Scalable Convolutional Neural Network for Image Compressed Sensing

Wuzhen Shi[1], Feng Jiang[1,2], Shaohui Liu[1,2], and Debin Zhao[1,2]
[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[2]Peng Cheng Laboratory, Shenzhen, China
{wzhshi, fjiang, shliu, dbzhao}@hit.edu.cn

Abstract

Recently, deep learning based image Compressed Sensing (CS) methods have been proposed and demonstrated superior reconstruction quality with low computational complexity. However, the existing deep learning based image CS methods need to train different models for different sampling ratios, which increases the complexity of the encoder and decoder. In this paper, we propose a scalable convolutional neural network (dubbed SCSNet) to achieve scalable sampling and scalable reconstruction with only one model. Specifically, SCSNet provides both coarse and fine granular scalability. For coarse granular scalability, SCSNet is designed as a single sampling matrix plus a hierarchical reconstruction network that contains a base layer plus multiple enhancement layers. The base layer provides the basic reconstruction quality, while the enhancement layers reference the lower reconstruction layers and gradually improve the reconstruction quality. For fine granular scalability, SCSNet achieves sampling and reconstruction at any sampling ratio by using a greedy method to select the measurement bases. Compared with the existing deep learning based image CS methods, SCSNet achieves scalable sampling and quality scalable reconstruction at any sampling ratio with only one model. Experimental results demonstrate that SCSNet has the state-of-the-art performance while maintaining a comparable running speed with the existing deep learning based image CS methods.[1]

1. Introduction

Compressed Sensing (CS) [11] depicts a new paradigm for signal acquisition and reconstruction, which implements sampling and compression jointly. Given a sampling matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m << n$, CS states that a signal $x \in {}^{n \times 1}$, which can be represented sparsely in a transform domain, can be well reconstructed from its linear measurements $y = \Phi x$. Since the CS theory guarantees that a signal can be reconstructed with high quality at low sampling ratio when the signal is sparse in some domain, there has been significant interest in CS. Some works have been proposed to apply CS to image acquisition [12, 17], source coding [28, 15], wireless broadcast [37, 21], and so on.

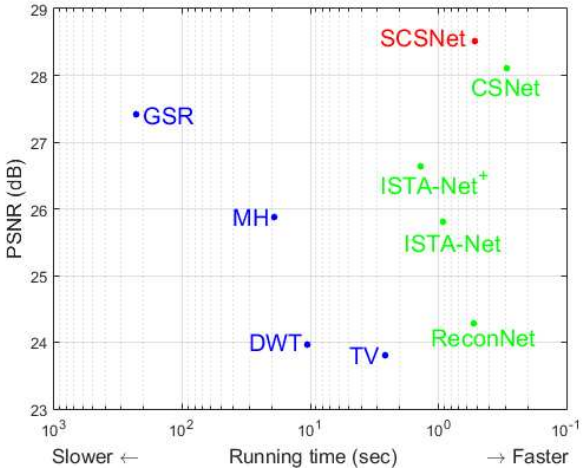

Figure 1. The reconstruction quality and running speed comparison on CPU. The compared traditional CS methods are marked with blue font, and the compared deep learning based CS methods are marked with green font. The chart is based on Set11 [20] results of sampling ratio of 0.1.

In the study of CS, the two main challengges are (1) the design of sampling matrix and (2) reconstructing the original signal from its linear measurements [14, 1]. To the first challenge, the representative sampling matices include: the random matrix [14], the binary matrix [1, 23], and the structural matrix [9, 15]. To the second challenge, the representative methods include: convex-optimization algorithms (e.g. [8, 35, 13]), the greedy algorithms (e.g. [24, 33, 30]), and the iterative thresholding algorithms (e.g. [16]). The iterative nature of these traditional methods lead to high computational complexity, which hampers their practical applications. Recently, a few deep learning based image CS reconstruction methods [26, 20, 4, 32, 39, 36] have been proposed. As shown in Figure 1, the deep learning based im-

[1]Test code is available at: https://github.com/wzhshi/SCSNet.

age CS methods can achieve better performance with lower computational complexity than the traditional methods.

The common problem of the existing deep learning based image CS methods is that they train different models for different sampling ratios, which increases the complexity of the encoder and decoder. Too many models consume considerable storage, memory bandwidth, and computational resources. Especially, these resource demands become prohibitive for embedded mobile applications. In addition, if the reconstruction quality at a given sampling ratio is not satisfied, the existing deep learning based image CS methods have to resample all measurements. This will lead to oversampling and harm the object being captured (e.g. medical imaging). Furthermore, some works [37, 21] investigate the image CS for wireless broadcast, in which different users will decode different quality images from different amount of measurements based on their channel conditions. Thus, scalable reconstruction is preferred. Both these two cases (medical imaging and wireless broadcast) expect scalable sampling and scalable reconstruction, which are not considered by the existing deep learning based image CS methods.

In this paper, we propose a scalable convolutional neural network (dubbed SCSNet) to achieve scalable sampling and scalable reconstruction that provides both coarse and fine granular scalability with only one model. For coarse granular scalability, SCSNet is designed as a single sampling matrix plus a hierarchical reconstruction network that contains a base layer (BL) and multiple enhancement layers (EL). The same with the coarse granular scalability of H.264 and H.265 [5], the BL of SCSNet provides the basic reconstruction quality. The ELs reference the lower layers and gradually improve the reconstruction quality. For fine granular scalability, SCSNet achieves sampling and reconstruction at any sampling ratio by using a greedy algorithm to select the measurement bases[2]. Compared with the existing deep learning based methods, SCSNet implements scalable sampling and scalable reconstruction at any sampling ratio with only one model. Experimental results show that SCSNet has the state-of-the-art reconstruction quality while maintaining a comparable running speed with the existing deep learning based image CS methods.

