Block Image Compressive Sensing with Local and Global Information Interaction

Xiaoyu Kong[1], Yongyong Chen[1]*, Feng Zheng[2], Zhenyu He[1]*

[1] Harbin Institute of Technology (Shenzhen)
[2] Southern University of Science and Technology
{xiaoykong15,YongyongChen.cn, zfeng02}@gmail.com, zhenyuhe@hit.edu.cn

Abstract

Block image compressive sensing methods, which divide a single image into small blocks for efficient sampling and reconstruction, have achieved significant success. However, these methods process each block locally and thus disregard the global communication among different blocks in the reconstruction step. Existing methods have attempted to address this issue with local filters or by directly reconstructing the entire image, but they have only achieved insufficient communication among adjacent pixels or bypassed the problem. To directly confront the communication problem among blocks and effectively resolve it, we propose a novel approach called **B**lock **R**econstruction with **B**locks' **C**ommunication **N**etwork (BRBCN). BRBCN focuses on both local and global information, while further taking their interactions into account. Specifically, BRBCN comprises dual CNN and Transformer architectures, in which CNN is used to reconstruct each block for powerful local processing and Transformer is used to calculate the global communication among all the blocks. Moreover, we propose a global-to-local module (G2L) and a local-to-global module (L2G) to effectively integrate the representations of CNN and Transformer, with which our BRBCN network realizes the bidirectional interaction between local and global information. Extensive experiments show our BRBCN method outperforms existing state-of-the-art methods by a large margin. The code is available at https://github.com/XYkong-CS/BRBCN

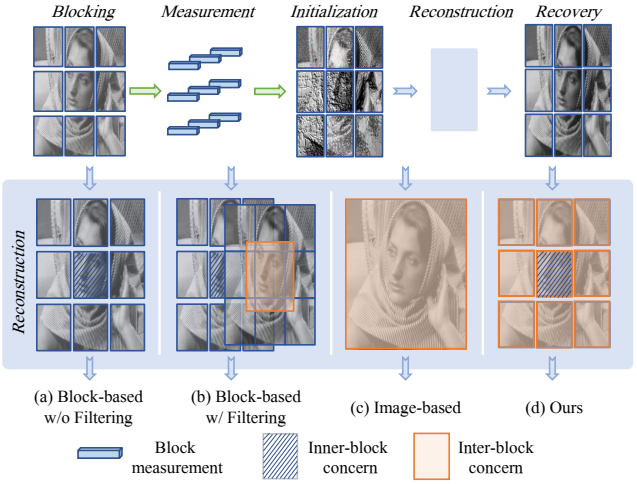

Figure 1: The procedure of block CSI. The whole image is split into non-overlapping blocks and is reconstructed with different strategies. (a): Block-based reconstruction without interaction among blocks. (b): Block-based reconstruction with local filtering operations, lacking sufficient blocks communication. (c): Image-based reconstruction with global concern, lacking inner-block concern. (d): Block-based reconstruction with local and global interaction.

Introduction

Compressive Sensing (CS) (Donoho 2006) is a signal processing technique capable of recovering high-dimensional signals from limited measurements with high probability. Demonstrating the potential for enhancing sampling speed and reducing storage and transmission costs, CS has aroused significant interest in various applications like medical imaging (Michailovich, Rathi, and Dolui 2011), single-pixel imaging (Duarte et al. 2008), image encryption (Li, Zhang, and Xie 2019), and snapshot compressive imaging (Meng et al. 2021; Wu, Zhang, and Mou 2021).

To handle arbitrary image resolution with fast processing, block compressive sensing imaging (CSI) divides the image into small non-overlapping blocks and samples and reconstructs each block independently (Gan 2007) as shown in Fig. 1. Traditional CSI methods (Dong et al. 2014; Gao et al. 2015) imported different priors and reconstructed the original image by non-linear iterative algorithms (Boyd et al. 2011; Donoho, Maleki, and Montanari 2009). Moreover, some researchers extended these traditional methods into deep learning, referred to deep unfolding networks (DUNs) (Fan, Lian, and Quan 2022; Yang et al. 2018; Zhang and Ghanem 2018; Zhang, Zhao, and Gao 2020). However, as depicted in Fig. 1(a), these algorithms reconstruct each block separately, thereby only focusing on local information.

