# Compressed Sensing Network based on Wavelet Transform

Zhu Yin [1, 2 +], ZhongCheng Wu [1, 2], Jun Zhang [1] and Fang Li[1]

[1] High Magnetic Field Laboratory, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

[2] University of Science and Technology of China, Hefei 230026, China

**Abstract.** The traditional compressed sensing algorithm realizes the optimization of image reconstruction through multiple iterative calculation from limited measurements, which cost high computational complexity and long reconstruction time. As the development of deep learning, it is proposed to combine the technology with compressed sensing(CS) which shows great advantages in accurate and fast CS reconstruction. In this paper, we propose a novel algorithm synthesize the advantages of the two technology as well as add another sparse prior technique based sym8 wavelet, which dubbed WCS-Net, is focus on two parts: sampling network based on sparse representation and deep reconstruction elastic network. Experimental results show the WCS-Net has the advanced performance at measurement rates 0.01, 0.04, 0.1, and 0.25, respectively, while maintaining the same running speed as existing image compression methods based on deep learning.

**Keywords:** compressed sensing; sampling network; sparse representation; deep learning.

## 1. Introduction

The conventional information transmission follows the Nyquist sampling, which sampling frequency should be no less than twice the highest frequency of the signal [1]. In 2006, Compressed Sensing (CS) theory [2] proposed by Candes, et al, which is based on sparse representation, complete the compression of information at the same time of sampling that breaks the limited bandwidth of traditional information transmission. CS has been applied in many practical applications, such as single-pixel cameras, accelerating magnetic resonance imaging (MRI), and wireless broadcast.

Currently, the CS methods are divided into two categories, namely traditional CS algorithm and CS algorithm based on deep learning technology. The traditional CS algorithm include convex optimization algorithm [5], greedy matching pursuit algorithm [4, 6], Bayesian algorithm [7] and the gradient descent methods [8]. These iterative calculations based on traditional methods have high complexity, time-consuming, and poor real-time performance. CS technology is based on sparse prior, and sparsity has been proven to be effective in many fine-structure image reconstructions, such as discrete wavelet transform(DWT) [9] that used as prior knowledge of the original image signals, which are beneficial to compressive sensing image reconstruction. The multi-hypothesis (MH) [13] prediction is applied to CS reconstruction of both still images and video sequences.

The CS algorithm based on deep learning is three orders of magnitude faster than the traditional CS algorithm and is an end-to-end difficult to interpret black box network. Typical deep learning-based CS methods include denoising encoder, convolutional neural network, residual network and generated model network. The precision of image reconstruction can be improved through amount of training and learning. In [16], Mousavi et al. proposed a stacked denoising autoencoder (SDA) to learn the signal mapping between CS certain measurements and improve signal reconstruction performance. In [15], Kulkarni et al. used a CNN (ReconNet) for image block reconstruction and an off-the-shelf denoiser for deblocking. In [18], the DR$^2$-Net proposed by Yao et al., which used the residual network to enhance the reconstruction quality of image. All of these block-by-block reconstruction methods [16,15,18] adopt random Gaussian matrix for measurements sampling and do not combine with image prior information to further enhance the accuracy of sampling

measurements. At the same time, there are block artifacts caused in these methods, which affect the quality of reconstruction.

In this paper, we focus on how to achieve more efficient and high quality reconstruction and recovery with less and more accurate measurements data. The model will have better feature representation and generalization performance around these representational purposes. Therefore, the article is roughly divided into the following points:

(i) The sparsity is proved to be effective in mage reconstruction [17]. In terms of sparse processing, wavelet sparse processing has a good effect on denoising. Therefore, this paper adopts the one-layer wavelet transform to carry out the image sparsity processing.

(ii) The sampling network is designed adaptively according to the image sparsity, and the sampling values with more image characteristics are extracted to further improve the reconstruction quality.

(iii) The reconstruction network adaptively combines the sampling values obtained by the sampling network and optimizes the overall reconstruction quality through training. The reconstruction network is divided into two subnetworks: the initial reconstruction network and the deep reconstruction network. The initial reconstruction network reconstructs the whole image, and the deep reconstruction network further improves the effect and precision of the final image reconstruction.

