

# Learning Memory Augmented Cascading Network for Compressed Sensing of Images

Jiwei Chen<sup>1</sup>[0000-0002-7835-5704], Yubao Sun<sup>1</sup>[0000-0002-0462-3729],  
Qingshan Liu<sup>1</sup>[0000-0002-5512-6984], and Rui Huang<sup>2,3</sup>[0000-0002-7950-1662]

<sup>1</sup> Jiangsu Key Laboratory of Big Data Analysis Technology, Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology

{jiweichen1994, sunyb, qslu}@nuist.edu.cn

<sup>2</sup> School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen  
ruihuang@cuhk.edu.cn

<sup>3</sup> Shenzhen Institute of Artificial Intelligence and Robotics for Society

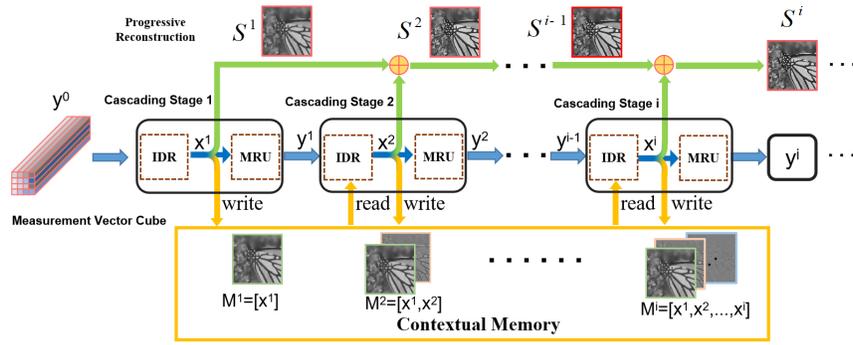
**Abstract.** In this paper, we propose a cascading network for compressed sensing of images with progressive reconstruction. Specifically, we decompose the complex reconstruction mapping into the cascade of incremental detail reconstruction (IDR) modules and measurement residual updating (MRU) modules. The IDR module is designed to reconstruct the remaining details from the residual measurement vector, and MRU is employed to update the residual measurement vector and feed it into the next IDR module. The contextual memory module is introduced to augment the capacity of IDR modules, therefore facilitating the information interaction among all the IDR modules. The final reconstruction is calculated by accumulating the outputs of all the IDR modules. Extensive experiments on natural images and magnetic resonance images demonstrate the proposed method achieves better performance against the state-of-the-art methods.

**Keywords:** Compressed Sensing, Cascading Network, Contextual Memory, Progressive Reconstruction

## 1 Introduction

Compressed sensing (CS) [5] is a well-known signal sensing technology, which attempts to directly sense the compressed signal. The basic principle of CS is that a  $N$ -dimensional sparse signal  $x$  with  $K$  non-zero transforming coefficients can be recovered from only about  $O(K \log(N/K))$  linear projection measurements [3, 4]. The CS technology has also achieved great success in many imaging systems, such as shortwave-infrared cameras [13, 22], compressive magnetic resonance imaging (MRI) [20, 39, 40], transmission electron microscopy, and snapshot compressive imaging [19, 37], because it can potentially improve the imaging systems by reducing the numbers of measurement, imaging time and storage space.

The core problem of compressed sensing of images is how to reconstruct the underlying image from the received measurement. Many approaches have been



**Fig. 1.** The cascading network architecture of the proposed MAC-Net.

proposed for this problem, and they can be broadly divided into two categories. The first category is to solve a sparsity-regularized nonlinear problem based on iterative optimization, including greedy pursuit methods [26, 33, 35] and convex relaxation methods [6, 8, 23, 34]. Specifically, greedy pursuit methods estimate the ideal sparse signal by selecting one or more columns in the sensing matrix that are most relevant to the measurement vector, and then calculate the residual measurement vector for subsequent iterations. Convex relaxation methods take the  $l_1$  norm as the sparsity metric, and the reconstruction can be obtained by means of many well-developed convex optimization methods. Since large-scale matrix multiplications are needed in the convex optimization, convex relaxation methods are very time-consuming. The second category of methods is to directly learn an inverse reconstruction mapping from the measurement vector to the original image with a deep network [16, 18, 30]. In order to improve the mapping accuracy, a lot of works leverage on designing more sophisticated and large networks. However, simply stacking complex network modules does not necessarily make the reconstruction result more accurate.

To deal with these issues, we propose a new cascading network for progressive CS image reconstruction, named Memory Augmented Cascading Network (MAC-Net). As shown in Fig. 1, the proposed network is composed of three kinds of modules, i.e., incremental detail reconstruction (IDR) modules, measurement residual updating (MRU) modules and a contextual memory (CM) module. The IDR module is dedicated to predicting the incremental reconstruction from the input measurement vector. The MRU is designed to calculate the residual measurement based on the reconstruction results of the former IDR module and feed it into the next IDR module. The CM module is used to augment the capacity of the IDR modules. Specifically, the current IDR module can easily access informative features in the memory through the reading operation, therefore facilitating the reconstruction of the remaining details. All these modules are trained in an end-to-end manner under the guidance of a unified loss function. The final reconstruction is obtained by accumulating outputs of all IDR modules. The code

of MAC-Net are available at <https://github.com/DFLYan/MAC-Net>. Our main contributions are summarized below.