To involve global information, there are two popular solutions: block-based methods with post-processing of local filters and image-based methods. The former added a local filter acting on the whole image after block reconstruction (Gan 2007; Zhang et al. 2020) as shown in Fig. 1(b). Meanwhile, image-based methods (Shi et al. 2019; Song,

*These authors are corresponding authors.

Chen, and Zhang 2021) reconstruct the image on a global level as shown in Fig. 1(c). Although they successfully relieved the blocking artifacts, local filters brought insufficient global information and the image-based methods ignored the local information, thus causing a decline in reconstruction quality. Ideally, an effective approach would incorporate both local pattern representations inner each block as well as the global information inter blocks, yet current research tends to focus on only one or the other. Moreover, appropriately integrating these types of information remains a challenging and unexplored area of study.

In this paper, we propose a **B**lock **R**econstruction with **B**locks' **C**ommunication **N**etwork (BRBCN). Our BRBCN is in DUN framework, as shown in Fig. 2, and it facilitates simultaneous local reconstruction and global communication in each stage. The reconstruction component comprises two branches: a CNN to process each block locally, and a Transformer for computing global communication between all blocks. Moreover, how block representation is fused with blocks' communication needs to design carefully, as a well-executed fusion can result in a mutually beneficial outcome, while an ill-conceived approach might impede progress. To address this, we implement the CNN and Transformer in parallel and propose a global-to-local (G2L) module and a local-to-global (L2G) module, realizing an effective bi-directional interaction. As shown in Fig. 1(d), our BRBCN concurrently considers both local and global information, enabling detailed and global image reconstruction. Our main contributions are summarized as follows:

- We propose BRBCN for block-based compressive sensing, leveraging full interaction between local and global information to enhance block reconstruction quality.
- We propose local-to-global and global-to-local modules to establish bi-directional interaction between local block representation and global communication information.
- Extensive experiments demonstrate that our approach surpasses existing state-of-the-art networks in quantitative evaluation and visual comparison, particularly in the low sampling ratio which has practical values.

Related Work

Block-based Block Compressive Sensing

Block CSI (Gan 2007; Mousavi, Patel, and Baraniuk 2015) was proposed to address the slow sampling and reconstruction stemming from the large measurement matrix of high-resolution images. It divided the image into small non-overlapping blocks and samples them independently. Some traditional methods (Li et al. 2013; Zhang, Zhao, and Gao 2014) introduced extra regularizers and some deep methods (Kulkarni et al. 2016; Mousavi and Baraniuk 2017) used DNN to improve the block reconstruction quality. Moreover, some researches (Gregor and LeCun 2010; Metzler, Maleki, and Baraniuk 2016) handled the CS problem from the perspective of denoising. To further improve the reconstruction quality, some methods combined deep learning with traditional non-linear iterative algorithms such as ISTA-Net (Zhang and Ghanem 2018) and ADMM-CSNet (Yang et al. 2018). The above schemes with theoretical convergence guarantee and local concern, however, ignored the information contained among blocks. To mitigate this issue, AMP-Net (Zhang et al. 2020) designed a block denoising procedure followed by some CNN layers on the whole image. However, these solutions only focus on realizing the interaction at the boundary between the adjacent blocks.

Image-based Block Compressive Sensing

The image-based methods utilized deep networks to map the corrupted image into a clean one. Unlike the traditional optimization algorithms on which block-based methods rely, deep networks offered greater flexibility. This means that block-wise measurements don't necessarily need to be recovered block by block. CSNet (Shi et al. 2019) initialized the entire image from the block measurements and then fed it into a CNN to directly obtain the whole image reconstruction result. OPINE-Net (Zhang, Zhao, and Gao 2020) and MADUN (Song, Chen, and Zhang 2021) used deep networks as deep prior, enabling their models to be trained and tested on both blocks and images. Although these methods avoided the problem of block isolation, they put little attention on specific inner-block information.

