J. Meas. Sci. Instrum., 2025, 16(4): 537-546.
DOI: 10.62756/jmsi.1674-8042.2025052
http://jmsi.publish.founderss.cn/
jmsi@nuc.edu.cn

Image compressed sensing reconstruction network based on self-attention mechanism

LIU Yuhong*, LIU Xiaoyan, CHEN Manyin

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

***Corresponding author:** LIU Yuhong (liuyuhong@mail.lzjtu.cn)

Received: March 15, 2022 **Revised:** May 21, 2022 **Accepted:** June 10, 2022

Abstract: For image compression sensing reconstruction, most algorithms use the method of reconstructing image blocks one by one and stacking many convolutional layers, which usually have defects of obvious block effects, high computational complexity, and long reconstruction time. An image compressed sensing reconstruction network based on self-attention mechanism (SAMNet) was proposed. For the compressed sampling, self-attention convolution was designed, which was conducive to capturing richer features, so that the compressed sensing measurement value retained more image structure information. For the reconstruction, a self-attention mechanism was introduced in the convolutional neural network. A reconstruction network including residual blocks, bottleneck transformer (BoTNet), and dense blocks was proposed, which strengthened the transfer of image features and reduced the amount of parameters dramatically. Under the Set5 dataset, when the measurement rates are 0.01, 0.04, 0.10, and 0.25, the average peak signal-to-noise ratio (PSNR) of SAMNet is improved by 1.27, 1.23, 0.50, and 0.15 dB, respectively, compared to the CSNet+. The running time of reconstructing a 256×256 image is reduced by 0.147 3, 0.178 9, 0.231 0, and 0.252 4 s compared to ReconNet. Experimental results showed that SAMNet improved the quality of reconstructed images and reduced the reconstruction time.

Key words: convolutional neural network; compressed sensing; self-attention mechanism; dense block; image reconstruction

0 Introduction

Under the traditional digital signal processing framework, if the analog signal is restored undisputed from the sampled discrete signal, the sampling frequency of the data is at least twice the signal bandwidth[1]. However, with the increasing demand for information, the signal bandwidth is becoming wider and wider. In the process of information acquisition, higher and higher requirements for measurement rate and processing speed are put forward. The theory of compressive sensing (CS) proposed by Donoho et al.[2] makes it possible that the measurement rate is less than twice the signal bandwidth, and the theory of CS can complete the signal compression while realizing the signal sampling process, combining the sampling and compression process, greatly saving the storage space of the collected signal. In the research of CS, this theory has brought a revolutionary breakthrough in data acquisition technology, and has been widely concerned. Compression sensing has been widely studied and applied in the fields of signal processing, communication, automatic control, and artificial intelligence[3,4].

Traditional iterative compressive sensing reconstruction algorithms include basis pursuit (BP)[5], matching pursuit (MP)[6], and bayesian compressive sensing (BCS)[7]. However, these iterative algorithms have disadvantages of high computational complexity and poor reconstruction quality. Deep learning has received extensive attention in the field of computer vision, such as image super-resolution reconstruction, image semantic segmentation, image denoising. Deep learning has been applied to image compressed sensing[8-11]. The sampling part uses Gaussian random measurement matrix, adaptive measurement matrix, fully connected layer and other methods to obtain compressed sensing measurement value. In the reconstruction part, DR^2-Net blocks the input image, and performs compression sensing measurement on each block image, and then sends the measured value into the reconstruction network to reconstruct these image blocks, and finally splices each image block to generate the final reconstructed image. It can be seen that the block-by-block reconstruction method[12,13] only uses the information in the block to reconstruct a block, which seriously damages the integrity of the image structure information, thus causing block effect.

As mentioned above, in the research of compressed sensing image reconstruction, there are obvious block effects, low quality of reconstructed image, long reconstruction time, and high computational complexity. We proposed an image compression sensing reconstruction network framework based on the self-attention mechanism. Through the improvement and optimization of the sampling part and the reconstruction part, it solved the problems of low accuracy of the reconstructed image, serious block effect, and long reconstruction process in this field, and made the CS measurement retain more image structure information to obtain better reconstruction quality.

The contributions of this paper are mainly reflected in the following aspects.

1) The self-attention mechanism is introduced in the compressed sensing image reconstruction. A structure combining convolution and self-attention mechanism is designed in the sampling network, which is called self-attention convolution (SA Conv). First, the compressed sensing measurement vector of the segmented image is obtained through convolution sampling, and then the self-attention mechanism. Because the output of the self-attention mechanism integrates each input feature, SA Conv can capture more features, make the compressed sensing measurement value retain more image structure information, and overcome the problem of insufficient image feature extraction. Secondly, in the sampling part, the measured vector are obtained by block measurement, and in the reconstruction part, the whole image is reconstructed by using all the measured values of an image. This method effectively reduces the block effect.

2) For the reconstruction network, a reconstruction module based on residual block, dense block, and bottleneck transformer is proposed. Since the dense block realizes the reuse of features in the channel dimension by establishing dense connections of all convolutional layers, so the transfer of image features is strengthened. The amount of parameters, the computational complexity and the reconstruction time are reduced.

3) The convolution-based architecture needs to stack many convolution layers to extract the global information in the image. The self-attention mechanism is to model the whole image and has the ability of fast parallel processing. Therefore, the bottleneck transformer can solve the problems of low learning efficiency and long reconstruction time caused by too many convolution layers.