1. The proposed MAC-Net decomposes the complex reconstruction mapping as progressive reconstruction, which can effectively reduce the learning difficulty and boost the reconstruction quality.
2. The CM module is designed to augment the capacity of IDR modules and enrich the interactions between them, therefore enhancing the reconstruction performance of each IDR module.
3. MAC-Net is flexible to achieve scalable reconstruction by using different numbers of modules according to practical demand. The extensive experiments on natural images and MRI images verify its promising performance.

## 2 Related Work

Mathematically, the compressed measurement vector  $y \in R^M$  of an image  $x \in R^N$  can be represented as the linear observation equation  $y = \Phi x + \varepsilon$ , where  $\Phi \in R^{M \times N}$  is the sensing matrix with  $M \ll N$ , such as the random Gaussian matrix and partial fourier transform matrix, and  $\varepsilon$  is the measurement noise.  $M/N$  is termed as the measurement rate. Compressed sensing reconstruction refers to the problem of recovering the original image  $x$  from the measurement vector  $y$ . Many methods have been proposed to solve this problem, and they can be grouped into the following two categories.

**Iterative optimization based methods** This category of methods try to find the sparsity solution to the underdetermined linear observation Eq. (1) by iterative optimization. When the  $l_0$  norm is used for the sparsity metric, the induced reconstruction is a non-convex and NP-hard problem [9]. Some greedy pursuit algorithms, such as the orthogonal matching pursuit (OMP) [33], regularized OMP (ROMP) [27] and CoSaMP [26], are adopted to find an approximate solution based on the greedy rule. Greedy pursuit algorithms have the advantages of fast calculation and easy implementation. However, they usually require a high measurement rate for reliable reconstruction.

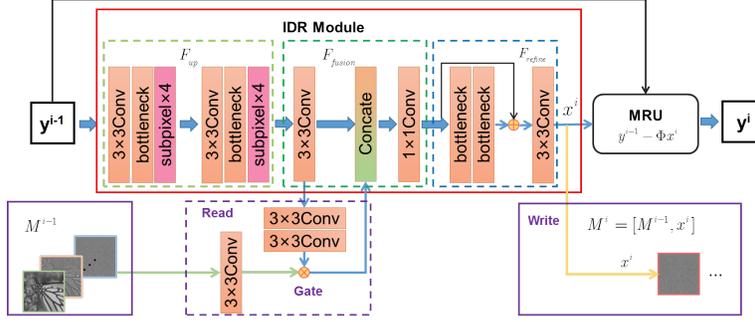
Different from the greedy methods, the convex optimization methods relax the reconstruction problem by replacing the non-convex  $l_0$  norm with the convex  $l_1$  norm [10]. The convex optimization of  $l_1$  norm solution can usually recover a good reconstruction to the original signal [7]. Many reconstruction methods are developed based on the principle of well-developed convex optimization algorithms, such as  $l_1$  norm minimization [2] by using the iterative shrinkage thresholding algorithm (ISTA), total variation minimization by augmented lagrangian and alternating direction algorithms [6]. Convex optimization methods can obtain superior reconstruction quality over greedy optimization methods. However, each iteration in convex optimization methods involves large-scale matrix multiplication, so they are computationally expensive.

**Deep network based methods** The main idea of deep network based methods is to learn the inverse mapping from the compressed measurement to the reconstruction, so that the test image can be fast reconstructed by simply passing the low-dimensional measurement vector through the learnt network. Many deep networks have been designed for CS reconstruction. DeepInverse [24] uses a deep convolutional network to predict the reconstruction. DAGAN [39] learns a deep de-aliasing generative adversarial network for CS-MRI (Magnetic Resonance Imaging) reconstruction. Specifically, a U-Net based generator is used to refine the reconstruction by predicting the incremental details. ReconNet [18] uses a deep convolutional network to directly learn the mapping relationship between the measurement vector and image blocks, and obtains the finally reconstructed image by assembling each block’s reconstruction. SCGAN [30] proposes a sub-pixel convolutional generative adversarial network for the reconstruction, where adversarial learning is beneficial for capturing the inherent image distribution for reconstruction. DR<sup>2</sup>-Net [41] exploits the linear layer and multiple residual blocks to learn the reconstruction mapping. NLR-CSNet [31] attempts to learn a network for reconstructing image sequences from the measurement vectors without pre-training. In order to combine the merits of iterative optimization based methods and deep learning based methods, ISTA-Net [16] unrolls the classical ISTA optimization into learnable network modules and learns all these network modules in an end-to-end manner. ADMM-Net [40] converts the alternating direction method of multipliers (ADMM) algorithm to the corresponding deep architectures. The above methods mostly leverage a single deep network to learn the mapping relationship between the measurement vector and the original image. According to the CS theory, since the image residuals are often more compressible, it is easier to reconstruct the residuals than the original image [4, 15, 32]. So, instead of one step prediction, MAC-Net proposed in this paper constructs multiple stages to continuously approximate the measurement vector for reconstructing the underlying image. Each stage learns the mapping relationship between the residual of the measurement vector and the residual of the reconstructed image. The progressive reconstruction can reduce the difficulty of network learning, which is conducive to improving the reconstruction quality.

### 3 Memory Augmented Cascading Reconstruction

#### 3.1 Network Architecture

MAC-Net aims to recover the original image from the given measurement vector progressively. Fig.1 illustrates the proposed MAC-Net, which decomposes the complex reconstruction mapping into multiple cascading stages. At each cascading stage, the IDR module predicts the current incremental reconstruction and engenders a residual, i.e, the remaining residual of the target image that has not been approximated. The measurement vector is then updated to reflect the remaining part and fed into the next IDR module, the new incremental reconstruction can be yielded.