Fusion of CNN and Transformer

CNNs (Krizhevsky, Sutskever, and Hinton 2012) have secured a dominant position in computer vision owing to their powerful local representation ability. Meanwhile, driven by the success of the Transformer (Vaswani et al. 2017) in natural language processing, Vision Transformer (ViT) (Dosovitskiy et al. 2020) has shown the effectiveness of attention mechanism for image processing. Recent works (Chen et al. 2022a,b; Mehta and Rastegari 2021; Wu et al. 2021) have shown the benefits of combining CNNs with ViT. BoTNet (Srinivas et al. 2021) used self-attention blocks in ResNet (He et al. 2016), enhancing both object detection and segmentation performance. Mobile-Former (Chen et al. 2022a) and MobileViT (Mehta and Rastegari 2021) successfully devised lightweight models for image processing backbones. These recent models have highlighted the potential of utilizing CNNs for local processing and Transformers for encoding global interactions. This inspiration led us to incorporate them into block CSI, enabling simultaneous processing of inner-block patterns and inter-block communication. However, existing CNN and Transformer combination methods were typically intended for the entire image. The effective fusion of block representations obtained by CNNs and the global information acquired by Transformers remains unexplored and presents a challenging aspect.

Methods

Preliminary

The goal of CSI is to recover the original image from its linear measurements. Mathematically, given a linear measurement Y, the original image X can be recovered by solving the following optimization problem:

$$\arg\min_{X} \frac{1}{2}\|\Phi X - Y\|_2^2 + \lambda G(X), \tag{1}$$

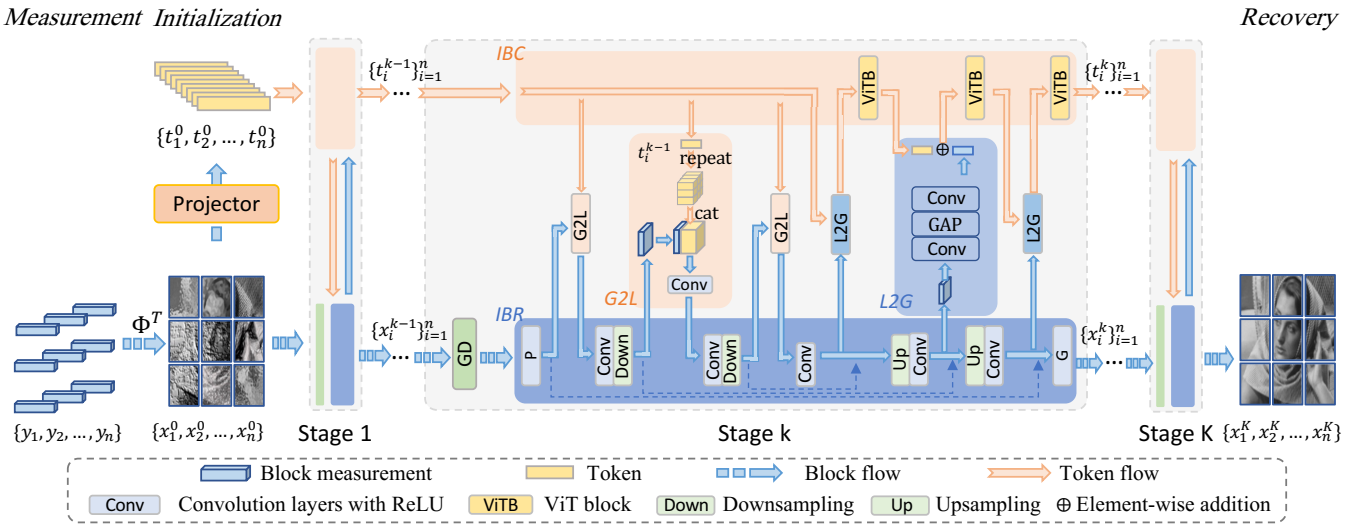

Figure 2: The pipeline of our proposed BRBCN. BRBCN takes the block measurements as input and outputs the recovery image by concatenating the reconstructed blocks. The reconstruction includes K stages and the block reconstruction with blocks communication happens within each stage.

where Φ is the measurement matrix and $G(X)$ is the regularization term with a weight λ. To solve this optimization problem, DUNs (Yang et al. 2018; Zhang and Ghanem 2018) unroll the network with a gradient descent (GD) : $R^k = X^{k-1} - \rho\Phi^T(\Phi X^{k-1} - Y)$ and a proximal mapping : $X^k = Prox(R^k)$.