1 Related work

1.1 Compressive sensing reconstruction network based on deep learning

In 2015, Mousavi et al.[14] first proposed a compressive sensing reconstruction algorithm based on stacking denoising automatic encoder (SDA). SDA is composed of a series of denoising automatic encoders stacked. The intermediate layer of a denoising auto-encoder is used as the input of the next layer, which is stacked layer by layer to form a multi-layer neural network. These networks form a stack denoising automatic encoder. It automatically learns the compressed representation from the original input data in an unsupervised way. In 2016, Kulkarni et al.[15] used convolutional neural network for the first time in the problem of compressed sensing image reconstruction, and proposed the ReconNet network. In order to reduce the complexity of the network, the input image was segmented, and the random Gaussian matrix was used as the measurement matrix. The network used multiple convolution layers to realize non-iterative compression sensing reconstruction. Compared with SDA method, this network reduces the complexity of the network, retains a lot of details and achieves better reconstruction quality at a lower compression ratio. In 2018, Lu et al.[16] proposed a convolution compression sensing (ConvCSNet) network based on deep convolution neural network, which used convolution filter to recover the entire image. In 2019, Du et al. [17] proposed the full convolution measurement network, which improved the measurement part block by block and used the full convolution layer to obtain the measurement value from the complete image. This ensures the integrity of the original image structure information. In 2020, Shi et al.[18] proposed an image compression sensing framework based on convolutional neural network (CSNet+). The sampling network adaptively learns the sampling matrix from the training image. The initial reconstruction network is composed of convolutional layer and self-built composite layer. The depth reconstruction network uses the depth residual network to further refine the initial reconstruction image. Sun et al.[19] proposed a dual-path attention network (DPA-Net) for compression sensing image reconstruction. The network designs two paths. The structure path restores the background information and edge structure of the image. The texture path reconstructs the image details. The texture attention module connects the two paths.

1.2 Dense network

Huang et al.[20] proposed a dense convolutional network (DenseNet). DenseNet has achieved great success in image detection, semantic segmentation and other aspects. DenseNet makes use of the advantages of fast connection. Specifically, the output of each layer is composed of feature maps of all previous layers, and then transmitted to each subsequent layer, so that each layer can directly obtain the output of all previous layers.

Zhang et al.[21] proposed dual-channel reconstruction network (DCNet) for image compressed sensing, in which a single-channel DR^2-Net network composed of four residual blocks was analyzed. DCNet points out that when the measurement rate is 0.01 and 0.10, the second and fourth residual blocks of DR^2-Net have similar performance, but the performance of the fourth residual block is poor compared to the second residual block. Therefore, DCNet replaced some residual blocks with dense blocks. A dual-channel reconstruction network is proposed, in which residual blocks and dense blocks are connected in parallel. SAMNet also introduces dense blocks.

1.3 Attention mechanism

The attention mechanism used for convolution neural network can be roughly divided into two directions. One is to enhance feature aggregation, and the other is to combine channel and spatial attention[22-24]. Bello et al.[25] proposed AA-Net that can jointly participate in space and feature subspace. They obtain the attention weight through matrix operation, assign multiple spaces through multiple operations, and carry out attention point multiplication in multiple spaces to realize the self-attention mechanism. Srinivas et al.[26] proposed the BoTNet skeleton network, which replaced the last three bottleneck layers of ResNet with the self-attention mechanism. The network incorporated the self-attention mechanism[27,28] into various computer vision tasks, including image classification, target detection, and instance segmentation. The indicators have been greatly improved.

According to the development of the attention mechanism, SAMNet combines the self-attention mechanism with the convolution structure, and uses the convolutional layer with the self-attention mechanism in the sampling part to simulate the compression sampling process, which is beneficial to extract richer features, so that the compressed sensing measurement value retains more image structure information. For the reconstruction part, the bottleneck transformer is used to model the global information, which solves the problem of low learning efficiency and long reconstruction time caused by too many convolutional layers.

2 Image compressed sensing reconstruction network framework based on self-attention mechanism

Fig. 1 is a structural frame diagram of the image compressed sensing reconstruction network based on the self-attention mechanism ($\oplus$ represents the sum operation, © represents the Concat operation). The SAMNet first blocks the input image in the sampling network, and then performs self-attention convolution SA Conv operation from the non-overlapping block images, which makes the compressed sensing measurement value retain more image structure information, and is more conducive to improving the quality of image reconstruction. The reconstruction network includes an initial reconstruction network and a deep reconstruction network, which are used to learn the end-to-end mapping between CS measurements value and reconstructed images. The initial reconstruction network uses convolutional layers and combines layers (Reshape+Concat) to generate the initial reconstructed image. The size is 256×256, which is used as the input of the next layer. The method of using all measured values of an image for reconstruction is different from block-by-block reconstruction[29], which avoids the block effect introduced by the reconstructed image block, so the reconstruction of the entire image effectively reduces the block effect. The deep reconstruction network is composed of two paths. The first path includes convolution layer, residual block (Res Block)[30,31], and BoTNet. The residual blocks can enhance the quality of the reconstructed image. The self-attention mechanism can simulate global information. And the bottleneck transformer has solved the problems such as low learning efficiency, excessive computational cost, and long reconstruction time caused by too many convolution layers. The second path is dense block, which realizes the multiplexing of features in the channel dimension and strengthens the transfer of image features. The output images of the initial reconstruction network are input into the two paths, respectively. After completing a series of operations, the two paths finally obtain a reconstructed image of 256×256 with

64 channels, and the last layer uses a convolution kernel of size $f\times f\times d$ to generate the final reconstructed image.