**Fig. 2.** Illustration of one cascading stage and its interaction with the CM module.

In order to further the reconstruction quality, all the incremental reconstructions of the former stages are written into contextual memory and a reading operation is designed to access the associated information for the current IDR module to promote the reconstruction of the remaining details. With the involvement of more stages, the reconstruction is continuously improved, and finally, high-quality reconstruction can be achieved. The above computation flow can be expressed as

$$\begin{aligned} x^i &= G^i(y^{i-1}, M^{i-1}), y^i = U^i(x^i, y^{i-1}), \\ M^i &= [M^{i-1}, x^i], 1 \leq i \leq K. \end{aligned} \quad (1)$$

$G^i$  denotes the IDR module of the  $i$ -th stage, which exploits both the residual measurement vector  $y^{i-1}$  and the contextual memory  $M^{i-1}$  to generate the incremental reconstruction  $x^i$ .  $U^i$  denotes the MRU module to generate the new measurement residual  $y^i$  according to  $x^i$ , and  $M^i$  denotes the contextual memory information at the  $i$ -th cascading stage.  $K$  is the total number of cascading stages. For the first cascading stage,  $y^0$  is the given measurement vector and  $M^0$  has empty memory. The final reconstruction accumulates the incremental reconstructions of all cascading stages and is calculated by

$$S^K = \sum_{i=1}^K x^i. \quad (2)$$

Due to this cascading architecture, MAC-Net can obtain scalable reconstruction results by choosing different  $K$  according to practical demands. In the following, we will illustrate the detailed architecture of each cascading stage.

### 3.2 Single Cascading Stage

Fig. 2 illustrates a single cascading stage in detail, which mainly includes an IDR module, a MRU module and the interaction with the CM module. The  $i$ -th cascading stage takes the measurement vector  $y^{i-1}$  as the input, and the IDR module outputs the incremental reconstruction  $x^i$ , while the MRU module calculates the associated residual measurement vector  $y^i$  for the next stage.

**The IDR Module** predicts the incremental reconstruction  $x^i$  from the measurement vector  $y^{i-1}$  with the assistance of contextual memory  $M^{i-1}$ . The IDR module consists of three processing blocks, i.e., the resolution upsampling  $F_{up}$ , contextual fusion  $F_{fusion}$ , and high-resolution refinement  $F_{refine}$ , and its whole process can be formulated as,

$$x^i = F_{refine}(F_{fusion}(F_{up}(y^{i-1}), M^{i-1})). \quad (3)$$

$F_{up}$  accepts the measurement vector  $y^{i-1}$  as the input and increases the resolution of the feature maps to make it the same as the original image. Different from the conventional setting of taking CS measurement over the whole image with size  $W \times H$ , we divide the image evenly into  $\lceil W/16 \rceil \times \lceil H/16 \rceil$  sub-blocks with size  $16 \times 16$  and perform an  $m$ -dimensional CS measurement on each sub-block. These sensed measurement vectors are rearranged as a data cube with the size of  $\lceil W/16 \rceil \times \lceil H/16 \rceil \times m$  according to the order of their spatial positions. This way, we can represent the measurement  $y^{i-1}$  as a 3-D data cube as in [38], and use a  $3 \times 3$  convolution instead of the fully connected layer to process the whole measurement, which can reduce the number of network parameters. Then, a bottleneck residual block [11] is used to extract the feature maps and upsample their resolutions by a sub-pixel convolution layer [28] with a scaling factor of 4. Repeating this process twice can generate the feature maps with the same resolution as the original image.

$F_{fusion}$  is designed to interact with the CM module to read informative features and concatenates them with the features from the output of  $F_{up}$ . The detailed interaction with the CM module will be described in sub-section 3.3.

$F_{refine}$  processes the feature maps by two bottleneck residual blocks and employs a convolution operation to adjust the number of channels of the feature maps, and finally outputs the incremental reconstruction  $x^i$ . It should be noted that we use a large  $5 \times 5$  convolution kernel for coarse reconstruction in the first stage, and a small  $3 \times 3$  convolution kernel for detailed reconstructions in subsequent stages. The  $1 \times 1$  convolution in the fusion layer remains unchanged. Although the CS measurement is conducted in a block-wise manner, the IDR module outputs the entire reconstruction  $x^i$  in a single forward computation, thereby eliminating the blocking artifacts.

**The MRU Module** updates the measurement vector according to the current incremental reconstruction so that it can reflect the residual. With regard to the  $i$ -th stage, the residual measurement vector is updated as:

$$y^i = y^{i-1} - \Phi x^i. \quad (4)$$

$\Phi$  has a block-diagonal structure corresponding to the block-by-block CS measurement, which can significantly reduce the computation complexity of the updating equation. In fact, by tracking back to the first stage with  $y^0$  as the input, Eq. 4 can be rewritten as  $y^i = y^0 - \Phi S^i$ . Therefore, as the cascading stage continues, the norm (e.g.  $l_1$  norm) of the residual measurement vector will decrease continually, making it possible to pursue all the information involved in the measurement vector for reconstruction.

### 3.3 Contextual Memory Augmentation

In order to enrich the information flow in our cascading network, we augment the sequential links between the network modules with the CM module, so that the current IDR module can access all the former predictions and extract the relevant information for reconstruction [25, 29]. We define the writing and reading operations of the CM module to solve the problem of memory update and usage. Therefore, each cascading stage can utilize the contextual memory more efficiently, and the information interaction will be facilitated between the cascading stages in the network.