For efficient sample and reconstruction, block CSI splits the image into non-overlapping blocks to sample and reconstruct independently. However, the block-wise strategy ignores the communication among different blocks. In this paper, we propose BRBCN, which focuses on improving the block reconstruction quality with the help of interaction between local and global information.

Overview of BRBCN

The overall architecture of our BRBCN is shown in Fig. 2. BRBCN takes the block measurements $\{y_i\}_{i=1}^n \in \mathbb{R}^m$ of image blocks $\{x_i\}_{i=1}^n \in \mathbb{R}^{B\times B}$ as input and outputs the recovery X, where m is the length of the measurements, $B \times B$ is the block size, and $\frac{m}{B\times B}$ is the sampling ratio.

Before the reconstruction step, an initialization step converts 1D measurements into 2D images. For easy implementation, the initialized blocks $\{x_i^0\}_{i=1}^n$ are obtained via the transposed sampling matrix Φ^T. Meanwhile, for the input of the following Transformer architecture, tokens $\{t_i^0\}_{i=1}^n \in \mathbb{R}^{C_T}$ are initialized from the flattened block initialization with a fully connected layer projector, where C_T is the tokens dimension. The number of blocks and tokens is the same, which means they are in one-to-one correspondence.

As shown in Fig. 2, the reconstruction part follows the general DUN framework with several repeated stages whose inputs are $\{x_i^0\}_{i=1}^n$. At the beginning of each stage k, since the GD step is the trivial solution for the fidelity term in Eq. (1), we keep the use of it and set the step size ρ as a stage-wise trainable parameter, which can be written as:

$$r_i^k = x_i^{k-1} - \rho^k\Phi^T(\Phi x_i^{k-1} - y_i), \tag{2}$$

whose input $\{x_i^{k-1}\}_{i=1}^n$ is the output of the last $(k-1)$-th stage. Instead of processing each block separately in the $Prox$ operator, BRBCN utilizes the communication among all the blocks to help reconstruct each block.

Reconstruction with Information Interaction

The key points of this section can be divided into four parts: reconstruct each block's pattern locally, make full global communication among blocks, use the global information to help block reconstruction, and make the local representation feedback to the global communication. Specifically, given the outputs $\{r_i^k\}_{i=1}^n$ from GD, the reconstruction is conducted within the following two steps:

$$x_i^k = IBR(r_i^k, G2L(r_i^k, t_i^{k-1})), \tag{3}$$

$$t_i^k = IBC(t^{k-1}, L2G(t_i^{k-1}, x_i^k)), \tag{4}$$

where the inner-block reconstruction (IBR) module, the inter-block communication (IBC) module, the global-to-local $(G2L)$ module, and the local-to-global $(L2G)$ module correspond to the four parts mentioned above. The four parts are closely linked together and, for a better understanding, we will first introduce the IBR and IBC because both $G2L$ and $L2G$ involve their intermediate products.

Local inner-block reconstruction module. Compared with the whole image reconstruction, the reconstruction of each block is also important since the block reconstructions are related to corresponding block measurements. Therefore, instead of directly recovering the concatenated image, IBR module follows the block-wise strategy with CNN architecture which has a good ability for local processing. IBR begins with a CNN projector P to project the single-channel $\{r_i^k\}_{i=1}^n$ into multi-channel feature map $F_{i\_enc\_0}$. In

the end, an inverse-projector G is used to project the feature map back into a single-channel image. Between them, an UNet architecture is adopted to extract and reconstruct the block representation, with representing encoder features as $\{F_{i\_enc\_1}, F_{i\_enc\_2}\}$, middle feature as $F_{i\_dec\_0}$, and decoder features as $\{F_{i\_dec\_1}, F_{i\_dec\_2}\}$ as shown in Fig. 2. Similar to many UNet-based structures (Lin et al. 2021; Mou, Wang, and Zhang 2022), multiple bypasses are added to avoid information loss due to the downsampling operation. The multi-scale IBR can effectively extract the representation of each block as much as possible because the block size B is always small, for example, 32.