2.1 Sampling network

According to Fig. 1, SA Conv is used in the sampling network. SA Conv first divides the image into non-overlapping image blocks. The size of each image block is $B\times B\times l$. The size of each convolution kernel in the sampling network is also $B\times B\times l$. And the stride of the convolutional layer is $B\times B$. So after the convolution operation, each convolution kernel outputs a measurement vector $\boldsymbol{y}_i(1,2,\cdots,n_B)$, which completes the convolution operation of non-overlapping image patches. And then these measurement vector $\boldsymbol{y}_i(1,2,\cdots,n_B)$ are fed into the self-attention mechanism.

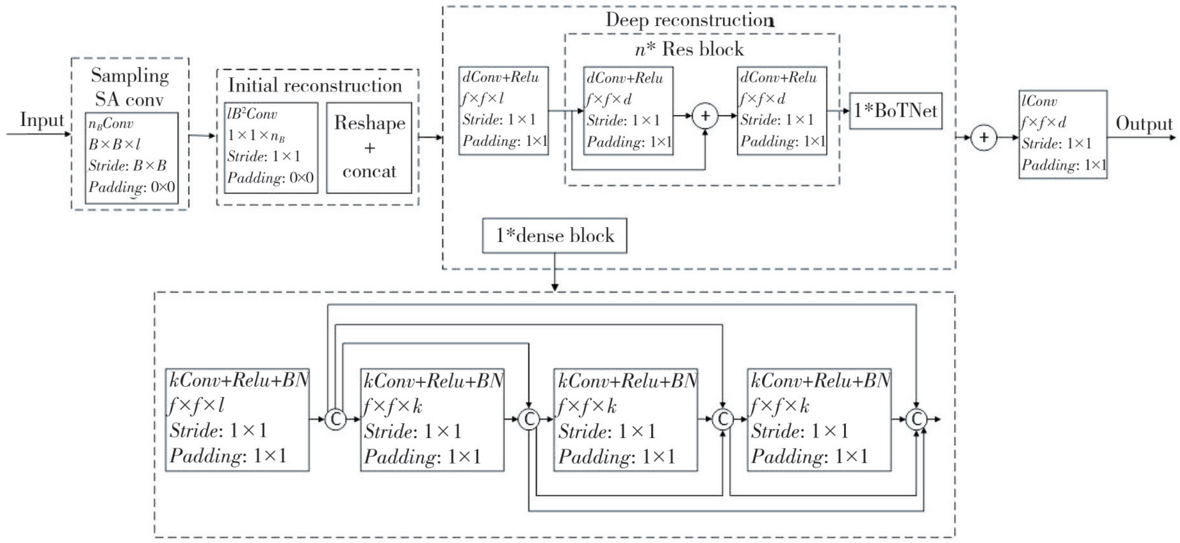


Fig. 1 SAMNet structure frame diagram

According to the principle of self-attention mechanism, there are three different vectors, as shown in Fig. 2. They are $\boldsymbol{Q}_i$ (Query) vector, $\boldsymbol{K}_i$ (Key) vector, and $\boldsymbol{V}_i$ (Value) vector, respectively, size is 1×1.

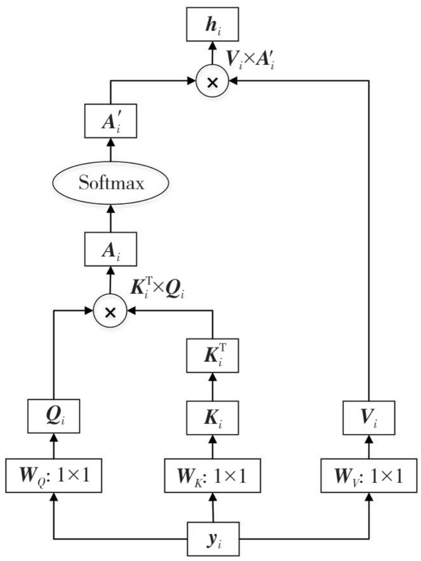


Fig. 2 Self-attention mechanism

The first step of self-attention is to calculate the three vectors $\boldsymbol{Q}_i$, $\boldsymbol{K}_i$, $\boldsymbol{V}_i$ by measuring the value $\boldsymbol{y}_i(1,2,\cdots,n_B)$.

$$\begin{cases}\boldsymbol{Q}_i=\boldsymbol{W}_Q\boldsymbol{y}_i,\\ \boldsymbol{K}_i=\boldsymbol{W}_K\boldsymbol{y}_i,\\ \boldsymbol{V}_i=\boldsymbol{W}_V\boldsymbol{y}_i,\end{cases} \tag{1}$$

where $\boldsymbol{Q}_i$ and $\boldsymbol{K}_i$ are used to calculate the importance of the feature, $\boldsymbol{V}_i$ is used to represent the input feature, and $\boldsymbol{W}_Q$, $\boldsymbol{W}_K$, $\boldsymbol{W}_V$ are the corresponding parameter matrices that can be learned. Then Softmax is used to normalize the vector

$$A_i{}'=Softmax\left(\boldsymbol{Q}_i,\boldsymbol{K}_i^{\mathrm{T}}\right). \tag{2}$$

The resulting value is multiplied by $\boldsymbol{V}_i$, which gives

$$\boldsymbol{h}_i=Attention\left(A_i{}',\boldsymbol{V}_i\right)=Attention\left(\left(\boldsymbol{Q}_i,\boldsymbol{K}_i^{\mathrm{T}}\right),\boldsymbol{V}_i\right). \tag{3}$$

In this way, the attention output $\boldsymbol{h}_i(1,2,\cdots,n_B)$ is obtained, and $\boldsymbol{h}_i$ is the same size as the input measurement value $\boldsymbol{y}_i$.