The reading operation is to extract the informative features for aiding the processing of the IDR module. Specifically, we use the feature maps  $f_q$  in the IDR module as a query and embed the query by convolution operations. The reading gate is designed to filter out the informative context  $f_m$  from the memory according to the embedded query from the IDR module, which is computed as,

$$\begin{cases} A = \text{softmax}(\text{conv}(\text{conv}(f_q))), \\ f_m = A \odot \text{conv}(M^{i-1}), \end{cases} \quad (5)$$

where  $\text{conv}$  is a  $3 \times 3$  convolution operation. The function of  $\text{softmax}$  is to normalize the embedded query into a probability distribution  $A$ .  $\odot$  denotes the element-wise multiplication operation. It should be noted that  $\text{softmax}$  is used in the channel dimension, and it means that all the values in one pixel's position of a 3-D feature map add up to 1. In fact,  $A$  can be seen as the 3-D attention maps. Different from the 1-D channel attention [12] and 2-D spatial attention [14], and the simple tensor product of the spatial and channel attention [36],  $A$  can attend each entry of 3-D feature maps  $\text{conv}(M^{i-1})$  adaptively. Thereby, the reading gate can extract informative context  $f_m$  into the  $F_{fusion}$  block of the IDR module, which helps the IDR module to reconstruct the remaining residual.

The writing operation of the  $i$ -th stage is to add the current incremental image  $x^i$  into the CM module and update the memory as  $M^i = [M^{i-1}, x^i]$ .

### 3.4 Network Loss and Learning

Assuming the training dataset consists of  $N$  pairs  $\{x_p, y_p^0 = \Phi x_p\}_{p=1}^N$ , and given the measurement vector  $y_p^0$ , the output of MAC-Net is denoted as  $S_p^K$ , where  $K$  is the number of cascading stages.  $y_p^i (1 \leq i \leq K)$  denotes the measurement residual of the  $p$ -th training image induced in the  $i$ -th stage of MAC-Net. Each stage is supposed to produce meaningful incremental reconstruction and contribute to the improvement of the reconstruction quality. Therefore, we impose weak supervision on each stage by minimizing the  $l_1$  norm of the updated residual measurement vector, so that the quality of the cumulative reconstruction can be

continuously improved. The unified loss function of MAC-Net is defined as,

$$\begin{aligned} \min_{\{G^i, M^i\}} \sum_{p=1}^N \left\{ \|x_p - S_p^K\|_1 + \lambda \sum_{i=1}^K \beta_i \|y_p^i\|_1 \right\} \\ \text{s.t. } x_p^i = G^i(y_p^{i-1}, M_p^{i-1}), y_p^i = U^i(x_p^i, y_p^{i-1}), \\ M_p^i = [M_p^{i-1}, x_p^i], S_p^K = \sum_{i=1}^K x_p^i, i = 1, \dots, K, \end{aligned} \quad (6)$$

where  $\lambda$  is the regularization parameter to balance the importance of the reconstruction error term and the measurement error term, and  $\beta_i$  is the weight associated with the  $i$ -th stage. Mean absolute error (MAE) is used to measure the reconstruction error and measurement error. Although there are no parameters to be optimized in the MRU module, we still need to define the gradient operation for the MRU module due to the chain rule. According to Eq. 4, the gradient of MRU with respect to  $x^i$  is the negative transpose of sensing matrix  $-\Phi^T$  and the gradient with respect to  $y^{i-1}$  is simply the constant 1. The Adam optimizer [17] is adopted to update the parameters of MAC-Net.

## 4 Experimental Results and Analysis

The proposed MAC-Net is evaluated on two kinds of image datasets. One is natural image datasets including Set11 [18] and BSD68 [21], and the other is the magnetic resonance image dataset, i.e., the MICCAI 2013 grand challenge dataset. Multiple state-of-the-art CS reconstruction methods are also tested for comparison, including TVAL3 [6], ReconNet [18], SCGAN [30], DR<sup>2</sup>-Net [41], ISTA-Net<sup>+</sup> [16], DeepADMM [40] and DAGAN [39].

In the natural image reconstruction experiments, followed by ReconNet [18] and DR<sup>2</sup>-Net [41], we choose their training set, Train91 dataset, for the network training. This image set contains 91 color images. Firstly, we convert these images into grayscale images by extracting the luminance component. In order to increase the number of samples, we randomly crop the image blocks with the size of  $96 \times 96$  from 91 grayscale images, and by conducting flipping (up-down and left-right) and rotating ( $0^\circ, 90^\circ, 180^\circ$  and  $270^\circ$ ) operations the training set is further augmented. In the magnetic resonance image reconstruction experiments, the MICCAI 2013 dataset has a training set with 15992 2D images including brain tissues. To ensure the integrity of the structure, we use complete MR images instead of cropped blocks. Data augmentation is also used during the training. SET11 [18], BSD68 [21] and MICCAI 2013 grand challenge dataset are used for testing. The Peak Signal to Noise Ratio (PSNR) is adopted as a quantitative evaluation criterion of the experimental results. The proposed MAC-Net is implemented upon the tensorflow platform [1] and the Adam optimizer [17] is adopted for updating network parameters with an initial learning rate of 0.0001. During training, the learning rate is reduced to 80% every 50 epochs, and we used 200 epochs in total. We run the experiments on Ubuntu Linux with GeForce GTX TITAN X GPU.