## 2. Algorithm Overview

As shown in Fig. 1, WCS-Net is takes the Compressively Sensed (CS) measurements of the wavelet transformed image as input, then use adaptive sampling network [19] to obtain more accurate measurement data, further more we conduct refactoring network to automatically realize high-quality reconstruction. The refactoring network contains two networks: initial reconstruction network and deep reconstruction network. Among them, the deep reconstruction network adopts residual network [11], which is the further improvement of $DR^2$-Net. WCS-Net combines the sparse priori with deep learning which can celebrate the network convergence speed, boost the network performance and has good performance with low computational complexity.
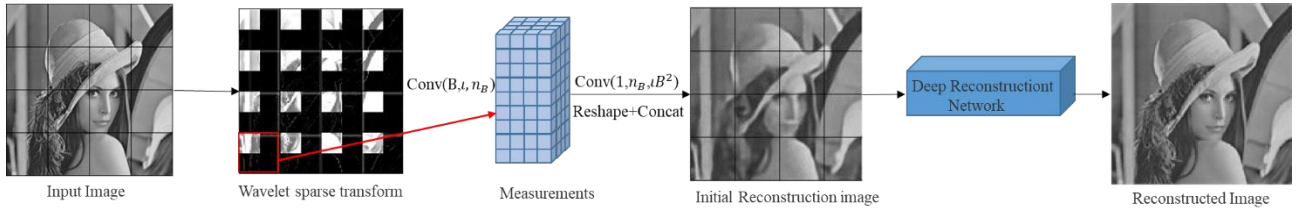


Fig. 1: The network structure schematic of WCS-Net.

### 2.1. Sparse Wavelet Transform Based on CS

CS technology aims recovery the signal $x \in R^{n \times 1}$ from its measurements $y \in R^{m \times 1}$. This process can be expressed as:

$$y = \Phi x \tag{1}$$

Here, $\Phi \in R^{m \times n}$ is a linear random projection (matrix), this process is a typical ill-posed problem, m/n is defined as CS ratio(m<<n).

The structured sparse signal reconstruction algorithm [20] by introducing the structure of the model to choose presupposition to constrain the feasible solution space, reduce the understanding of spatial dimension, and the necessary number of sampling measurement, which not only get fast speed, but also obtain better refactor. Based on the different edge information of the image, we first make the image x sparsely divided into blocks using the Sym8 wavelet basis ($\Psi = [\psi_1, \psi_2, \cdots, \psi_N]$) as shown below:

$$x = \Psi s \tag{2}$$

$s \in R^{n \times 1}$ is the signal form of x after the wavelet sparse transformation, when $\| s \|_0 = K$, that is, signal x can be considered to be K sparse under the wavelet basis.

Wavelet transform makes the image from two-dimensional spatial domain to high and low frequency wavelet domain, which effectively improves the image processing effect of CS technology.
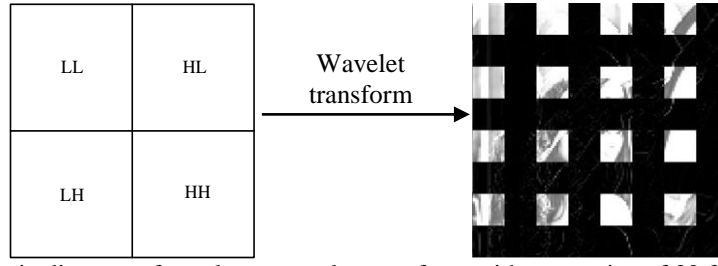
Fig. 2: A schematic diagram of one-layer wavelet transform with a step size of 32 for the image.

From Fig. 2 the left picture, the image is divided into four sub-band modules: LL, HL, LH and HH respectively represent the approximate components, horizontal details, vertical details and diagonal details of the image. Among them, HL has column sparsity in the horizontal direction, LH has row sparsity in the vertical direction, and HH has sparsity in the diagonal direction. In our method, we carry out a layer of sym8 wavelet transform to the image with the step size 32x32, and the transformed image is the right picture shown in Fig. 2 the right picture.

Combined with equation (2), equation (1) can be expressed as:

$$y = \Phi x = \Phi \Psi s \tag{3}$$

Through the sparse image processing, it is better for sampling to obtain more accurate measurements data, so as to achieve better reconstruction.