This paper sets the measurement rate to M/N, then the sampling point is $n_B=B^2M/N$, so there are n_B filters of size $B\times B\times l$ in the sampling layer. In the experiment, this paper set $B=32$, $l=1$. The self-attention mechanism of the sampling part is a single head, and when the measurement rate M/N is 0.1, it can be obtained according to the calculation that there are 102 filters.

2.2 Initial reconstruction network

The initial reconstruction network in Fig. 1 takes the output $\boldsymbol{h}_i$ of the self-attention convolution sampling layer as input. And $\boldsymbol{h}_i$ is $1\times1\times n_B$ vector, which is first convolved through the filter of $1\times1\times n_B$ with a stride of 1×1 to reconstruct each block. Each block constructed is still a vector, therefore, this paper designs a combination layer (Reshape+Concat). Reshape reconstructs each $\boldsymbol{B}^2\times1$ reconstructed vector into a $B\times B$ block, then through the Concat operation, the features of different scales are fused into the initial reconstructed image of 256×256 as the input of the next layer. In algorithms such as ReconNet and DR^2-Net, since a single image block is sampled and reconstructed block by block, there is no Concat operation. In the SAMNet algorithm, the Concat operation is used in the initial reconstruction part and is combined with the Reshape operation, so that all the block measurements of an image are used to restore the complete image at the same time, which effectively reduces the block effect.

2.3 Deep reconstruction network

In Section 2.2, only the initial reconstructed image is obtained. Next, the initial reconstructed image is used as input to further reconstruct the image through two paths, respectively. This is the depth reconstruction section in Fig. 1. The first path consists of convolutional layers, residual blocks, and bottleneck bransformer. The second path consists of dense blocks. Specifically, in the first path, there are first d filters with a convolution kernel size $f\times f\times l$ to generate d feature maps, followed by n residual blocks and a bottleneck transformer. This paper sets $f=3$, $d=64$, $n=2$, and stride and padding are both 1×1. The output $\boldsymbol{X}_m$ of the traditional network at layer m is

$$\boldsymbol{X}_m=\boldsymbol{H}_m(\boldsymbol{x}_{m-1}). \tag{4}$$

The difference between the residual block and the dense block is that the connection between the feature maps is different, as shown in Fig.3 and Fig.4, and the input $\boldsymbol{x}_0$ is an image block of size 32×32.

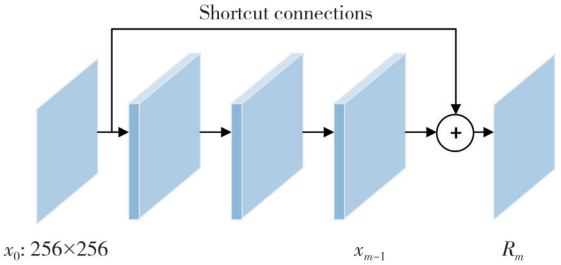


Fig. 3 Residual block

As shown in Fig.3, the input features enter the residual block in the form of two branches. The output generated by the linear branch of multiple convolution layers and the features transmitted through the shortcut connection branch are superimposed together and then input to the next layer. Therefore, the output $\boldsymbol{R}_m$ of the residual block after adding the input from the previous layer is

$$\boldsymbol{R}_m=\boldsymbol{H}_m(\boldsymbol{x}_{m-1})+\boldsymbol{x}_{m-1}. \tag{5}$$

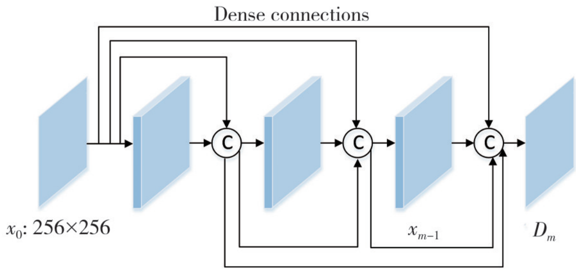


Fig. 4 Dense block

As shown in Fig. 4, in the dense block, in order to maximize the information flow between layers, all the previous dense layers are connected as inputs, that is, the input of a layer is the output of all the previous layers, so the output $\boldsymbol{D}_m$ of the m layer of the dense block is

$$\boldsymbol{D}_m=\boldsymbol{H}_m([\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{m-1}]), \tag{6}$$

where $[\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{m-1}]$ is the cascade of the previous $0,1,\cdots,m-1$ layers, $\boldsymbol{H}_m(\cdot)$ represents the nonlinear transformation function.

In this paper, the size of the feature map of each layer in the dense block is the same, and the nonlinear transformation function $\boldsymbol{H}_m(\cdot)$ adopts the structure of $BN+Relu+3\times3Conv$. If each function $\boldsymbol{H}_m$ generates k feature maps, then the number of characteristic graphs k_m of layer m is

$$k_m=k_0+k\times(m-1), \tag{7}$$

where k_0 is the number of channels in the input layer, and k is called the growth rate of the network. We set $k=16$. The dense block finally gets a reconstructed image of 256×256 with 64 channels. The last layer uses a convolution kernel of size $f\times f\times d$ to generate a feature map.

As shown in Fig. 5, the 3×3 convolution kernel in the residual block is replaced by the multi-head self-attention mechanism MHSA. This gets the bottleneck transformer. The multi-head self-attention is similar to the multi-channel of CNN. According to the principle of the self-attention mechanism, the multi-head self-attention H can be obtained by

$$H=Concat(H_1,H_2,H_3,H_4). \tag{8}$$

The two 1×1 convolutions in BoTNet are similar to the computation in Res Block. BoTNet first goes through 1×1 convolution to get the output of 64 channels. Next is MHSA. Unlike this, here multiple individual heads will be copied, and the number of these heads is set to 4. After the calculation of multi-head self-attention is completed, the Concat operation is used to output 64 channels. The final step is still to perform 1×1 convolution.