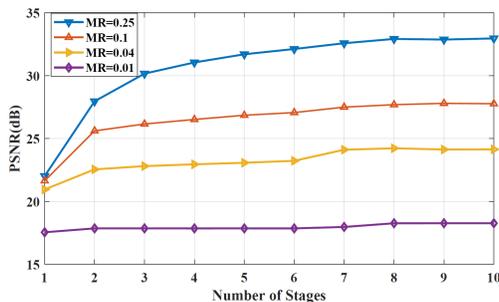


Fig. 3. The curves of average PSNR versus the number of cascading stages on SET11.

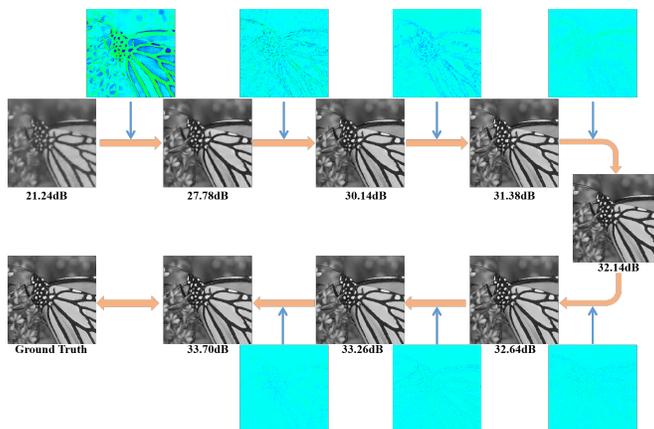
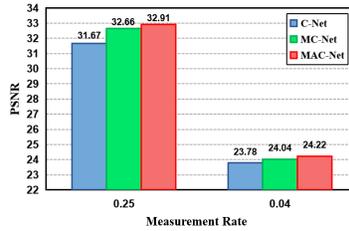


Fig. 4. Visualization of cascading reconstruction of the image Monarch from SET11.

#### 4.1 Ablation studies

In order to better understand the behaviour of MAC-Net, we conduct two groups of ablation studies, the first group is to take a deeper insight to the procedure of cascading reconstruction, and the second group is to evaluate the influence of memory augmentation on the reconstruction performance.

We first evaluate the performance of cascading reconstruction. As described in Section 3, using more cascading stages is helpful for improving the reconstruction quality, but the complexity of the network will be also increased. Fig.3 illustrates average PSNR curves as a function of the number of cascading stages at four measurement rates on the SET11 dataset. At the measurement rates of 0.25 and 0.1, each additional stage can bring some PSNR improvement. Specially, there is a sharp rise from 22.01 dB to 32.91 dB at the measurement rate of 0.25. Although the PSNR increments per stage at the measurement rates of 0.04 and 0.01 are smaller than those at the measurement rates of 0.25 and 0.1, the cascading reconstruction is also beneficial for exploiting the limited information in the measurement vector for reconstruction. We can also find out that 8 stages



**Fig. 5.** The ablation studies of the contextual memory and the reading operation.

can well balance the quality and efficiency of network reconstruction. More than 8 stages will only bring very slight performance improvement and even reduce the reconstruction quality. Taking the Monarch image from the SET11 dataset as an example, Fig.4 visualizes the procedure of its cascading reconstruction at the measurement rate of 0.25. The images above the arrows represent the incremental reconstruction generated by each stage, while the images along the direction of the arrows represent the cumulative reconstruction at each stage. It can be seen that the incremental details generated by each stage can refine the reconstruction progressively and the PSNR value increases stage by stage.

The second group of ablation studies are designed for determining the effect of contextual memory augmentation on the reconstruction performance. MAC-Net uses the contextual memory to augment the links between the cascading stages. In particular, the current stage can read all the previous reconstructions through the CM module, and the reading operation is used to extract the informative features for boosting the reconstruction of current stage.

To verify the benefits of the contextual memory and reading operation, we remove the CM module and reading operation from MAC-Net to form two simplified versions, including MAC-Net without the CM module (named C-Net), MAC-Net with the CM module but without the reading operation (named MC-Net). In the case of without the reading operation, the reading gate is not used and the convolutional feature maps from the contextual memory are directly concatenated with the feature maps in the IDR module. Fig. 5 shows the average PSNR values of MAC-Net and its two simplified versions at the measurement rates of 0.25 and 0.04 upon the SET11 dataset. It can be seen that C-Net has the lowest PSNR values, and it in turn demonstrates that the CM module is beneficial to improve the reconstruction performance. MAC-Net is superior to MC-Net, which shows that the use of reading gate in the read operation can further improve the reconstruction quality. With both the CM module and the reading gate, the PSNR value of MAC-Net is significantly improved by 1.24dB and 0.44dB at the measurement rates of 0.25 and 0.04 respectively.