Global inter-block communication module. We argue that global information among different blocks is another important factor in block CSI. Compared with the limited receptive field of CNN, Transformer has full global sense and is therefore more suitable for global information modelling. Thus, IBC builds the communication among all blocks with Transformer for global communication encoding. Notice that we initialize a group of tokens in the initialization step, IBC updates these tokens within each stage with three standard ViT blocks (Dosovitskiy et al. 2020) which contain a multi-head self-attention operation and FFN operation, representing as $\{T^k_{i\_1}, T^k_{i\_2}, T^k_{i\_3}\}$ and $t^k_i = T^k_{i\_3}$.

Global-to-local module. To utilize the inter-block communication information to help the inner-block reconstruction, multiple $G2L$ modules are inserted into the encoder of IBR as shown in Fig. 2. Unlike recent works of CNN-Transformer fusion methods (Chen et al. 2022a,b) in which both CNN and Transformer representations come from the whole image, BRBCN focuses on passing the global information into each block. Given the token $t^{k-1}_i \in \mathbb{R}^{C_T}$ for the $i\text{-}th$ block, it is first repeated into the same resolution as the block encoder feature $\{F_{i\_enc\_n}\}^2_{n=0} \in \mathbb{R}^{C_{enc} \times W \times H}$ and then concatenated together to form the feature fusion maps $\in \mathbb{R}^{(C_{enc}+C_T)\times W \times H}$. Then, we use convolution layers with $ReLU$ activation to extract the fused features. In this manner, each pixel of the block feature maps can have a sense of the entire global information brought by the token. Meanwhile, with multi-scale interaction, the information passing from global to local can be more sufficient.

Local-to-global module. Although the Transformer-based IBC module can update each token with global communication, it lacks local attention and image inductive bias. To relieve these problems, $L2G$ passes the local information extracted by CNN to strengthen the representation of tokens. As shown in Fig. 2, the $L2G$ module contains two 1×1 convolution layers with a global average pooling to fuse the decoder feature $\{F_{i\_dec\_n}\}^2_{n=0}$ with the $i\text{-}th$ token. In this way, the latest pattern activation contained in each block can be always passed to each token timely. Moreover, the t^k_i contains the multi-scale information before the inverse-projection G which may cause serve information loss due to the channel reduction. This information is then provided to the $(k+1)\text{-}th$ stage the with $G2L$ module, which means the $L2G$ module also works as a cross-stage connection to compensate for the information loss within iterations.

Loss Function

Given the training data $\{X_j\}^{N_d}_{j=1}$, BRBCN takes image X_j as input and outputs the reconstructed result with a trainable sampling matrix Φ. Our network is trained in an end-to-end way with the commonly used $\mathcal{L}_2$ loss function as:

$$\mathcal{L}_{rec} = \frac{1}{N_d}\sum_{j=1}^{N_d} \|X^*_j - X_j\|^2_2, \tag{5}$$

where $X^*_j = Cat(x^K_1, x^K_2, ..., x^K_n)$ is the recovered image with the concatenation operation $Cat()$ and K denotes the number of iterations.

Enhanced version of BRBCN: BRBCN$^+$

Although BRBCN greatly strengthens the interaction between local and global representations, the global representation obtained by Transformer only focuses on the texture information but has a weak sense of the position information. However, the relation between one block and its all adjacent blocks is also important, without which would lead to blocking artifacts. To this end, the enhanced version BRBCN$^+$ is proposed by adding an image-level IBR to process the whole image which is concatenated with the recovered blocks $\{x^K_i\}^n_{i=1}$. Notice that the extra IBR is inserted only once to the end of BRBCN, instead of being added in each stage which may greatly smooth the fine-grained details in the block reconstruction.

Experiments

Implementation Details

Since Transformer is data-hungry, We use ImageNet to train BRBCN and all images are converted to gray-scale and resized into 256×256. The training epochs and batch size are three and eight, respectively. The Adam optimization strategy is applied, with a learning rate of 10^{-4} for the early two epochs and then reduced to 10^{-5} for the last epoch. Five sampling ratios are investigated, including low ratios 0.01 and 0.04, middle ratios 0.1 and 0.25, and a higher ratio 0.5. The default iteration time K and block size B are set to be 8 and 32. The tokens dimension C_T is set as 128.