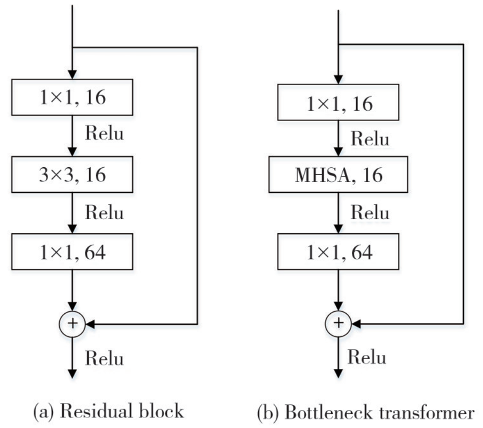


Fig. 5 Structure comparison diagram of residual block and bottleneck transformer

The output of the deep reconstruction network is expressed as $O(x)$, which can be obtained by

$$O(x)=C(x)\otimes nR(x)\otimes B(x)\oplus D(x), \tag{9}$$

where $C(x)$ represents the first convolutional layers, $\otimes$ represents the cascade operation, $\oplus$ represents the sum operation of the two paths, $R(x)$, $B(x)$, and $D(x)$ represent the residual block, bottleneck transformer, and dense block, respectively. Finally, SAMNet obtains the final reconstruction result through a convolutional kernel.

3 Experiment

In this section, extensive experiments are conducted on the Pytorch platform to test the performance of the proposed SAMNet and compare it with other algorithms, such as ReconNet, ConvCSNet, MSRNet, FDC-Net, DPA-Net and CSNet+. The computer is equipped with an Intel Core i5-8250U CPU clocked at 1.8 GHz and a Nvidia GeForce GTX 1080Ti GPU, and the network model runs on the Windows 10 operating system.

3.1 Training data and evaluation indicators

SAMNet selects DIV2K training dataset, which contains a total of 800 images. The size of training images is 256 × 256, they are further divided into 64 image patches of size 32 × 32. SAMNet uses two test datasets: Set5 (5 images) [32] and Set11 (11 images) [33]. Set5 and Set11 are shown in Fig. 6 and Fig. 7, respectively.

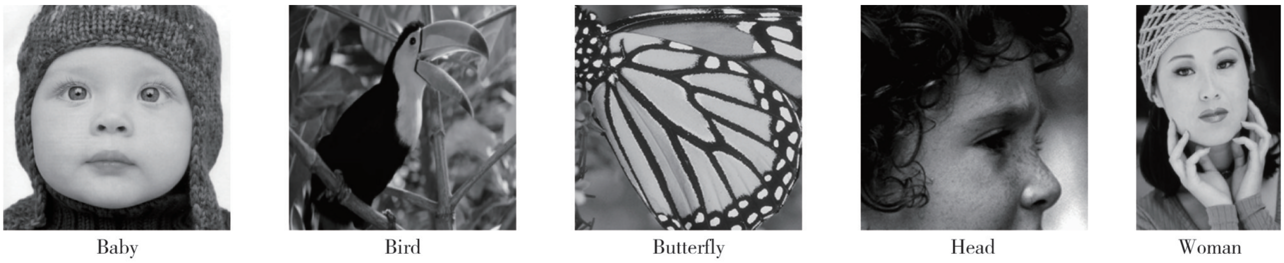


Fig. 6 Set5 dataset

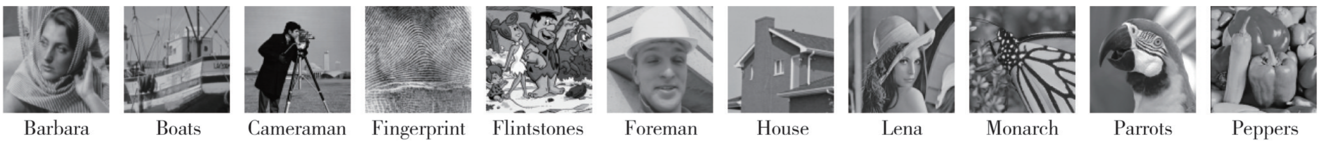


Fig. 7 Set11 dataset

This paper employs four different sampling rates (0.01, 0.04, 0.10, and 0.25) to sample the image blocks, each of which contains 1 024 pixels. The evaluation indicators of the final reconstructed image quality include time complexity, PSNR, and structural similarity (SSIM). The loss function is the mean squared error (MSE) loss function. The size of two images is $H\times W$. $X(i,j)$ represents the reconstructed image. $Y(i,j)$ represents the input image. Then MSE represents the MSE of X and Y.

$$L_{\mathrm{MSE}}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{H}\left(X(i,j)-Y(i,j)\right)^{2}. \tag{10}$$

3.2 Comparison with other methods

The present network structure is compared with ReconNet, ConvCSNet, MSRNet, FDC-Net, DPA-Net, CSNet+. PSNR and SSIM are selected as evaluation criteria. First, comparative experiments are done on Set5 and Set11. Tables 1 and 2 show the comparison with each algorithm when the MR is 0.01, 0.04, 0.10, and 0.25. The best results are highlighted in bold.

According to Table 1, it can be concluded that compared with CSNet+ , the average PSNR and SSIM of the proposed SAMNet on Set5 are improved by 0.79 dB and 0.032 3, respectively. It can be seen that the reconstruction effect of SAMNet is the best at all measurement rates. It illustrates the effectiveness and rationality of the SAMNet framework incorporating a self-attention mechanism and dense blocks.

