Image Compressed Sensing and Reconstruction of Multi-Scale Residual Network Combined with Channel Attention Mechanism

Chunyan Zeng¹, Zhenghui Wang¹, Zhifeng Wang ²*, Kang Yan¹ and Yan Yu¹

¹ Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan, 430068, China
² School of Educational Information Technology / Hubei Research Center for Educational Informationization, Central China Normal University, Wuhan, 430079, China

*Corresponding author’s e-mail: zfwang@mail.ccnu.edu.cn.

Abstract. Most of the existing compressed sensing reconstruction algorithms based on deep learning only use simple stacked convolutional layers to extract image feature information, which cannot extract sufficient image information; and each feature maps are treated equally in the reconstruction process, which lacks the ability of discriminative learning across feature channels and can’t make full use of the representation capabilities of convolutional neural network. This paper proposes a new multi-scale image compressed sensing reconstruction network based on residual network and channel attention mechanism. We use the residual network to increase the network depth, which reduces the information loss in the convolutional process through the skip connections in the residual block. It can obtain richer image information and protect the integrity of data information. We also add channel attention mechanism to adaptively scale each feature maps to enhance effective feature information and suppress invalid feature information. The experimental results show that the performance of image reconstruction is greatly ameliorated compared with the existing the state-of-the-art methods.

1. Introduction
Compressed sensing (CS) theory is an important milestone in data sampling and reconstruction. In sparse domain, it can reconstruct original signal with high probability from a lower sampling rate. However, traditional compressed sensing algorithms have high computational complexity and low reconstruction quality, which hinders the practical application of compressed sensing.

Recently, deep learning technology has made great achievements in various visual fields such as face recognition [1], object detection [2], semantic segmentation [3] and image super resolution [4]. Therefore, Mousavi et al. [5] applies deep learning technology to CS for image reconstruction. It employs the SDA model to realize encoding and decoding by fully-connected, which improves reconstruction performance. Because convolutional neural networks (CNN) can effectively process high-dimensional data compared with fully-connected, CNN is used to reconstruct the image from the measurement in [6] and [7]. In order to extract richer image information to improve the quality of image reconstruction. In [8], Yao et al. use the residual network to replace the general convolutional layer in the reconstruction network to increase the network depth, which further improve the reconstruction
performance. Due to deep learning lacks understanding of their internal functioning, Yang et al. propose a new compressed sensing reconstruction architecture naming ADMM-CSNet [9] which combines traditional CS method and deep learning method to reconstruct image from sparsely measurements. In [10], Zhang et al. are inspired by the iterative shrinkage-thresholding algorithm (ISTA) to propose ISTA-Net which casts ISTA into deep network form for image CS reconstruction.

Most of the above methods use simple stacked convolutional layers to extract the feature information of the image, which it only can obtain single-scale spatial information. In convolutional process, each feature maps are treated equally, which lacks the ability of discriminative learning across feature channels and cannot make full use of the representation capabilities of CNN.

This paper proposes a new multi-scale image compressed sensing reconstruction network combining residual network and attention mechanism. In deep reconstruction part, we apply convolution kernels of different scales in the residual block to obtain the deep multi-scale feature information of the image, which can obtain richer image information and protect the integrity of data information. In the network, we also add channel attention mechanism to adaptively scale each feature maps to enhance effective feature information and suppress invalid feature information.

2. Related work

2.1. Residual network

Recently, researchers found that the performance of the network can be improved when appropriately increasing the number of network layers. But when the number of network layers increases to a certain level, the network will be degraded and performance of model deteriorates. In 2015, He et al. [11] proposed a new network named ResNet to solve this problem well. The residual network is actually composed of multiple residual blocks. The residual block is shown in Figure 1(a), which has a skip connection compared the previous convolutional operation. If the output of previous two convolutional layers is \( H(x) \), Then the output of the residual block can be expressed as:

\[
H(x) = F(x) + x
\]

2.2. Channel attention mechanism

Previous CNN-based compressed sensing reconstruction algorithms treat each channel feature equally. Although it effectively improves the reconstruction performance, it does not make full use of the representation ability of CNN. Hu et al. [12] propose Squeeze-and-Excitation block that can effectively solve this problem. Its structure is shown in Figure 2.
First, the input feature is $X$, the output feature is $U$. The convolutional operation process is as follows:

$$F_c: X \rightarrow U, X \in H \times W \times C, U \in H \times W \times C$$ (2)

Therefore, the output of $U$ can be expressed as $U = [u_1, u_2, \cdots, u_c]$. Next, the compression operation is performed. Then we perform a global average pooling operation. Each two-dimensional feature channel is transformed into a real number, and the compressed one-dimensional feature represents the global distribution on the feature channel. The expression is as follows:

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j)$$ (3)

Where $W$ and $H$ are respectively the width and height of the feature maps. Then, the excitation operation is used to obtain the dependency from the compressed one-dimensional features. The expression is as follows:

$$F_{ex}(u_c, z) = f(W_d \delta(W_z z))$$ (4)

Where $f(\cdot)$ represents the sigmoid function, and $\delta(\cdot)$ represents the ReLU activation function. In order to improve the generalization of the model, down-sampling and up-sampling operations are respectively performed for one-dimensional features, $W_1$ and $W_2$ represent the weights of them. The $s$ represents obtained one-dimensional features which will be used as the weight of feature channel fusion.