The main contributions of this paper are as follows:

- A scalable convolutional neural network (dubbed SCSNet) is proposed to achieve scalable sampling and scalable reconstruction with only one model.
- Coarse granular scalable sampling and scalable reconstruction using CNN is presented, in which the BL provides the basic reconstruction quality and the ELs gradually improve the reconstruction quality.
- Fine granular scalable sampling and scalable reconstruction is introduced, which employs a greedy method to select the measurement bases. The fine granular scalability can sample and recover the image at any sampling ratio.

[2]Each row of the sampling matrix is called as a measurement base in this paper.

2. Related work and motivation

We review the related work by grouping the existing methods into traditional CS methods and deep learning based CS methods. Generally, the traditional CS methods recover a signal from the CS measurements by solving a sparsity-regularized optimization problem. The well-known methods include: the convex optimization methods [7], the greedy algorithms [24, 33], and the gradient-descent methods [8, 35, 13]. For image CS, some methods introduce image prior as a regularization item. For example, Li et al. [22] used the total variation (TV) regularized constraint to replace the sparsity-based one for enhancing the local smoothness. In [40], Zhang et al. proposed group sparse representation (GSR) for image CS recovery by enforcing image sparsity and non-local self-similarity simultaneously. In addition, the block based compressed sampling (BCS) and projected Landweber based CS reconstruction methods [14, 27, 6] have also been proposed, in which additional optimization criteria can be easily incorporated. In [27], discrete wavelet transform (DWT) is used to encourage image sparsity. In [6], multi-hypothesis (MH) predictions is considered for CS reconstruction of both still images and video sequences.

Recently, some deep learning based image CS methods have been explored. These methods can be roughly divided into block-by-block reconstruction methods [26, 20, 39, 36] and end-to-end reconstruction methods [32]. In [26], Mousavi et al. proposed a stacked denoising autoencoder (SDA) to capture statistical dependencies between the different elements of certain signals and improve signal reconstruction performance. In [20], Kulkarni et al. used a CNN (ReconNet) for image block reconstruction and an off-the-shelf denoiser for deblocking. In [39], Zhang et al. cast the iterative shrinkage-thresholding algorithm as CNN (ISTA-Net). In [36], Xu et al. proposed a Laplacian pyramid reconstructive adversarial network (LAPRAN) that generates multiple outputs with different resolution simultaneously. These block-by-block reconstruction methods [26, 20, 39, 36] will cause blocking artifact. Compared with these methods, CSNet [32] can avoid blocking artifact by learning an end-to-end mapping between measurements and the whole reconstructed images. However, the existing deep learning based CS methods need to train different models for different sampling ratios, which increases the complexity of the encoder and decoder.

Image CS has been explored for many kinds of applications such as image acquisition [12, 17], image/video source

coding [28, 15], medical imaging [31], and wireless broadcast [37, 21]. The existing deep learning based CS methods use different models for different sampling ratios. This will cause difficulty for storage and hardware implementation. In addition, some applications need scalable sampling and scalable reconstruction. In medical imaging, oversampling may harm the object being captured. Scalable reconstruction is preferred in wireless broadcast. However, scalable sampling and scalable reconstruction are not considered by the existing deep learning based image CS methods.

3. Proposed method

3.1. Overview of SCSNet

Figure 2 shows the network structure schematic of SCSNet with two ELs. SCSNet uses a convolution layer with specific filter size and stride to implement BCS. The reconstruction network of SCSNet has a BL and multiple ELs. Both BL and EL have an initial reconstruction network and a deep reconstruction network. The initial reconstruction network of BL directly generates the initial reconstructed image from the measurements. The initial reconstruction networks of ELs first use the measurements to obtain the supplementary information (i.e. residual), and then add the initial reconstruction of the lower reconstruction layers to generate the initial reconstruction of ELs. A deep reconstruction network is used to refine the initial reconstruction in each reconstruction layer. This hierarchical reconstruction network structure is similar to the decoder architecture of scalable video coding [5], and provides coarse granular scalability.

To implement fine granular scalable sampling and scalable reconstruction, SCSNet first recognizes the importance of each measurement base offline. The higher reconstruction layers reference the initial reconstruction of the lower reconstruction layers, so the measurement bases in the lower reconstruction layers are more important than those in the higher reconstruction layers. The importance of the measurement bases in the same reconstruction layer are determined by a greedy method. SCSNet achieves sampling and reconstruction at any sampling ratio by removing some unimportant measurement bases and the corresponding connection to the reconstruction network.

3.2. Coarse granular scalability

3.2.1 BCS with a convolution layer

BCS divides the images into non-overlapping blocks of size $B \times B \times l$, where B and l are the spacial size and the amount of channel, respectively. Note that all our experiments are conducted on grayscale images, so $l = 1$ in this paper. To the j^{th} block x_j, BCS is represented as $y_j = \Phi_B x_j$, where Φ_B is the sampling matrix of size $n_B \times lB^2$ (for sampling ratio α, $n_B = \lfloor \alpha l B^2 \rfloor$). This process can be converted to a convolution layer with specific filer size and stride as

$$y = \mathrm{S}(x) = W_s * x \tag{1}$$

where W_s corresponds to n_B filters of support $B \times B \times l$. This convolution layer is represented as $Conv(B, l, n_B)$ in Figure 2, . There is no bias in the layer, and no activation function after this layer. To ensure the fixed sampling ratio, the stride of this convolution layer is $B \times B$ to implement non-overlapping sampling. With this specific convolution layer, the sampling matrix can be learned by jointly optimizing this convolution layer and the reconstruction network.

