–1837.
- [12] Y. Yang, J. Sun, H. Li, and Z. Xu, "Admm-csnet: A deep learning approach for image compressive sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 521–538, 2020.
- [13] 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, 2020.
- [14] X. Yuan, J. Yang, P. Llull, X. Liao, G. Sapiro, D. J. Brady, and L. Carin, "Adaptive temporal compressive sensing for video," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 14–18.
- [15] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Scalable convolutional neural network for image compressed sensing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 290–12 299.
- [16] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.
- [17] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1447–1457.
- [18] C. Metzler, A. Mousavi, and R. Baraniuk, "Learned d-amp: Principled neural network based compressive image recovery," in *Advances in Neural Information Processing Systems*, 2017, pp. 1772–1783.
- [19] K. Zhang, W. Zuo, Y. Chen, 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, 2016.
- [20] J. H. R. Chang, C. Li, B. Poczos, and B. V. K. V. Kumar, "One network to solve them all - solving linear inverse problems using deep projection models," in *International Conference on Computer Vision*, 2017, pp. 5889–5898.
- [21] Z. Zha, X. Yuan, J. T. Zhou, J. Zhou, B. Wen, and C. Zhu, "The power of triply complementary priors for image compressive sensing," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 983–987.
- [22] Z. Chen, X. Hou, C. Gong, and X. Qian, "Compressive sensing reconstruction for compressible signal based on projection replacement," *Multimedia Tools and Applications*, vol. 75, no. 5, pp. 2565–2578, 2016.
- [23] Z. Chen, X. Hou, L. Shao, and S. Wang, "Revising regularisation with linear approximation term for compressive sensing improvement," *Electronics Letters*, vol. 55, no. 7, pp. 384–386, 2019.
- [24] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *16th IEEE International Conference on Image Processing*, 2009, pp. 3021–3024.
- [25] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45, 2009, pp. 18 914–18 919.
- [26] S. Li and H. Qi, "A douglas–rachford splitting approach to compressed sensing image recovery using low-rank regularization," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4240–4249, 2015.
- [27] 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.
- [28] 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.
- [29] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3929–3938.
- [30] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.
- [31] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 1799–1808.
- [32] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE information theory workshop on information theory (ITW 2010, Cairo)*. IEEE, 2010, pp. 1–5.
- [33] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv:1609.04747 [cs.LG]*, 2016.
- [34] R. Sathish, B. Thierry, and U. Michael, "Monte-carlo sure: a black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, 2008.
- [35] M. Zhussip, S. Soltanayev, and S. Y. Chun, "Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 247–10 256.
- [36] M. Mardani, Q. Sun, D. L. Donoho, V. Pappas, H. Monajemi, S. S. Vasanawala, and J. M. Pauly, "Neural proximal gradient descent for compressive imaging," in *Advances in Neural Information Processing Systems*, 2018, pp. 9573–9583.
- [37] R. Timofte, R. Rothe, and L. V. Gool, "Seven ways to improve example-based single image super resolution," in *Computer Vision and Pattern Recognition*, 2016, pp. 1865–1873.
- [38] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [39] 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.
- [40] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017.
- [41] 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 *IEEE International Conference on Computer Vision*, 2002, pp. 416–423.
- [42] R. Timofte, E. Agustsson, L. V. Gool, M. H. Yang, and G. Qi, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1110–1121.
- [43] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "Imagenet: a large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [44] J. Zhang, S. Liu, R. Xiong, S. Ma, and D. Zhao, "Improved total variation based image compressive sensing recovery by nonlocal regularization," in *2013 IEEE International Symposium on Circuits and Systems*, 2013, pp. 2836–2839.
- [45] T. T. Do, G. Lu, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, 2011.
- [46] 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.
- [47] 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.



**Zan Chen** received the B.S. degree and Ph.D from Xi'an Jiaotong University in 2012 and 2019, respectively. He was a visiting scholar at the University of East Anglia (UEA), Norwich, UK, in 2018. Now, he is an associate professor with the College of Information Engineering, Zhejiang University of Technology. His research interests include compressive sensing, computer vision, and medical image processing.



**Wenlong Guo** received his B.S. degree in the College of Information Engineering from Zhejiang University of Technology, Hangzhou, China, in 2020. Currently, he is a postgraduate focused on control science and engineering at Zhejiang University of Technology. His research interests include compressive sensing and medical image processing.



**Ling Shao** is the CEO and Chief Scientist of the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. His research interests include computer vision, machine learning and medical imaging. He is an associate editor of IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems, and several other journals. He is a fellow of the IAPR, the IET and the BCS.



**Yuanjing Feng** holds a Ph.D. in control science and engineering from Xi'an Jiaotong University and an M.S. in mechanical design and theory from Northwest A&F University. Currently, he is the Director of the Institute of Information Processing and Automation and is working as a professor at Zhejiang University of Technology. His interests include data-driven modeling and optimization in the fields of intelligent transportation systems and medical image analysis.

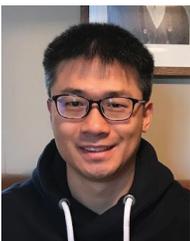


**Yongqiang Li** received his B.S. degree and Ph.D. degree in control theory and control engineering from Beijing Jiaotong University, Beijing, China, in 2007 and 2014, respectively. Currently, he is working as an associate professor in the School of Information Engineering at Zhejiang University of Technology, Hangzhou, China. His research interests are nonlinear control and machine learning.



**Changchen Zhao** received his Ph.D. degrees from Beihang University, Beijing, China, in 2018 and from National Chung Hsing University, Taichung, Taiwan, in 2018.

He is currently a lecturer in the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His research interests include computer vision, image processing, and machine learning.



**Yi Ren** received his Ph.D. degree in information communication and technology from the University of Agder, Norway, in 2012. He was with the Department of Computer Science, National Chiao Tung University (NCTU), Hsinchu, Taiwan, as a Postdoctoral Fellow and an Assistant Research Fellow from 2012 to 2017. He is currently a Lecturer (Assistant Professor) in the School of Computing Science at University of East Anglia (UEA), Norwich, U.K. He received the Best Paper Award at IEEE MDM 2012.