Finally, Output features can be adaptively scaled by the channel weight. The expression is as follows:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c$$ (5)

Where $\tilde{X}_c = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_c]$ and $F_{scale}(u_c, s_c)$ represent the product between the channels feature maps $u_c \in \mathbb{R}^{W \times H}$ and the scalar $s_c$. The channel attention mechanism can suppress weakly related information and highlight strongly related feature information.

3. Proposed method

This paper proposes a new network framework for CS measurement and reconstruction. The network has three parts including image measuring, initial reconstruction and depth reconstruction shown in Figure 3.

3.1. Network architecture

In the traditional compressed sensing measurement methods, the measurements can be obtained by $y = \Phi x$. The $\Phi$ represents random measurement matrix. However, this paper applies the convolutional network to obtain the measurements. The measurement matrix can be learned automatically in network. In order to reconstruct the original image from the measurements. First, the image data needs to be up-sampling to improve the dimension of the image data. The deconvolution is used to obtain the initial reconstructed image. Second, to get high-quality reconstructed image, we need to perform deep reconstruction. In the deep reconstruction network, we obtain multi-scale deep feature information through three parallel convolution channels. The residual block in each parallel convolution channel is
composed of general convolutional kernel and dilated convolutional kernel with the same receptive field, and we can obtain richer image information through multiple stacked residual blocks.

We introduce the channel attention mechanism before the fusion of different scale feature information. We apply the channel attention mechanism module to learn a new weight for each feature maps through network training, then the weight is multiplied by each feature maps to adaptively scale each feature maps to enhance effective feature information and suppress invalid feature information. Then, the newly obtained feature information of different scales is fused and input into the two final convolutional layers to obtain the final reconstructed image.

3.2. Network training
We use mean square error (MSE) as loss function liking the previous reconstruction methods. The mean square error loss function is represented as

$$\min \frac{1}{2N} \sum_{i=1}^{N} \| f(x_i, \theta) - x_i \|^2$$

(6)

Where $N$ represents the total number of samples, $\theta$ is the parameter we need to learn through network training, $f(x_i, \theta)$ is Reconstructed image, and $x_i$ is the label. Finally, we train all network parameters by the Adam optimizer.

4. Experimental results and analysis

4.1. Experimental setting
In the measurement period, the convolutional stride and the size of convolution kernel are both set to 32, The stride and kernel size of the initial reconstruction layer are also set to 32. Each residual block is composed of a general convolutional kernel and dilated convolutional kernel with the same receptive field, and the dilated factors of the three parallel convolution channels are 1, 2, and 3 respectively. The size of the convolutional kernel of the last two reconstruction layers is $3 \times 3$. We use ReLU as the activation function excepting for the measuring layer and last layer.

The experiment uses 400 images from the BSDS500 to train network completed in Tensorflow. The epoch of maximum number is set to 100. The learning rates of the first 50 epochs, the 51 to 80 epochs, and the other 20 epochs are 0.001, 0.0001, and 0.00001, respectively.

4.2. Comparisons with state-of-the-art methods
We compare our method with three methods based on deep learning: ReconNet [6], DR2-Net [8] and ISTA-Net [10]. We use peak signal to noise ratio (PSNR) as evaluation indicator and complete the test on 11 test images from set11. The experimental results are shown in Table 1. It can be seen from Table 1 that the average PSNR of our is higher than that of previous methods.