3.2.2 Hierarchical initial reconstruction network

The measurements obtained by the sampling layer can be treated as n_B feature maps that are divided into multiple groups as marked with different colors in Figure 2. BL uses only one group of measurements to get the initial reconstruction. Each EL uses one group of measurements to generate a reconstruction residual, and reference the lower layers to improve the initial reconstruction quality.

Given the measurements y, CSNet [32] obtains the initial reconstruction by using a convolution layer and a combination layer that is expressed as

$$\tilde{\mathrm{I}}(y) = W_{\mathrm{int}} * y \tag{2}$$

$$\tilde{x} = \mathrm{I}(y) = \kappa \begin{pmatrix} \gamma\left(\tilde{\mathrm{I}}_{11}(y)\right) & \cdots & \gamma\left(\tilde{\mathrm{I}}_{1w}(y)\right) \\ \vdots & \ddots & \vdots \\ \gamma\left(\tilde{\mathrm{I}}_{h1}(y)\right) & \cdots & \gamma\left(\tilde{\mathrm{I}}_{hw}(y)\right) \end{pmatrix} \tag{3}$$

where W_{int} corresponds to lB^2 filters, $\tilde{\mathrm{I}}_{ab}(y)$ is a $1 \times 1 \times lB^2$ vector, a and b are the space indices of $\tilde{\mathrm{I}}(y)$, h and w represent the numbers of blocks in row and column respectively, $\gamma(\cdot)$ is the reshape function that converts the $1 \times 1 \times lB^2$ vector to a $B \times B \times l$ block, $\kappa(\cdot)$ is the concatenation function that concatenates all these blocks to generate a whole image.

In this work, BL uses Eq.(2) and Eq.(3) to get the initial reconstruction, but each EL just uses Eq.(2) and Eq.(3) to get a reconstruction residual as shown in Figure 2. The initial reconstruction in the i^{th} EL is the reconstruction residual of the i^{th} EL plus the initial reconstruction of the $(i-1)^{th}$ EL or BL. Suppose the i^{th} group of measurements are obtained with n_i measurement bases, W_{int} in Eq.(2) corresponds to lB^2 filters of support $1 \times 1 \times n_i$.

3.2.3 Hourglass-shape deep reconstruction network

After getting the initial reconstruction, there is a nonlinear reconstruction process in the traditional BCS methods [14, 27]. In this work, a deep reconstruction network

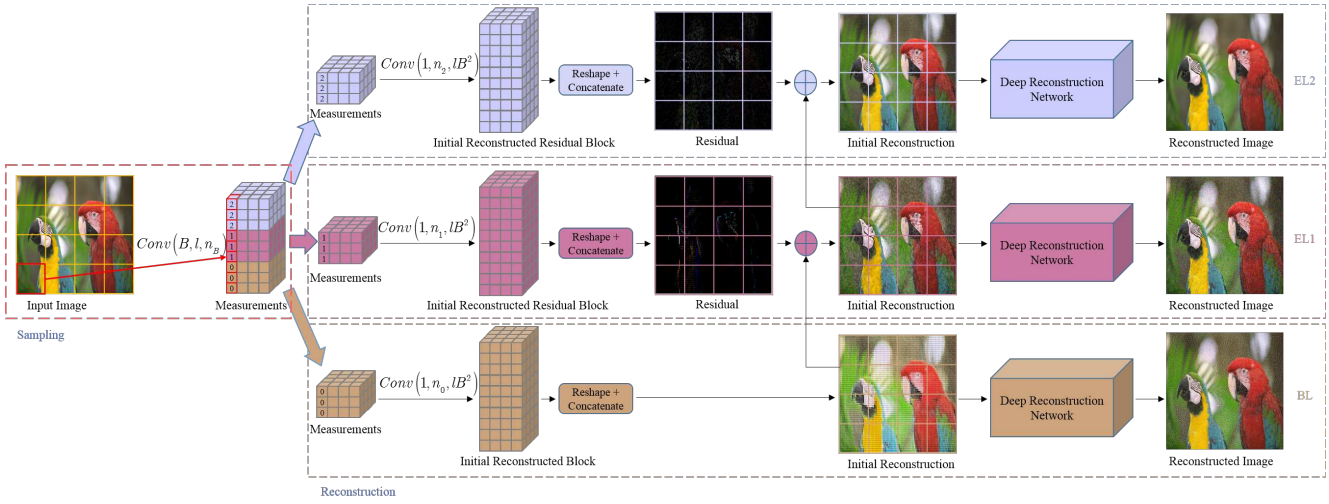

Figure 2. The network structure schematic of SCSNet with two ELs.

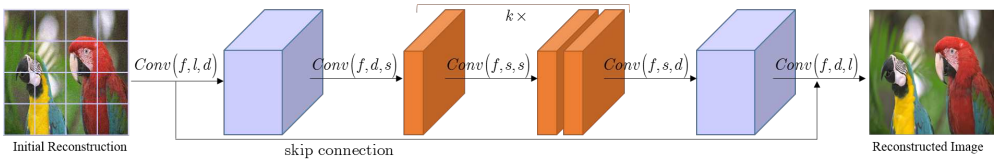

Figure 3. The network structure of Deep Reconstruction.

is used to further refine the reconstructed images in BL and ELs. Dong et al. [10] shows that hourglass-shape network has good performance with low computational complexity. In addition, many works [18] show residual learning can accelerate the network convergence speed and boost the network performance. Based on the existing works, our deep reconstruction network is a hourglass-shape residual learning network as shown in Figure 3.