Table 1 Comparison of average PSNR and SSIM of different algorithms on Set5

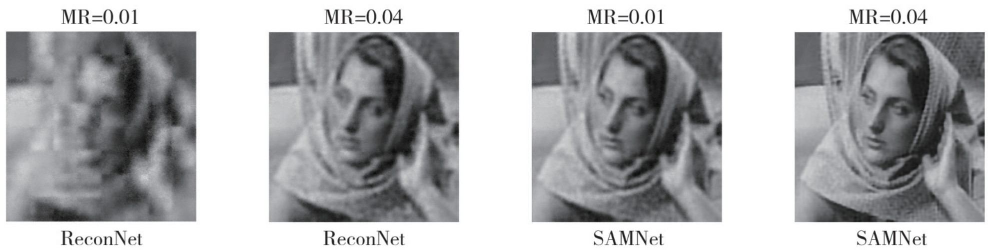
MR	ReconNet		DPA-Net		CSNet+		SAMNet	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
0.01	18.46	0.449 2	22.75	0.573 1	24.18	0.647 8	**25.45**	**0.743 7**
0.04	22.99	0.606 9	26.63	0.776 7	28.11	0.850 1	**29.34**	**0.853 3**
0.10	25.95	0.722 8	30.32	0.871 3	32.59	0.901 5	**33.09**	**0.929 8**
0.25	28.87	0.852 3	31.83	0.890 4	33.27	0.930 6	**33.42**	**0.932 5**
Average	24.07	0.657 8	27.88	0.777 9	29.54	0.832 5	**30.33**	**0.864 8**

Table 2 Comparison of average PSNR and SSIM of different algorithms on Set11

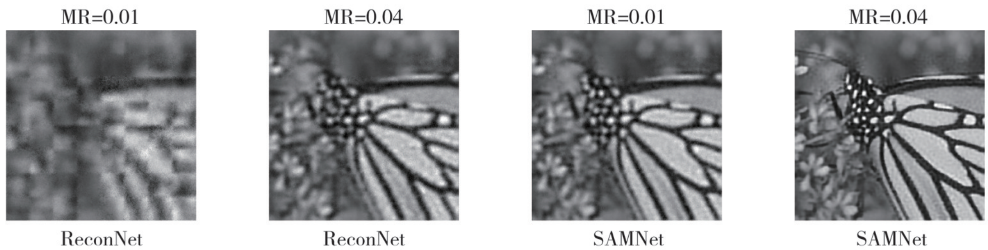
MR	ReconNet		MSRNet		DPA-Net		SAMNet	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
0.01	17.27	0.408 4	17.76	0.453 5	18.05	0.501 1	**22.43**	**0.657 3**
0.04	20.63	0.525 9	21.79	0.616 7	23.50	0.720 5	**25.39**	**0.723 8**
0.10	24.28	0.640 6	25.59	0.759 8	26.99	0.835 4	**28.95**	**0.864 0**
0.25	25.60	0.758 9	30.39	0.869 8	31.74	0.923 8	**31.96**	**0.925 7**
Average	21.95	0.583 5	23.88	0.675 0	25.07	0.745 2	**27.18**	**0.792 7**

According to Table 2, it can be concluded that the average SSIM of SAMNet is 0.156 2, 0.003 3, 0.028 6, and 0.001 9 higher then DPA-Net on the Set11 test dataset. It can be seen that SAMNet has the highest average SSIM at all measurement rates. It shows that SAMNet improves the quality of reconstructed images and obtains better reconstruction performance.

In order to observe the elimination of block effect, Barbara and Monarch in Set11 are selected for comparison with ReconNet, respectively, and the results are shown in Fig. 8. The image reconstructed by ReconNet is relatively blurry and has obvious block effect. The ReconNet obtains the reconstructed image by reconstructing and splicing block by block, and SAMNet reconstructs all the measured values of the entire image. The image reconstructed by the SAMNet effectively reduces the block effect and obtains a better visual effect.



(a) Barbara



(b) Monarch

Fig. 8 Barbara and Monarch reconstructed by ReconNet and SAMNet

For a comprehensive understanding of the performance of the algorithm, Table 3 shows the reconstruction results of 7 images in Set11. The SAMNet achieves high PSNR at four measurement rates, and the best results are highlighted in bold show.

Under the given four measurement rates, this paper compares SAMNet with ReconNet, ConvCSNet, MSRNet, FDC-Net, DPA-Net, and CSNet+ algorithms. The PSNR of each reconstructed image under SAMNet has been improved to varying degrees, which shows that the reconstruction quality of SAMNet is the best.

In order to verify the good applicability of the SAMNet, the experiments are compared with the ReconNet and MSRNet on the dataset BSD500, and the best results are highlighted in bold. As shown in Table 4, the average PSNR and SSIM of SAMNet are significantly higher than those of ReconNet and MSRNet under different measurement rates, and the highest values are 29.73 dB and 0.895 7, which indicates that the reconstructed image quality of the SAMNet is good.