## 4.2 Results on natural images

In the following experiments, we evaluate the reconstruction performance of MAC-Net on natural images and compare it with several state-of-the-art meth-

**Table 1.** Reconstruction results on SET11 at four measurement rates.

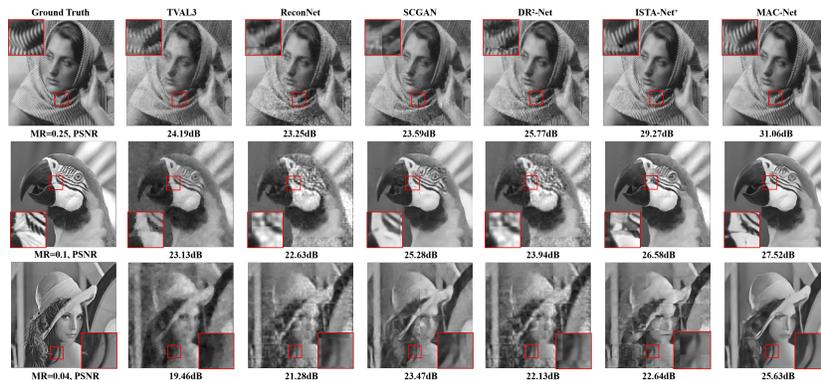
Algorithm	Measurement Rate			
	0.25	0.10	0.04	0.01
TVAL3 [6]	27.84	22.84	18.39	11.31
ReconNet [18]	25.54	22.68	19.99	17.27
SCGAN [30]	27.19	24.80	22.18	<b>18.43</b>
DR <sup>2</sup> -Net [41]	28.66	24.32	20.63	17.27
ISTA-Net <sup>+</sup> [16]	32.57	26.64	21.31	17.34
MAC-Net	<b>32.91</b>	<b>27.68</b>	<b>24.22</b>	18.26

**Table 2.** Reconstruction results on BSD68 at four measurement rates.

Algorithm	Measurement Rate			
	0.30	0.25	0.10	0.04
TVAL3 [6]	22.68	21.91	19.84	18.28
ReconNet [18]	27.53	25.31	24.15	21.66
SCGAN [30]	26.22	25.91	24.10	22.25
ISTA-Net <sup>+</sup> [16]	<b>30.34</b>	29.36	25.33	22.17
MAC-Net	30.28	<b>29.42</b>	<b>25.80</b>	<b>23.62</b>

ods including TVAL3 [6], ReconNet [18], SCGAN [30], DR<sup>2</sup>-Net [41] and ISTA-Net<sup>+</sup> [16]. Among these methods, TVAL3 is the representative iteration optimization based method, ISTA-Net<sup>+</sup> is the most recent deep learning based reconstruction method. Table 1 and Table 2 reports the experimental results of these methods on Set11 and BSD68 respectively. It can be seen from these two tables that MAC-Net can almost surpass all other competing methods. Especially at the measurement rate of 0.04, MAC-Net outperforms ISTA-Net<sup>+</sup> with a large margin of 2.91dB on the SET11 dataset and 1.45dB on the BSD68 dataset.

Some reconstructed images of Set11 and BSD68 are also presented in Fig.6 and Fig.7 respectively. Some patches are zoomed in for a clear comparison of local image structures. MAC-Net can recover richer image structures and texture details than all the other methods. For instance, in the Barbara image, it is relatively difficult to reconstruct the texture on the headscarf, due to the complex variation of the pixel value. ReconNet and SCGAN fail to recover the texture pattern. The basic texture pattern can be reconstructed by TVAL3, but a lot of noise is introduced. Although DR<sup>2</sup>-Net and ISTA<sup>+</sup>-Net can recover the main structure, the reconstructed images still lack fine details. The reconstructed images by MAC-Net have more texture details with the best visual quality. At the low measurement rate of 0.04, MAC-Net shows an obvious advantage over the other methods, and it can still recover meaningful image structures. The superiority of MAC-Net is mainly due to the network architecture of cascading reconstruction and memory augmentation. In addition, due to block-wise reconstruction, all the other deep-learning based methods suffer from blocking artifacts, especially at the measurement rates of 0.1 and 0.04. MAC-Net predicts



**Fig. 6.** Visual comparisons between multiple algorithms upon Barbara, Parrot and Lena images from SET11 at CS measurement rate of 0.25 (the top group), 0.1 (the middle group) and 0.04 (the bottom group).

the ensemble image from the measurement vectors of all blocks through a single forward computation, thereby getting rid of blocking artifacts.

In the practical application of compressed sensing, the robustness to noise is also a criterion for measuring the quality of the algorithm. Therefore, we perform a series of experiments based on the Set11 dataset. Gaussian noise with a mean value of zero and five levels of variances including 0.01, 0.05, 0.1, 0.25 and 0.5 is added to the measurement vectors, and we then feed them into the trained network which has been trained with the clean data. The result is shown in the Fig.8 and we can see that MAC-Net is more robust to the noise.

### 4.3 Compressive MRI reconstruction

We further conduct the reconstruction tests on Magnetic Resonance Imaging, which is one of the most widely used fields of compressed sensing. The MR image is sampled in the  $k$ -space. In order to reduce the acquisition time, compressive MRI undersamples the  $k$ -space by partial Fourier sampling. In this case, the sensing matrix  $\Phi$  is defined as  $\Phi = PF$ , where  $P$  is the under-sampling matrix and  $F$  denotes the discrete Fourier transform. Due to this specific sampling mechanism of Compressive MRI, we slightly adjust the network architecture of MAC-Net. Specifically, we add a layer to obtain the initial reconstruction by the inverse Fourier transform from the zero-filled undersampled  $k$ -space measurements. The initial reconstruction has the same spatial resolution as the original images, the operations for resolution enhancement are not required anymore. The cascading architecture and the contextual memory augmentation are kept the same.