The hourglass-shape residual learning network includes six kinds of operations, i.e. feature extraction, shrinking, non-linear mapping, expanding, feature aggregation, and skip connection. All these operations are convolution layers with different size filters except skip connection. This forms a symmetric structure, thick at the ends, and thin in the middle. We represent a convolution layer as $conv(f, in, out)$, where f, in, and out are the spacial size of the filters, the amount of the input channels, and the amount of the output channels, respectively. Then, feature extraction, shrinking, non-linear mapping, expanding and feature aggregation are represented as $conv(f, l, d)$, $conv(f, d, s)$, $conv(f, s, s)$, $conv(f, s, d)$, and $conv(f, d, l)$, where l is the amount of image channel. Note that $d >> s$, which ensures that the deep reconstruction is a compact hourglass-shape network. To increase the network non-linear, the non-linear mapping is cascaded k times. All these convolution layers are followed with a ReLu [29] activation layer except the last convolution layer. A skip connection is added between the initial and the final reconstruction.

3.2.4 Loss function

Suppose the reconstruction network has T initial reconstructions and T final reconstructions, we have $2T$ objectives to minimize. We adopt the mean square error (MSE) as the loss function to supervise each initial reconstruction and final reconstruction. With these constraints, all initial and final reconstructions are expected to correctly reconstruct the desired image, which can accelerate network convergence and boost the final reconstruction quality. Adaptive moment estimation (Adam) [19] is used to optimize all network parameters.

3.3. Fine granular scalability

Fine granular scalability is necessary, which increases the flexibility in applications. For example, if BL is for the sampling ratio of 0.01 and the first EL is for the sampling ratio of 0.05 in a well-trained model, it cannot be applied directly to reconstruct image at sampling ratio of 0.04. The fine granular scalability makes it possible to sample and reconstruct image at any sampling ratio with one model. To a given sampling ratio r, which is smaller than the sampling ration r_i of the i^{th} EL (for convenience, BL is treated as a specific EL) but larger than the sampling ratio r_{i-1} of the $(i-1)^{th}$ EL, we remove some unimportant measurement bases and the corresponding connections to the reconstruction network in the i^{th} EL. To obtain as good reconstruction as possible, we use a greedy algorithm to preserve the most

Algorithm 1 The greedy method for measurement base selection. M orders the measurement bases from least impact on PSNR to most impact.

Input: Validation set $\{x_j\}$, and the index i of EL
Output: the order M of measurement bases in the i^{th} EL

1: $M \leftarrow \emptyset, Z = \{1, 2, \cdots, L\}$;
2: **for** $k = 1$ to L **do**
3: $max\_psnr \leftarrow -\infty$;
4: **for** $z \in Z$ and $z \notin M$ **do**
5: $M' \leftarrow M \cup \{z\}$;
6: compute avg_psnr when the measurement bases indexed by M' are removed in the i^{th} EL;
7: **if** $avg\_psnr > max\_psnr$ **then**
8: $max\_psnr = avg\_psnr, max\_z \leftarrow z$;
9: **end if**
10: **end for**
11: $M \leftarrow M \cup \{max\_z\}$;
12: **end for**

important measurement bases.

Suppose the i^{th} EL uses L measurement bases that their indexes are represented as $Z = \{1, 2, \cdots, L\}$. When some measurement bases and their connections to the reconstruction network are removed, we hope the remaining measurements provide as good reconstruction as possible. Hence, to preserve the most important measurement bases, we solve the following optimization problem

$$\arg\min_{M} \sum_{j=1}^{N} \left(R_M^{(i)} \left(S_M^{(i)} \left(x_j \right) \right) + R^{(i-1)} \left(S^{(i-1)} \left(x_j \right) \right) - x_j \right)^2 \quad s.t.\ M \subset Z = \{1, 2, \cdots, L\} \tag{4}$$

where x_j is a validation sample, $S^{(i-1)}$ and $R^{(i-1)}$ are the measurements and the reconstruction of the $(i-1)^{th}$ EL, $S_M^{(i)}$ and $R_M^{(i)}$ are measurements and the reconstructed residual of the i^{th} EL after removing those measurements indexed by a subset M of Z.

We use a greedy method to solve Eq.(4). The idea is to select a best option in each step. For a given amount of measurement bases, the solution of Eq.(4) is those measurement bases that provide highest average PSNR. As illustrated in Algorithm 1, M is the order set of the measurement bases in the i^{th} EL, and it is empty in the beginning. In Step 3 to Step 11, we select only one index to move into M that has less impact on the average PSNR. That is, the index be moved into M in first is more unimportant than the index be moved into M in later. After L iterations, we obtain the order of the measurement bases in the i^{th} EL based on their importance to the reconstruction quality.

As the importance order of the measurement bases can be obtained by using Algorithm 1, it is easy to implement sampling and reconstruction at any desired sampling ratio with only one model and provides as good reconstruction as possible, which provides fine granular scalability.

4. Experiments

4.1. Dataset and implementation details

Similar to the traditional image CS methods [27, 22, 6, 40], the block size is set $B = 32$ and $l = 1$. In our experiment, SCSNet contains one BL and six ELs that corresponds to sampling ratio of 0.01, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. The filter size in the initial reconstruction is computed based on the sampling ratio and the block size as introduced in Subsection 3.2.2. In the deep reconstruction network, we set $f = 3$, $l = 1$, $d = 128$, $s = 32$, and $k = 13$ respectively. To optimize the network parameters, the learning ratios of the first 50, the 51 to 80 and the last 20 epochs are 10^{-3}, 10^{-4}, and 10^{-5}, respectively. The training data is the same with CSNet. That is, the training set (200 images) and test set (200 images) of the BSDS500 database [2] form the training dataset. Each image is cut into multiple patches of size 96×96. Finally, only 89600 patches are used to optimize the network parameters.