## 2.2. Sampling Network

The sampling network designed in our method, which uses BCS measurement projection divides the sparsely processed images into non-overlapping blocks of size $B \times B \times l$, where B and $l$ represent the spacial size and the number of channel, as well as the value are set as 32 and 1, respectively. We denote the $F^S(\cdot)$ a convolution layer as the sampling network, which have the $n_B$ filters of size $B \times B \times l$, no bias and the stride of this convolution layer is $B \times B$. The whole procession devoted as $\text{Conv}(B, l, n_B)$ in Fig 1 would be shown as the following equation:

$$y = F^S(s, W^S) = W^S * s = \quad n_B = \frac{m}{n} l B^2 \tag{4}$$

which the weight $W^S$ indicated as $n_B$ filters of size $B \times B \times l$. By training the whole algorithm framework with Adam [12] method, the optimal sampling matrix can be obtained adaptively.

## 2.3. Refactoring Network

1) Initial Reconstruction Network

By the sampling network, the optimally measurements y would be obtained, then the convolution layer and concatenation layer are used to construct y to the initial reconstruction $\tilde{x}$, that is expressed as the following equation:

$$\tilde{x} = F^I(y, W^I) = F^I(F^S(s, W^S), W^I)$$

$$= C \begin{pmatrix} \gamma(F_{11}^I(y, W^I)) & \cdots & \gamma(F_{1w}^I(y, W^I)) \\ \vdots & \ddots & \vdots \\ \gamma(F_{h1}^I(y, W^I)) & \cdots & \gamma(F_{hw}^I(y, W^I)) \end{pmatrix}$$

$$= C \begin{pmatrix} \gamma(F_{11}^I(F^S(s, W^S), W^I)) & \cdots & \gamma(F_{1w}^I(F^S(s, W^S), W^I)) \\ \vdots & \ddots & \vdots \\ \gamma(F_{h1}^I(F^S(s, W^S), W^I)) & \cdots & \gamma(F_{hw}^I(F^S(s, W^S), W^I)) \end{pmatrix} \tag{5}$$

Which $W^I$ corresponding to the $1 \times 32 \times 32$ filters, $F_{hw}^I(y, W^I)$ represent $1 \times 1 \times 32 \times 32$ vector, h and w respectively indicate the number of blocks in rows and columns. The symbol, $\gamma(\cdot)$, means reshape function that converts the 1024 vector is into $32 \times 32$ a two-dimensional image blocks. $C(\cdot)$ is the function that concatenates all blocks to form the whole image.

2) Deep Reconstruction Network:

The deep reconstruction network $F^D(\tilde{x}, W^D)$ is designed similar to the DR²-Net algorithm, reconstructs the image through the sampling network and the initial reconstruction network to obtain the approximate original image, then obtains high quality reconstructed images by learns the residuals of the image from the residual network that consist of four residual submodules and each submodules comprise three convolutional layers in Fig. 3. The four residual submodules are represented by green, purple, blue and magenta respectively.
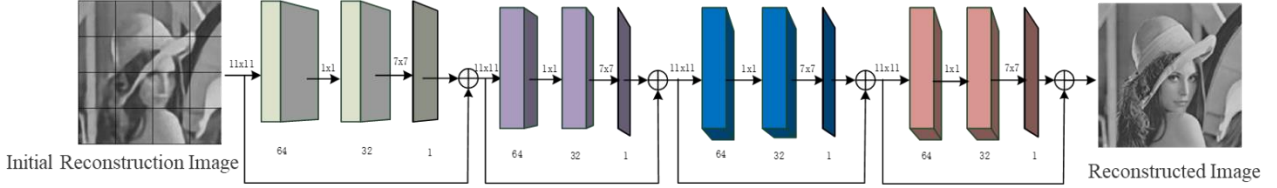


Fig. 3: The framework of Deep Reconstruction Network.

The first layer is green residual submodule which takes initial reconstruction as input and generates 64 feature maps with kernel of size 11 ×11, the second convolution layer generates 32 feature maps with 1×1 kernel, the third convolution layer generates one feature map (residual image) with 7×7 kernel, then fuses it with the initial reconstructed image and used as the input of the next residual module. Then other modules with the same residual module are processed successively, and the final result is the reconstructed image. In the convolution process, the step size is 1 and add corresponding padding on each layer to keep the original image size, the final reconstructed image is $\tilde{x}'$.