Table 3 PSNR of test images under different measurement rates for different algorithms

Image name	Algorithm	PSNR/dB			
		MR=0.01	MR=0.04	MR=0.10	MR=0.25
Barbara	ReconNet	19.07	20.20	22.51	23.55
	ConvCSNet	18.35	21.00	23.01	25.98
	MSRNet	18.90	21.28	23.06	26.91
	FDC-Net	21.71	23.71	24.52	28.00
	DPA-Net	20.03	23.89	24.65	28.50
	CSNet+	21.79	24.18	24.41	28.62
	SAMNet	**22.24**	**24.57**	**26.32**	**29.19**
Cameraman	ReconNet	17.49	19.73	21.67	23.61
	ConvCSNet	17.92	20.01	22.69	25.26
	MSRNet	18.79	19.95	22.27	26.06
	FDC-Net	20.21	23.08	25.31	29.04
	DPA-Net	21.29	24.64	26.99	30.19
	CSNet+	21.65	26.33	28.15	30.57
	SAMNet	**21.90**	**26.77**	**28.21**	**30.82**
Flintstones	ReconNet	14.07	16.56	19.20	22.60
	ConvCSNet	15.32	17.58	19.82	26.53
	MSRNet	14.10	17.40	21.81	26.89
	FDC-Net	16.49	20.54	24.55	28.91
	DPA-Net	16.53	21.63	**25.76**	29.13
	CSNet+	16.65	21.82	24.31	**29.21**
	SAMNet	**16.89**	**22.59**	25.08	29.06
Lena	ReconNet	18.07	21.83	24.51	26.55
	ConvCSNet	18.16	22.08	25.61	27.32
	MSRNet	18.35	23.06	26.41	30.37
	FDC-Net	21.71	26.38	28.93	32.91
	DPA-Net	22.15	27.15	29.12	33.09
	CSNet+	22.45	27.63	29.15	32.86
	SAMNet	**22.73**	**28.02**	**30.86**	**33.25**
Monarch	ReconNet	15.47	18.33	21.51	25.05
	ConvCSNet	16.81	19.18	22.01	26.71
	MSRNet	15.61	19.48	24.17	29.04
	FDC-Net	18.46	23.52	27.83	31.97
	DPA-Net	19.44	21.48	28.04	32.21
	CSNet+	19.85	23.61	29.39	**32.87**
	SAMNet	**20.43**	**25.57**	**30.21**	32.56
Parrot	ReconNet	18.31	20.06	23.25	26.22
	ConvCSNet	18.15	21.18	24.85	26.38
	MSRNet	18.05	21.08	23.97	28.43
	FDC-Net	22.25	24.74	27.84	31.94
	DPA-Net	22.83	25.23	28.30	32.49
	CSNet+	23.08	26.42	28.00	33.01
	SAMNet	**23.89**	**26.63**	**29.05**	**33.63**
Peppers	ReconNet	16.98	20.00	22.68	25.15
	ConvCSNet	18.01	21.08	23.66	26.51
	MSRNet	17.33	21.16	25.18	29.86
	FDC-Net	20.56	24.72	27.21	32.92
	DPA-Net	21.31	25.31	28.49	32.94
	CSNet+	21.94	26.28	29.39	33.27
	SAMNet	**22.37**	**26.88**	**29.92**	**33.73**

3.3 Time complexity

SAMNet is significantly less than ReconNet and MSRNet in terms of running time. The relevant results are shown in Table 5, where the best results are highlighted in bold.

The running time of SAMNet is shorter than ReconNet and MSRNet, which shows that the reconstruction time of SAMNet is significantly less than other algorithms.

Table 4 Average PSNR and SSIM of each algorithm on BSD500

Algorithm	MR=0.25		MR=0.10		MR=0.04		MR=0.01	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
ReconNet	25.48	0.724 1	23.28	0.612 1	21.40	0.514 9	19.17	0.424 7
MSRNet	27.93	0.818 2	24.73	0.683 7	22.25	0.569 6	19.35	0.454 1
SAMNet	**29.73**	**0.895 7**	**27.09**	**0.758 8**	**25.38**	**0.719 7**	**22.76**	**0.587 4**

Table 5 Running time of reconstructed 256×256 images

Algorithm	Running time/s			
	MR=0.25	MR=0.10	MR=0.04	MR=0.01
ReconNet	0.535 8	0.536 3	0.536 6	0.536 1
MSRNet	0.520 6	0.515 2	0.517 2	0.488 4
SAMNet	**0.388 5**	**0.357 4**	**0.305 6**	**0.283 7**

4 Conclusions

We proposed an image compressed sensing reconstruction network based on self-attention mechanism (SAMNet). The sampling network performed convolution and self-attention operations from non-overlapping patch images. It made the compressed sensing measurement value retain more image structure information, which was more conducive to image reconstruction. The initial reconstruction network used convolutional layers and combined layers to generate initial reconstructed images, and the deep reconstruction network used residual blocks, bottleneck transformer, and dense blocks. The experimental results on the commonly used Set5 and Set11 test data sets showed that SAMNet effectively reduced the block effect and shortened the reconstruction time.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Nos. 61261016, 61661025) and Science and Technology Plan of Gansu Province (No.20JR10RA273).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] SHANNON C E. Communication in the presence of noise. Proceedings of the IRE, 1949, 37(1): 10-21.

[2] DONOHO D L. Compressed sensing. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.

[3] YAO K. Research on OFDM channel estimation based on compressed sensing. Beijing: Beijing University of Posts and Telecommunications, 2021.

[4] WANG Q. Information processing method for wireless communication based on compressed sensing. China New Communications, 2021, 23(13): 9-10.

[5] NARAYANAN S, SAHOO S K, MAKUR A. Greedy pursuits assisted basis pursuit for reconstruction of joint-sparse signals. Signal Processing, 2018, 142: 485-491.

[6] LEE J, CHOI J W, SHIM B. Sparse signal recovery *via* tree search matching pursuit. Journal of Communications and Networks, 2016, 18(5): 699-712.