4.2. Comparison with the state-of-the-arts

4.2.1 Comparison with traditional methods

The compared traditional methods include: wavelet method (DWT) [27], total variation (TV) method [22], multi-hypothesis (MH) method [6], and group sparse representation (GSR) method [40]. CSNet [32] is also listed for comparison. All these methods are popular BCS methods. The implementation codes of the compared methods are downloaded from the author's websites and the default parameter settings are used in our experiments. We compare these methods on three popular test dataset, i.e. Set5 (5 images) [3], Set14 (14 images) [38] and the BSD100 (100 images) [25]. Note that all experiments are conducted on the Y channel of the YUV color space. Seven sampling ratios, i.e. 0.01, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5, are investigated. Both quantitative and qualitative comparisons are given.

The average PSNR and SSIM on the three test datasets are shown in Table 1. The best results are marked in bold font. The quantitative results show that SCSNet outperforms the five compared CS methods at all sampling ratios. Specially, compared with DWT, TV, MH, GSR, and CSNet, SCSNet gains by average 7.45 dB, 5.16 dB, 4.31 dB, 2.58 dB, and 0.50 dB, respectively, over seven sampling ratios and three datasets. The average SSIM also shows SCSNet is significantly superior to the five compared methods.

Figure 4 shows a visual quality comparison of image CS recovery in the case of sampling ratio of 0.2. We have magnified a subregion of each image to compare the reconstruction details of each image. Obviously, SCSNet achieves better visual quality than the traditional methods. Although the

Figure 4. Visual quality comparisons of CS recovery on *PPT3* from Set14 [38] in the case of sampling ratio = 0.2.

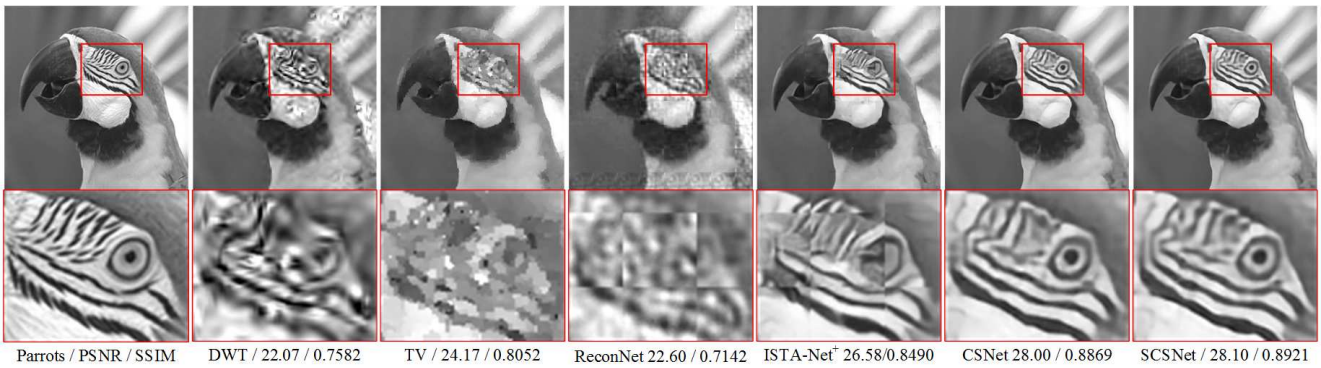

Figure 5. Visual quality comparisons of CS recovery on *Parrots* from Set11 [20] in the case of sampling ratio = 0.1.

visual differences between the reconstruction results of SCSNet and CSNet are small, SCSNet gets higher PSNR and SSIM values. All the experimental results demonstrate SCSNet not only has property of scalability but also has state-of-the-art performance.

4.2.2 Comparison with deep learning based methods

The compared deep learning based methods include: SDA [26], ReconNet [20], ISTA-Net [39], ISTA-Net$^+$ [39] and CSNet [32]. LAPRAN [36] gets high PSNR with a flexible resolution. However, the reconstruction results of LAPRAN have significant blocking artifact, and the pretrained models of LAPRAN have not been released. Therefore, we do not make comparison with LAPRAN in this paper. We follow [39] to use Set11 [20] and BSD68 [25] as the test images. Table 2 shows the average PSNR of different deep learning based methods on five sampling ratios. The results of SDA, ReconNet, ISTA-Net, and ISTA-Net$^+$ are taken from [39]. As shown, SCSNet obtains significantly higher average PSNR than the compared deep learning based methods at the five sampling ratios on Set11 and BSD68. Figure 5 shows a visual comparison between various image CS methods. As shown, both ReconNet and ISTA-Net$^+$ have blocking artifact. In contrast, SCSNet does not have blocking artifact and obtains better visual effect. SCSNet is the scalable extension of CSNet. SCSNet outperforms CSNet because it uses a better reconstruction network.

Table 2. Average PSNR comparison of different deep learning based image CS methods on Set11 [20] and BSD68 [25].

Data	Alg.	Sampling Ratio				
		0.5	0.4	0.3	0.1	0.01
Set11	SDA [26]	28.95	27.79	26.63	22.65	17.29
	ReconNet [20]	31.50	30.58	28.74	24.28	17.27
	ISTA-Net [39]	37.43	35.36	32.91	25.80	17.30
	ISTA-Net$^+$ [39]	38.07	36.06	33.82	26.64	17.34
	CSNet [32]	37.51	36.10	33.86	28.10	20.94
	SCSNet	**39.01**	**36.92**	**34.62**	**28.48**	**21.04**
BSD68	SDA [26]	28.35	27.41	26.38	23.12	-
	ReconNet [20]	29.86	29.08	27.53	24.15	-
	ISTA-Net [39]	33.60	31.85	29.93	25.02	-
	ISTA-Net$^+$ [39]	34.01	32.21	30.34	25.33	-
	CSNet [32]	34.89	32.53	31.45	27.10	22.34
	SCSNet	**35.77**	**33.86**	**31.87**	**27.28**	**22.37**