By training on the sampling and refactoring network simultaneously, we could get $W^I, W^S, W^D$ by solving the following equation, i.e. ,

$$W^I, W^S, W^D = arg \min_{w^i,w^s,w^d} \| x - \tilde{x}' \|_2^2 = arg \min_{w^i,w^s} \| x - F^D(\tilde{x}, W^D) \|_2^2$$
$$= arg \min_{w^i,w^s,w^d} \| x - F^D(F^I(y, W^I), W^D) \|_2^2$$
$$= arg \min_{w^i,w^s,w^d} \| x - F^D(F^I(F^S(s, W^S), W^I), W^D) \|_2^2 \qquad (6)$$

Different from the traditional block-based compressed sensing algorithm, our method avoids the block effect through the way of convolution, so there is no need to add BM3D to remove the block artifacts, which greatly improves the efficiency and is better than the DR²-Net method in effect.

## 3. Experiment

### 3.1. Training Data Generation and Implementation Details

We choose 400 images of the BSDS500 [3] database as the training dataset which consists of the training set (200 images) and the test set (200 images). Firstly, we cropped the training dataset that 400 images into the size of 96×96 pixel sub-images. Then these sub-images are subjected to the following data enhancement operations, such as rotation and flipped of the corresponding transformed images. Finally, the 89600 patches are chosen to optimize the network parameters. For the whole network, the number of training epochs is 100, which the learning rate of the first 50 epochs, the middle 30 epochs and the last 20 epochs are set as $10^{-3}$, $10^{-4}$ and $10^{-5}$, respectively. In the specific training procession, the colour space of image from RGB to YCbcr, and only training the luminance component Y.

### 3.2. Comparison with the State-of-the-art

1) Comparison with traditional methods

Table 1: Average PSNR and SSIM comparisons of different image CS algorithms on Set5 [19]

| Data | | DWT | | TV | | MH | | GSR | | WCS-Net | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ratio | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Set5 | 0.01 | 9.27 | 0.140 | 15.53 | 0.455 | 18.08 | 0.447 | 18.87 | 0.4909 | **23.96** | **0.643** |
| | 0.05 | 14.27 | 0.355 | 23.16 | 0.667 | 23.67 | 0.656 | 24.95 | 0.7270 | **29.35** | **0.838** |
| | 0.1 | 24.74 | 0.768 | 27.07 | 0.786 | 28.57 | 0.821 | 29.99 | 0.8654 | **32.30** | **0.902** |
| | 0.2 | 30.83 | 0.874 | 30.45 | 0.870 | 32.08 | 0.888 | 34.17 | 0.9257 | **35.66** | **0.944** |
| | 0.3 | 33.61 | 0.9.50 | 32.75 | 0.910 | 34.06 | 0.915 | 36.83 | 0.9492 | **37.89** | **0.962** |

As shown in Table 1, four typical and advanced traditional compressed sensing methods were selected for comparison which are wavelet method (DWT) [9], total variation (TV) method [10], group sparse representation (GSR) method [14] and multi-hypothesis (MH) method [13], all of them are the BCS methods. We selected Set5, three datasets which are widely used in the field of compressed sensing, and evaluated each algorithm by the two indexes of signal-to-noise ratio and structural similarity. We select 5 sampling ratio, i.e. 0.01,0.05,0.1,0.2,0.3 to have quantitative and qualitative comparison. All the compared methods would be downloaded from the authors' websites, and the experiments are accomplished in Matlab2018b on Windows 10 system, and runs on desktop computer with Intel Core i7-3770 CPU + NVIDIA GTX1080 GPU, except the authors has already provided the results of compared methods.

2) Comparison with deep learning methods

The compared deep learning based CS methods include: SDA, ReconNet and $DR^2$-Net.The average PSNR comparison on Set11 [15] of four sampling ratios are shown in Table 2 are taken from [18]. As shown in table 2, WCS-Net outperforms the other compared deep learning methods significantly at any ratio.

Table 2: Average PSNR comparison of different deep learning based image CS methods on Set11

| Data | Alg. | Ratio | | | |
|------|------|-------|-------|-------|-------|
|      |      | 0.01  | 0.04  | 0.1   | 0.25  |
| **Set11** | SDA | 17.29 | 19.96 | 22.43 | 24.72 |
|      | ReconNet | 17.27 | 19.99 | 22.68 | 25.54 |
|      | $DR^2$-Net | 17.44 | 20.80 | 24.32 | 28.66 |
|      | WCS-Net | **20.73** | **24.61** | **28.09** | **32.63** |

Fig.4 shows a visual quality comparison of image CS recovery in the case of sampling ratio of 0.01, 0.04，0.1 and 0.25. Obviously, the WCS-Net achieves better visual quality than the $DR^2$-Net methods.