In the experiments, we compare it with DeepADMM [40] and DAGAN [39], because both of them achieve great success on compressive MRI reconstruction and obtained state-of-the-art results on the MICCAI 2013 grand challenge dataset. The reconstruction results are shown in the Table 3. At the measurement

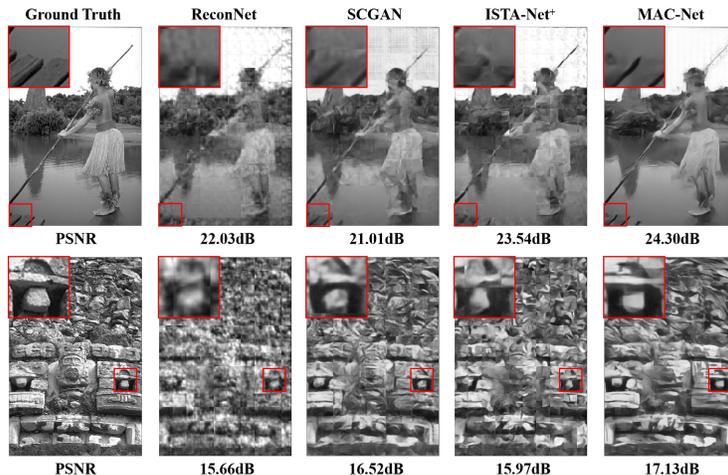


Fig. 7. Visual comparisons between multiple algorithms upon test002 and test067 images from BSD68 at CS measurement rate of 0.04.

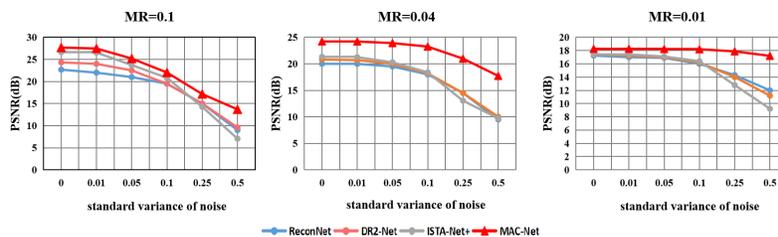


Fig. 8. Comparison of noise robustness on SET11.

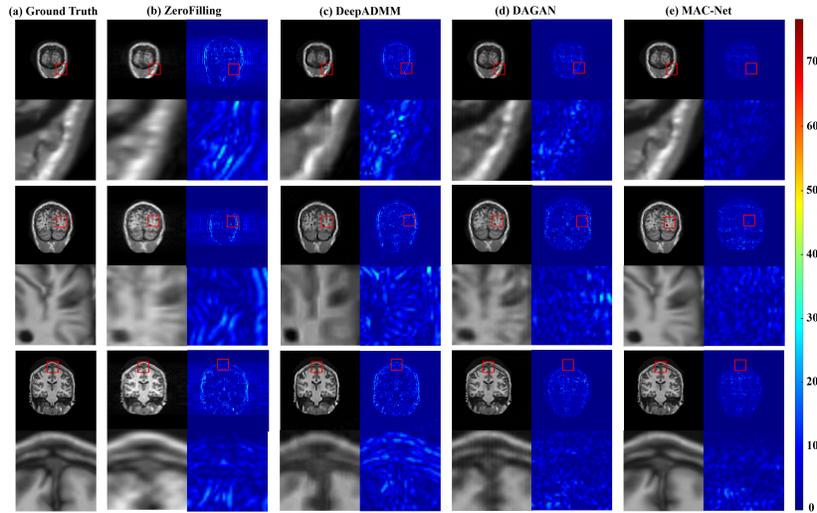
rates of 0.1, 0.2, 0.3, 0.4, and 0.5, MAC-Net is significantly better than the other two methods, even exceeds DAGAN 9.52dB at the measurement rate of 0.4. Fig. 9 shows the visual results, including reconstructed images and residuals between reconstructed images and ground truth images. We also zoom in texture block of every image to better reflect the reconstruction effect of MAC-Net. There are more details reconstructed by MAC-Net and fewer differences between original images and reconstructed images. Apart from this, compared to DeepADMM and DAGAN, less noise is generated during reconstruction. This indicates that through the full use of measurement vector in a cascading manner, MAC-Net can not only reconstruct images accurately but also suppress noise effectively.

## 5 Conclusions

In this paper, we propose a memory augmented cascading network for compressed sensing of images. In order to reconstruct images of better quality, the proposed network employs a cascading framework with multiple stages. Each

**Table 3.** Reconstruction results on MICCAI 2013 grand challenge dataset.

Algorithm	Measurement Rate				
	0.1	0.2	0.3	0.4	0.5
DeepADMM [40]	30.70	39.10	39.72	43.25	44.39
DAGAN [39]	33.79	39.44	40.20	44.83	47.83
MAC-Net	<b>35.94</b>	<b>46.07</b>	<b>46.93</b>	<b>54.35</b>	<b>56.34</b>



**Fig. 9.** Visual results of MAC-Net, DeepADMM and DAGAN upon three sample images from MICCAI 2013 grand challenge dataset at CS measurement rate of 0.3 using 1D Gaussian masks. The residuals between reconstructed images and ground truth images are also presented by heat maps. The two columns of (b)-(e) are reconstructed images and the residuals, respectively. The color bar on the right indicates the intensity value of each pixel in the residual images.

stage has two tasks: reconstructing the image and calculating the residual of the measurement vector of the current reconstructed image. At the same time, through the contextual memory augmentation and incremental learning of residual parts of images, the network can reconstruct high-quality images. The experimental results show that the proposed method is superior to other state-of-the-art methods on compressed sensing of natural images and MR images.

## 6 Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant 61672292, Grant 61825601, Grant 61532009, and funding from Shenzhen Institute of Artificial Intelligence and Robotics for Society. The corresponding author of this paper is Yubao Sun.