4.2.3 Running time comparison

Table 3 shows the average running time comparisons between various algorithms for reconstructing a 256×256 image at sampling ratio of 0.01 and 0.1. The running times of SAD and ReconNet are taken from [20]. The running times of ISTA-Net and ISTA-Net$^+$ are the average running time of seven sampling ratio taken from [39]. The running times

Table 1. Average PSNR and SSIM comparisons of different image CS algorithms on Set5 [3], Set14 [38] and BSD100 [25]

Data	Ratio	DWT [27]		TV [22]		MH [6]		GSR [40]		CSNet [32]		SCSNet	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	0.01	9.27	0.1402	15.53	0.4554	18.08	0.4472	18.87	0.4909	24.02	0.6378	**24.21**	**0.6468**
	0.05	14.27	0.3559	23.16	0.6678	23.67	0.6566	24.95	0.7270	29.32	0.8354	**29.74**	**0.8472**
	0.1	24.74	0.7680	27.07	0.7865	28.57	0.8211	29.99	0.8654	32.30	0.9015	**32.77**	**0.9083**
	0.2	30.83	0.8749	30.45	0.8709	32.08	0.8881	34.17	0.9257	35.63	0.9451	**36.15**	**0.9487**
	0.3	33.61	0.9050	32.75	0.9107	34.06	0.9158	36.83	0.9492	37.90	0.9630	**38.45**	**0.9655**
	0.4	35.32	0.9249	34.89	0.9363	35.65	0.9337	38.81	0.9626	39.89	0.9736	**40.44**	**0.9755**
	0.5	36.87	0.9409	36.75	0.9540	37.21	0.9482	40.65	0.9724	40.96	0.9784	**42.22**	**0.9820**
Set14	0.01	8.97	0.0989	15.26	0.3890	17.23	0.3970	17.87	0.4337	22.73	0.5556	**22.87**	**0.5631**
	0.05	14.52	0.2933	22.24	0.5815	21.64	0.5567	22.54	0.6140	26.65	0.7238	**26.92**	**0.7322**
	0.1	24.16	0.6798	25.24	0.6887	26.38	0.7282	27.50	0.7705	28.91	0.8119	**29.22**	**0.8181**
	0.2	28.13	0.7882	28.07	0.7844	29.47	0.8237	31.22	0.8642	31.86	0.8908	**32.19**	**0.8945**
	0.3	30.38	0.8389	30.12	0.8424	31.37	0.8694	33.74	0.9071	34.00	0.9276	**34.51**	**0.9311**
	0.4	31.99	0.8753	32.03	0.8837	33.03	0.9009	35.78	0.9336	35.95	0.9495	**36.54**	**0.9525**
	0.5	33.54	0.9044	33.84	0.9148	34.52	0.9239	37.66	0.9522	37.05	0.9607	**38.41**	**0.9659**
BSD100	0.01	9.63	0.1067	15.98	0.3995	18.21	0.4076	18.90	0.4431	23.69	0.5441	**23.78**	**0.5483**
	0.05	14.81	0.2935	23.05	0.5690	21.36	0.5169	22.16	0.5682	26.61	0.6908	**26.77**	**0.6972**
	0.1	23.46	0.6343	25.46	0.6612	25.16	0.6673	25.91	0.7071	28.40	0.7787	**28.57**	**0.7844**
	0.2	27.26	0.7516	27.58	0.7557	28.09	0.7746	29.18	0.8156	30.88	0.8681	**31.10**	**0.8731**
	0.3	29.23	0.8108	29.27	0.8191	29.85	0.8307	31.33	0.8723	32.89	0.9146	**33.24**	**0.9190**
	0.4	30.72	0.8524	30.86	0.8660	31.35	0.8695	33.20	0.9096	34.13	0.9250	**35.21**	**0.9470**
	0.5	32.17	0.8862	32.46	0.9019	32.86	0.9012	34.94	0.9359	36.09	0.9587	**37.14**	**0.9649**
Avg.		24.95	0.6535	27.24	0.7447	28.09	0.7513	29.82	0.7914	31.90	0.8445	**32.40**	**0.8507**

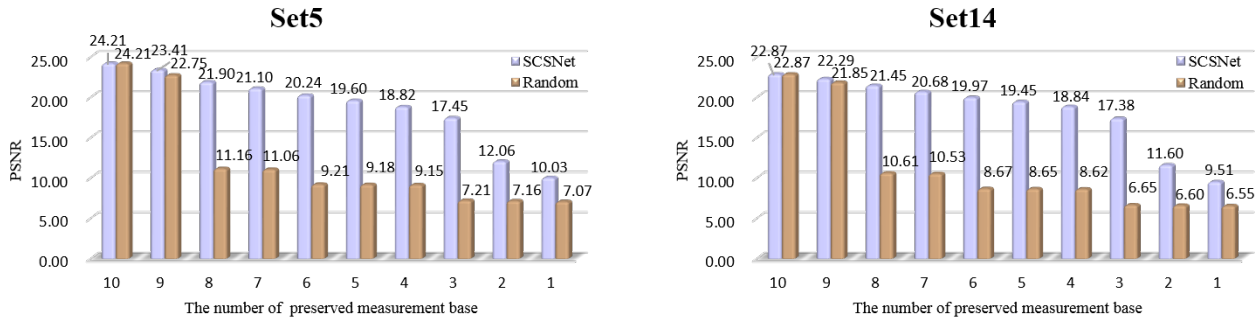

Figure 6. Average PSNR comparisons on Set5 [3] and Set14 [38] when preserving different number of measurement base of BL by the proposed method and the random method.

for DWT, TV, MH, GSR, and CSNet are the implementation times on the platform of an Intel Core i7-3770 CPU plus a NVIDIA GTX960 GPU with the codes download from the author's websites. SCSNet runs on the platform of an Intel Core i7-7700 CPU plus a NVIDIA GTX1080 GPU. As shown in Table 3, the four compared traditional methods take roughly several seconds to several minutes to reconstruct an image. This may be because they need repeated iterative operations. All the compared deep learning based methods run faster than those compared traditional methods. Although SCSNet runs slower than the compared deep learning based methods on GPU, SCSNet can run faster than ReconNet, ISTA-Net, and ISTA-Net$^+$ on CPU. In our experiment, SCSNet is implemented using the DagNN wrapper of MatConvNet package [34]. In practical application, we can reimplement SCSNet using the other deep learning framework for faster running speed.