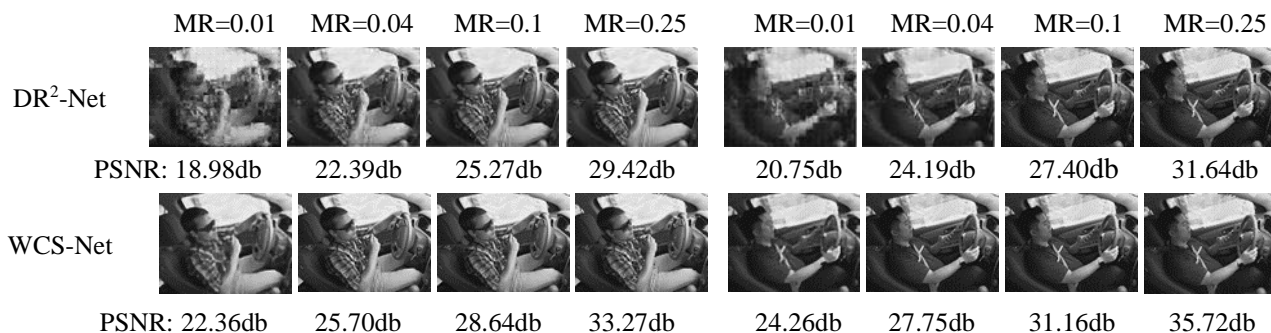| | MR=0.01 | MR=0.04 | MR=0.1 | MR=0.25 | MR=0.01 | MR=0.04 | MR=0.1 | MR=0.25 |
|---|---|---|---|---|---|---|---|---|
| $DR^2$-Net | | | | | | | | |
| PSNR: | 18.98db | 22.39db | 25.27db | 29.42db | 20.75db | 24.19db | 27.40db | 31.64db |
| WCS-Net | | | | | | | | |
| PSNR: | 22.36db | 25.70db | 28.64db | 33.27db | 24.26db | 27.75db | 31.16db | 35.72db |

Fig. 4: Visual quality comparison of image CS recovery on image from State farm [21] in different sampling ratio.

# 4. Conclusion

In this paper, we proposed a network model which is the improvement based on $DR^2$-Net algorithm, synthesize the advantages of the deep learning and the sparse prior, uses adaptive sampling network and refactoring network. Extensive experiments show that WCS-Net greatly improve upon the results of state-of-the-arts CS methods, while maintaining a fast runtime.

# 5. Acknowledgements

# 6. References

[1] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006)1289–1306.

[2] E. Candes, M. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.

[3] 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.

[4] S.G. Mallat , Z. Zhang , Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. SIAM Review, 3(1):129–159, 2001.

[6] J.A . Tropp , A .C. Gilbert , Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.

[7] Ji S, Xue Y, Carin L. Bayesian compressive sensing[J].IEEE Transactions on Signal Processing,2008,56（6）2346-2356.

[8] Mário A T 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.

[9] 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.

[10] 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/.

[11] S. Mun, J.E. Fowler, Residual reconstruction for block-based compressed sensing of video, in: Proceedings Data Comp. Conference (DCC), 2011, pp. 183–192.

[12] Kingma D , Ba J . Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.

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

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

[15] 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.

[16] 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.

[17] Y.C. Eldar, P. Kuppinger, H. Bolcskei, Block-sparse signals: Uncertainty relations and efficient recovery, IEEE Trans. Signal Process. 58 (6) (2010) 3042–3054.

[18] Hantao Yao, Feng Dai, Shiliang Zhang, Yongdong Zhang, Qi Tian, Changsheng Xu, DR$^2$-Net: Deep Residual Reconstruction Network for image compressive sensing, Neurocomputing[J], Pages 483-493,2019.d

[19] 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.

[20] Xiaoyu F, Qiusheng L, Baoshun S. Compressed sensing MRI based on image decomposition model and group sparsity[J]. Magnetic Resonance Imaging, 2019.

[21] State farm distracted driver detection. https://www.kaggle.com/c/state-farm-distracted-driver-detection.