## References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). pp. 265–283 (2016)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009)
3. Candès, E.J., Justin, R., Terence, T.: Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* **59**(8), 1207–1223 (2006)
4. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory* **52**(12), 5406–5425 (2006)
5. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. *IEEE Signal Processing Magazine* **25**(2), 21–30 (2008)
6. Chengbo, L., Wotao, Y., Hong, J., Yin, Z.: An efficient augmented lagrangian method with applications to total variation minimization. *Computational Optimization and Applications* **56**(3), 507–530 (2013)
7. David L., D.: For most large underdetermined systems of linear equations the minimal  $l_1$  norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* **59**(7), 907–934 (2006)
8. Dong, W., Shi, G., Li, X., Ma, Y., Huang, F.: Compressive sensing via nonlocal low-rank regularization. *IEEE Transactions on Image Processing* **23**(8), 3618–3632 (2014)
9. Dong, Z., Zhu, W.: Homotopy methods based on  $l_0$ -norm for compressed sensing. *IEEE Transactions on Neural Networks and Learning Systems* **29**(4), 1132–1146 (2017)
10. Guo, C., Yang, Q.: A neurodynamic optimization method for recovery of compressive sensed signals with globally converged solution approximating to  $l_0$  minimization. *IEEE Transactions on Neural Networks and Learning Systems* **26**(7), 1363–1374 (2014)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 (2016)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 7132–7141 (2018)
13. Itzler, M.A., Entwistle, M., Jiang, X., Owens, M., Slomkowski, K., Rangwala, S.: Geiger-mode apd single-photon cameras for 3d laser radar imaging. In: *IEEE Aerospace Conference*. pp. 1–12. IEEE (2014)
14. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: *Advances in neural information processing systems*. pp. 2017–2025 (2015)
15. James Edwin, F., Sungkwang, M., Eric, T.: Block-based compressed sensing of images and video. *Foundations and Trends in Signal Processing* **4**, 297–416 (2012)
16. Jian, Z., Bernard, G.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018)
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)

18. Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., Ashok, A.: Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 449–458 (2016)
19. Liu, Y., Yuan, X., Suo, J., Brady, D., Dai, Q.: Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(12), 2990–3006 (2019)
20. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine* **58**(6), 1182–1195 (2007)
21. Martin, D., Fowlkes, C., Tal, D., Malik, J.: 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 (ICCV). vol. 2, pp. 416–423. IEEE (2001)
22. McMackin, L., Herman, M.A., Weston, T.: Multi-spectral visible-to-shortwave infrared smart camera built on a compressive sensing platform. In: *Frontiers in Optics*. pp. FTh5C–4. Optical Society of America (2016)
23. Metzler, C.A., Maleki, A., Baraniuk, R.G.: From denoising to compressed sensing. *IEEE Transactions on Information Theory* **62**(9), 5117–5144 (2016)
24. Mousavi, A., Baraniuk, R.G.: Learning to invert: Signal recovery via deep convolutional networks. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 2272–2276. IEEE (2017)
25. Na, S., Lee, S., Kim, J., Kim, G.: A read-write memory network for movie story understanding. In: IEEE International Conference on Computer Vision (CVPR). pp. 677–685 (2017)
26. Needell, D., Tropp, J.A.: Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* **26**(3), 301–321 (2009)
27. Needell, D., Vershynin, R.: Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing* **4**(2), 310–316 (2010)
28. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1874–1883 (2016)
29. Sukhbaatar, S., Weston, J., Fergus, R., et al.: End-to-end memory networks. In: International Conference on Neural Information Processing Systems (NeurIPS). pp. 2440–2448 (2015)
30. Sun, Y., Chen, J., Liu, Q., Liu, G.: Learning image compressed sensing with sub-pixel convolutional generative adversarial network. *Pattern Recognition* **98**, 107051 (2020)
31. Sun, Y., Yang, Y., Liu, Q., et al.: Learning non-locally regularized compressed sensing network with half-quadratic splitting. *IEEE Transactions on Multimedia* (online, 2020)
32. Sungkwang, M., James Edwin, F.: Residual reconstruction for block-based compressed sensing of video. In: *Data Compression Conference* (2011)
33. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory* **53**(12), 4655–4666 (2007)
34. Vaswani, N.: Ls-cs-residual (ls-cs): compressive sensing on least squares residual. *IEEE Transactions on Signal Processing* **58**(8), 4108–4120 (2010)

35. Wang, J., Kwon, S., Shim, B.: Generalized orthogonal matching pursuit. *IEEE Transactions on Signal Processing* **60**(12), 6202–6216 (2012)
36. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: Cbam: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 3–19 (2018)
37. Wu, Y., Arce, G.: Snapshot spectral imaging via compressive random convolution. In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 1465–1468. IEEE (2011)
38. Xie, X., Wang, C., Du, J., Shi, G.: Full image recover for block-based compressive sensing. In: *IEEE International Conference on Multimedia and Expo (ICME)*. pp. 1–6. IEEE (2018)
39. Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., et al.: Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE Transactions on Medical Imaging* **37**(6), 1310–1321 (2017)
40. Yang, Y., Sun, J., Li, H., Xu, Z.: Admm-csnet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(3), 521–538 (2020)
41. Yao, H., Dai, F., Zhang, S., Zhang, Y., Tian, Q., Xu, C.: Dr<sup>2</sup>-net: Deep residual reconstruction network for image compressive sensing. *Neurocomputing* **359**, 483–493 (2019)