Table 3. Average running time (in seconds) of various algorithms for reconstructing a 256×256 image.

Algorithm	Ratio = 0.01		Ratio = 0.1	
	CPU	GPU	CPU	GPU
DWT [27]	10.3176	-	10.5539	-
TV [22]	2.3349	-	2.5871	-
MH [6]	23.1006	-	19.0405	-
GSR [40]	235.6297	-	230.4755	-
SDA [26]	-	0.0045	-	0.0029
ReconNet [20]	0.5193	0.0244	0.5258	0.0195
ISTA-Net [39]	0.9230	0.0390	0.9230	0.0390
ISTA-Net$^+$ [39]	1.3750	0.0470	1.3750	0.0470
CSNet [32]	0.2950	0.0157	0.2941	0.0155
SCSNet	0.5103	0.1050	0.5146	0.1332

4.3. Study of fine granular scalability

SCSNet provides measurement base level of fine granular scalability. The greedy method for measurement base selection introduced in Subsection 3.3 is a data driven method.

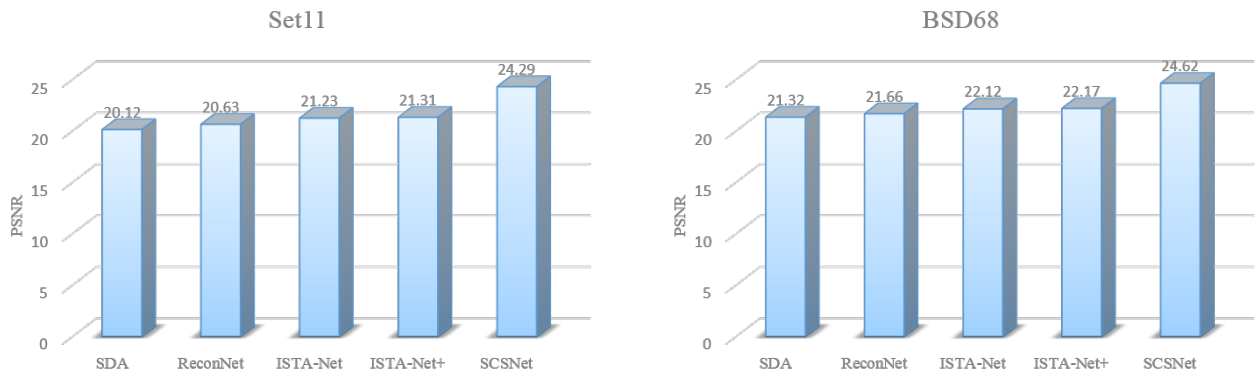

Figure 7. Average PSNR comparisons between different deep learning based image CS methods on Set11 [20] and BSD68 [25] in the case of sampling ratio = 0.04.

In our experiment, we use the 50 images from the validation set of the BSDS500 database [2] to determine the importance of each measurement base. For comparison, we also randomly determine the importance of each measurement base. The random seed is 10.

We implement progressive sampling and progressive reconstruction based on the importance of each measurement base. That is, the encoder acquires measurements with the most important measurement base in first, and gives to the decoder. The decoder gradually improves the reconstructed image quality as more measurements it receives. Figure 6 shows the results of progressive sampling and progressive reconstruction using the BL of SCSNet. As shown, the proposed greedy method is an effective way for measurement base selection because it provides significant higher average PSNR than the random selection method.

In addition, we also verify sampling and reconstruction at different sampling ratios by using one model. Table 4 shows the average PSNR at nine sampling ratios, which are obtained by using one well-trained model. The results of sampling ratio of 0.01, 0.05 and 0.1 correspond to the BL, the first and the second EL, respectively, while the other results are obtained by removing some measurement bases. As shown, SCSNet can provide good reconstruction for scalable sampling and scalable reconstruction with only one model. In Figure 7, the results of the compared deep learning based image CS methods are obtained by the models that are specially trained for sampling ratio of 0.04, while the results of SCSNet are obtained by removing some measurement bases of the well-trained model. As shown, SCSNet does not need to train the special model for sampling ratio of 0.04, but it still significantly outperforms the compared methods in this sampling ratio.

Table 4. The average PSNR of SCSNet at nine different sampling ratios using only one well-trained model.