Finally, we compare reconstructed visual effects of each algorithm respectively at three different measurement rates. The reconstructed images are shown in Figure 4. It can be seen from Figure 4 that the visual effect of our proposed is significantly better than that of other methods.
| Images   | Methods    | Measurement rate | Images   | Methods    | Measurement rate |
|----------|------------|------------------|----------|------------|------------------|
|          |            | 0.10             |          |            | 0.10             |
|          |            | 0.04             |          |            | 0.04             |
|          |            | 0.01             |          |            | 0.01             |
| Barbara  | ReconNet   | 21.89 20.38 18.61| Boats    | ReconNet   | 24.15 21.36 18.49|
|          | DR2-Net    | 22.69 20.70 18.65|          | DR2-Net    | 25.58 22.11 18.67|
|          | ISTA-Net   | 23.51 20.99 18.38|          | ISTA-Net   | 27.27 22.14 18.47|
|          | Our        | **24.73 23.92 21.93**|          | Our        | **29.32 26.41 22.12**|
| Fingerprints | ReconNet  | 20.75 16.91 14.82| Parrot   | ReconNet   | 22.63 20.27 17.63|
|          | DR2-Net    | 22.03 17.40 14.73|          | DR2-Net    | 23.94 21.16 18.01|
|          | ISTA-Net   | 22.45 17.31 14.78|          | ISTA-Net   | 26.21 22.24 17.90|
|          | Our        | **26.91 21.39 16.51**|          | Our        | **27.25 25.37 22.73**|
| Lena     | ReconNet   | 23.83 21.28 17.87| Foreman  | ReconNet   | 27.09 23.72 20.04|
|          | DR2-Net    | 25.39 22.13 17.97|          | DR2-Net    | 29.20 25.34 20.59|
|          | ISTA-Net   | 27.44 22.40 18.29|          | ISTA-Net   | 32.78 25.76 20.21|
|          | Our        | **28.94 26.33 22.72**|          | Our        | **34.45 31.62 26.97**|
| Monarch  | ReconNet   | 21.10 18.19 15.39| House    | ReconNet   | 26.69 22.58 19.31|
|          | DR2-Net    | 23.10 18.93 15.33|          | DR2-Net    | 27.53 23.92 19.61|
|          | ISTA-Net   | 25.58 19.40 14.99|          | ISTA-Net   | 30.13 24.03 19.80|
|          | Our        | **28.01 24.12 18.31**|          | Our        | **32.49 29.25 24.47**|
| Flinstones| ReconNet  | 18.92 16.30 13.96| Peppers  | ReconNet   | 22.15 19.56 16.82|
|          | DR2-Net    | 21.09 16.93 14.01|          | DR2-Net    | 23.73 20.32 16.90|
|          | ISTA-Net   | 23.39 17.43 14.00|          | ISTA-Net   | 26.66 20.64 16.88|
|          | Our        | **23.52 20.41 16.82**|          | Our        | **27.01 24.62 20.92**|
| Cameraman| ReconNet   | 21.28 19.26 17.11| Mean     | ReconNet   | 22.68 19.99 17.27|
|          | DR2-Net    | 22.46 19.84 17.08|          | DR2-Net    | 24.32 20.80 17.44|
|          | ISTA-Net   | 23.46 20.27 17.26|          | ISTA-Net   | 26.26 21.15 17.36|
|          | Our        | **25.51 22.98 20.45**|          | Our        | **28.01 25.13 21.27**|

Figure 4  From left to right is the original image and the reconstructed image of the ReconNet, DR²-Net, ISTA-Net, and Our algorithm at three different measurement rates
5. Conclusion
This paper proposes a multi-scale residual network with channel attention mechanism for CS image measurement and reconstruction. We use residual to extract image deep feature information. In multi-scale residual network, the channel attention mechanism module is also added to learn a new weight for each feature maps through network training, in which network can focus on more useful feature information to improve image reconstruction quality. The experimental results show that the residual and attention mechanism can enhance the representation ability of the convolutional network and greatly improve the image reconstruction performance.

Acknowledgments
This research is supported by National Natural Science Foundation of China (No.61901165, 61501199), Science and Technology Research Project of Hubei Education Department (No. Q20191406), Hubei Natural Science Foundation (No. 2017CFB683), and self-determined research funds of CCNU from the colleges basic research and operation of MOE (No. CCNU20ZT010).

References
[1] He, R., Cao, J., Song, L., Sun, Z. and Tan, T. (2020) Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence., 42: 1025-1037.
[2] Deng, S. et al. (2021) A Global-Local Self-Adaptive Network for Drone-View Object Detection. IEEE Transactions on Image Processing., 30: 1556-1569.
[3] Han, H.Y., Chen, Y.C., Hsiao, P. Y. and Fu, L.C. (2021) Using Channel-Wise Attention for Deep CNN Based Real-Time Semantic Segmentation With Class-Aware Edge Information. IEEE Transactions on Intelligent Transportation Systems., 22: 1041-1051.
[4] Wei, W., Nie, J., Li, Y. (2020) Deep Recursive Network for Hyperspectral Image Super-Resolution. IEEE Transactions on Computational Imaging., 6:1233-1244.
[5] Mousavi, A., Patel, A.B. and Baraniuk, R.G. (2015) A Deep Learning Approach to Structured Signal Recovery. 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 1336-1343.
[6] Kulkarni, K., Lohit, S., Turaga, P. (2016) ReconNet: Non-Iterative Reconstruction of Images from Compressively Sensed Measurements. Computer Vision and Pattern Recognition. pp. 449-458.
[7] Mousavi, A., Baraniuk, R.G. (2017) Learning to invert: Signal recovery via Deep Convolutional Networks. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 2272-2276.
[8] Yao, H., Dai, F., Zhang, S. (2019) DR2-Net: Deep Residual Reconstruction Network for Image Compressive Sensing. Neucomp. pp. 483–493.
[9] Yang, Y., Sun, J., Li, H. (2020) ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence., 42:521-538.
[10] Zhang, J. (2018) ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. Conference on Computer Vision and Pattern Recognition. pp. 1828-1837.
[11] He, K., Zhang, X., Ren, S. (2016) Deep Residual Learning for Image Recognition. Conference on Computer Vision and Pattern Recognition. pp. 770-778.
[12] Hu, J., Shen, L., Sun, G. (2018) Squeeze-and-Excitation Networks. Conference on Computer Vision and Pattern Recognition. pp. 7132-7141.