[7] JI S H, XUE Y, CARIN L. Bayesian compressive sensing. IEEE Transactions on Signal Processing, 2008, 56(6): 2346-2356.

[8] WANG B, MA S X. Improvement of Gaussian random measurement matrices in compressed sensing. Advanced Materials Research, 2011, 301/302/303: 245-250.

[9] MOUSAVI A, DASARATHY G, BARANIUK R G. DeepCodec: adaptive sensing and recovery *via* deep convolutional neural networks. 2017: arXiv: 1707.03386. https://arxiv.org/abs/1707.03386

[10] YAO H T, DAI F, ZHANG S L, et al. DR^2-Net: deep residual reconstruction network for image compressive sensing. Neurocomputing, 2019, 359: 483-493.

[11] LIAN Q S, FU L P, CHEN S Z, et al. A compressed sensing algorithm based on multi-scale residual reconstruction network. Acta Automatica Sinica, 2019, 45(11): 2082-2091.

[12] XIE X M, WANG Y X, SHI G M, et al. Adaptive measurement network for CS image reconstruction//Computer Vision. Singapore: Springer Singapore, 2017: 407-417.

[13] ZHANG J, GHANEM B. ISTA-net: interpretable optimization-inspired deep network for image compressive sensing//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 1828-1837.

[14] MOUSAVI A, PATEL A B, BARANIUK R G. A deep learning approach to structured signal recovery//2015 53rd Annual Allerton Conference on Communication, Control, and Computing, September 29-October 2, 2015, Monticello, IL, USA. New York: IEEE, 2016: 1336-1343.

[15] KULKARNI K, LOHIT S, TURAGA P, et al. ReconNet: non-iterative reconstruction of images from compressively sensed measurements//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 449-458.

[16] LU X T, DONG W S, WANG P Y, et al. ConvCSNet: a convolutional compressive sensing framework based on deep learning. 2018: arXiv: 1801.10342. https://arxiv.org/abs/1801.10342

[17] DU J, XIE X M, WANG C Y, et al. Fully convolutional measurement network for compressive sensing image reconstruction. Neurocomputing, 2019, 328: 105-112.

[18] SHI W Z, JIANG F, LIU S H, et al. Image compressed sensing using convolutional neural network. IEEE Transactions on Image Processing, 2020, 29: 375-388.

[19] SUN Y B, CHEN J W, LIU Q S, et al. Dual-path attention network for compressed sensing image reconstruction. IEEE Transactions on Image Processing, 2020, 29: 9482-9495.

[20] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.

[21] ZHANG Z Q, GAO D H, XIE X M, et al. Dual-channel reconstruction network for image compressive sensing. Sensors, 2019, 19(11): 2549.

[22] LI X, WANG W H, HU X L, et al. Selective kernel networks//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2020: 510-519.

[23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7132-7141.

[24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module. Computer Vision-ECCV 2018. Cham: Springer, 2018: 3-19.

[25] BELLO I, ZOPH B, LE Q, et al. Attention augmented convolutional networks//2019 IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019. Seoul, Korea. New York: IEEE, 2019: 3285-3294.

[26] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25, 2021. Nashville, TN, USA. New York: IEEE, 2021: 16514-16524.

[27] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module. 2018: arXiv: 1807. 06514. https://arxiv. org/abs/1807. 06514.

[28] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations. 2018: arXiv: 1803. 02155. https://arxiv. org/abs/1803. 02155.

[29] XIE X M, WANG C Y, DU J, et al. Full image recover for block-based compressive sensing//2018 IEEE International Conference on Multimedia and Expo, July 23-27, 2018, San Diego, CA, USA. New York: IEEE, 2018: 1-6.

[30] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.

[31] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 1132-1140.

[32] SHI W Z, JIANG F, ZHANG S P, et al. Deep networks for compressed image sensing//2017 IEEE International Conference on Multimedia and Expo, July 10-14, 2017, Hong Kong, China. New York: IEEE, 2017: 877-882.

[33] ZHOU S W, HE Y, LIU Y H, et al. Multi-channel deep networks for block-based image compressive sensing. IEEE Transactions on Multimedia, 2021, 23: 2627-2640.

基于自注意力机制的图像压缩感知重构网络

刘玉红*, 刘晓燕, 陈满银

兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070

摘　要： 在图像压缩感知重构中，大量算法使用了逐块重构图像块和堆叠很多个卷积层的方法，这通常会产生明显的块效应、计算复杂度高和重构时间长的问题。本文提出了一种基于自注意力机制的图像压缩感知重构网络(SAMNet)。在压缩采样部分设计了自注意力卷积，利于捕捉更丰富的特征，使压缩感知测量值保留更多的图像结构信息。在重构部分，在卷积神经网络中加入了自注意力机制，提出了包含残差块、瓶颈变压器和密集块的重构网络，加强了图像特征的传递，减少了参数量。在Set5数据集下，采样率分别为0.01、0.04、0.10和0.25时，算法的峰值信噪比(Peak signal-to-noise ratio, PSNR)平均值比CSNet+分别提高1.27、1.23、0.50和0.15 dB，重构一张256×256图像的运行时间比ReconNet降低了0.147 3、0.178 9、0.231 0和0.252 4 s。实验结果表明，SAMNet提升了重构图像的质量，并减少了重构时间。

关键词： 卷积神经网络；压缩感知；自注意力机制；密集块；图像重构

引用格式： LIU Yuhong, LIU Xiaoyan, CHEN Manyin. Image compressed sensing reconstruction network based on self-attention mechanism. Journal of Measurement Science and Instrumentation, 2025, 16(4): 537-546. DOI: 10.62756/jmsi.1674-8042.2025052