Data	Sampling ratio								
	0.01	0.02	0.03	0.04	**0.05**	0.07	0.08	0.09	**0.1**
Set5	**24.21**	24.57	26.10	27.83	**29.74**	29.93	30.78	31.76	**32.77**
Set11	**21.04**	21.55	22.84	24.29	**25.85**	26.22	27.00	27.82	**28.52**
Set14	**22.87**	23.20	24.42	25.65	**26.92**	27.20	27.89	28.58	**29.22**
BSD68	**22.37**	22.86	23.73	24.62	**25.43**	25.90	26.38	26.86	**27.28**
BSD100	**23.78**	24.14	24.98	25.88	**26.77**	27.07	27.57	28.09	**28.57**

5. Conclusion

While observing the limitation of the existing deep learning based image CS methods, we have proposed a scalable convolutional neural network (dubbed SCSNet) for image compressed sensing. SCSNet is the first to implement scalable sampling and scalable reconstruction using CNN, which provides both coarse granular scalability and fine granular scalability. For coarse granular scalability, SCSNet is designed as a single sampling matrix plus a hierarchical reconstruction network that has a BL and multiple ELs. BL provides the basic reconstruction quality, while ELs improve the reconstructed image quality by generating the reconstruction residual and referencing the lower layers. An effective greedy method for measurement base selection has also been introduced. By removing some unimportant measurement bases and their corresponding connection to the reconstruction network, SCSNet achieves fine granular scalable sampling and reconstruction. Compared with the existing deep learning based methods, SCSNet provides scalable sampling and scalable reconstruction that provides the chance to implement sampling and reconstruction at any sampling ratio with only one model. Experimental results show that SCSNet has the state-of-the-art reconstruction quality while maintaining a comparable running speed with the existing deep learning based image CS methods.

6. Acknowledgement

This work was supported by the National Basic Research Program of China (2015CB351804), the National Natural Science Foundation of China under Grants 61872116, 61672188, 61572155, and the Alibaba Innovative Research (AIR) Program.

References

[1] Arash Amini and Farokh Marvasti. Deterministic construction of binary, bipolar, and ternary compressed sensing matrices. *IEEE Transactions on Information Theory*, 57(4):2360–2370, 2011.

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2011.

[3] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

[4] Lei Bo, Hancheng Lu, Yujiao Lu, Jianwen Meng, and Wenzhe Wang. FompNet: Compressive sensing reconstruction with deep learning over wireless fading channels. In *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2017.

[5] Jill M Boyce, Yan Ye, Jianle Chen, and Adarsh K Ramasubramonian. Overview of SHVC: scalable extensions of the high efficiency video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1):20–34, 2016.

[6] C. Chen, E. W Tramel, and J. E. Fowler. Compressed-sensing recovery of images and video using multihypothesis predictions. In *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)*, pages 1193–1198. IEEE, 2011.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[8] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[9] K. Q. Dinh, H. J. Shim, and B. Jeon. Measurement coding for compressive imaging using a structural measuremnet matrix. In *2013 IEEE International Conference on Image Processing*, pages 10–13. IEEE, 2013.

[10] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision – ECCV 2016*, pages 391–407, Cham, 2016. Springer International Publishing.

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

[12] Marco F Duarte, Mark A Davenport, Dharmpal Takbar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

[13] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007.

[14] L. Gan. Block compressed sensing of natural images. In *2007 15th International Conference on Digital Signal Processing*, pages 403–406. IEEE, 2007.

[15] X. Gao, J. Zhang, W. Che, X. Fan, and D. Zhao. Block-based compressive sensing coding of natural images by local structural measurement matrix. In *2015 Data Compression Conference*, pages 133–142. IEEE, 2015.

[16] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Transactions on Information Theory*, 52(9):4036–4048, 2006.

[17] Ronan Kerviche, Nan Zhu, and Amit Ashok. Information-optimal scalable compressive imaging system. In *Computational Optical Sensing and Imaging*, pages CM2D–2. Optical Society of America, 2014.

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[19] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. ReconNet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 449–458, 2016.

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

[22] C. Li, W. Yin, and Y. Zhang. Tval3: Tv minimization by augmented lagrangian and alternating direction agorithm 2009. Available: `http://www.caam.rice.edu/$\sim$optimization/L1/TVAL3/`.

[23] Weizhi Lu, Tao Dai, and Shu-Tao Xia. Binary matrices for compressed sensing. *IEEE transations on signal processing*, 66(1):77, 2018.

[24] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[25] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE, 2001.

[26] Ali Mousavi, Ankit B. Patel, and Richard G. Baraniuk. A deep learning approach to structured signal recovery. In *Communication, Control, and Computing*, pages 1336–1343, 2016.

[27] S. Mun and J. E. Fowler. Block compressed sensing of images using directional transforms. In *2009 16th IEEE international conference on image processing (ICIP)*, pages 3021–3024. IEEE, 2009.

[28] Sungkwang Mun and James E Fowler. DPCM for quantized block-based compressed sensing of images. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1424–1428. IEEE, 2012.

[29] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the*

27th International Conference on Machine Learning (ICML-10), pages 807–814, 2010.

[30] Deanna Needell and Joel A Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[31] Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. *IEEE transactions on medical imaging*, 37(6):1488–1497, 2018.

[32] Wuzhen Shi, Feng Jiang, Shengping Zhang, and Debin Zhao. Deep networks for compressed image sensing. In *Multimedia and Expo (ICME), 2017 IEEE International Conference on*, pages 877–882. IEEE, 2017.

[33] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[34] A. Vedaldi and K. Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 689–692. ACM, 2015.

[35] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.

[36] Kai Xu, Zhikang Zhang, and Fengbo Ren. Lapran: A scalable laplacian pyramid reconstructive adversarial network for flexible compressive sensing reconstruction. In *European Conference on Computer Vision*, pages 491–507. Springer, 2018.

[37] Wenbin Yin, Xiaopeng Fan, Yunhui Shi, Ruiqin Xiong, and Debin Zhao. Compressive sensing based soft video broadcast using spatial and temporal sparsity. *Mobile Networks and Applications*, 21(6):1002–1012, 2016.

[38] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[39] Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2018.

[40] J. Zhang, D. Zhao, and W. Gao. Group-based sparse representation for image restoration. *IEEE Transactions on Image Processing*, 23(8):3336–3351, 2014.