Two gray-scale datasets including Set14 (14 images) (Zeyde, Elad, and Protter 2010) and BSD68 (68 images) (Sapiro 2008) and one color dataset Waterloo (4744 images) (Ma et al. 2016) are used. The CS reconstruction accuracies on all the datasets are evaluated with peak signal-to-noise ratio (PSNR) and structure similarity index measure (SSIM). In general, the reconstructed images with higher PSNR and SSIM values denote better reconstruction performance. We implement the model using PyTorch, and train and test it on Nvidia RTX 3090 GPU.

Comparison with State-of-the-arts Methods

We compare the proposed BRBCN with six state-of-the-art CSI methods, including ISTA-Net$^+$(Zhang and Ghanem 2018), CSNet$^+$(Shi et al. 2019), AMP-Net(Zhang et al. 2020), OPINE-Net$^+$(Zhang, Zhao, and Gao 2020), MADUN(Song, Chen, and Zhang 2021), and

Datasets	Methods	Sampling ratio					
		0.01	0.04	0.10	0.25	0.5	avg
Set14	ISTA-Net$^+$	18.26/0.4000	22.07/0.5687	25.92/0.7268	30.50/0.8688	35.87/0.9481	26.52/0.7025
	OPINE-Net$^+$	21.39/0.5243	25.47/0.7110	28.67/0.8279	32.98/0.9186	37.98/0.9661	29.39/0.7896
	CSNet$^+$	21.36/0.5189	24.81/0.6930	27.62/0.8152	31.74/0.9112	36.53/0.9625	28.41/0.7802
	AMP-Net	21.66/0.5412	25.44/0.6990	28.68/0.8168	33.10/0.9134	38.14/0.9649	29.40/0.7871
	MADUN	21.49/0.5376	25.45/0.7235	28.85/0.8424	33.15/0.9265	36.95/0.9625	29.18/0.7985
	DGUNet$^+$	21.87/0.5411	25.87/0.7249	29.35/0.8455	33.70/0.9294	38.83/0.9709	29.92/0.8024
	BRBCN	21.87/0.5332	26.17/0.7317	29.51/0.8460	34.28/0.9328	**39.56/0.9732**	30.28/0.8034
	BRBCN$^+$	**22.39/0.5606**	**26.43/0.7395**	**29.57/0.8486**	**34.35/0.9335**	39.44/0.9731	**30.44/0.8111**
BSD68	ISTA-Net$^+$	19.18/0.4201	22.34/0.5573	25.30/0.7001	29.31/0.8507	34.01/0.9421	26.03/0.6941
	OPINE-Net$^+$	21.88/0.5162	25.16/0.6841	27.81/0.8040	31.50/0.9062	36.32/0.9658	28.53/0.7753
	CSNet$^+$	22.05/0.5180	24.91/0.6783	27.16/0.7990	30.53/0.9046	34.89/0.9637	27.91/0.7727
	AMP-Net	22.28/0.5376	25.20/0.6764	27.86/0.7929	31.72/0.9047	36.81/0.9679	28.77/0.7759
	MADUN	22.20/0.5306	25.22/0.6973	27.92/0.8200	31.74/0.9186	35.22/0.9633	28.46/0.7860
	DGUNet$^+$	22.13/0.5215	25.45/0.6986	28.13/0.8165	31.97/0.9158	37.04/0.9718	28.94/0.7848
	BRBCN	22.54/0.5267	25.59/0.7000	28.22/0.8178	32.16/0.9182	**37.24/0.9724**	29.15/0.7870
	BRBCN$^+$	**22.81/0.5431**	**25.74/0.7066**	**28.29/0.8206**	**32.23/0.9193**	37.19/**0.9724**	**29.25/0.7924**
Waterloo	ISTA-Net$^+$	18.94/0.4780	23.05/0.6351	27.17/0.7832	32.30/0.9044	37.80/0.9647	27.85/0.7531
	OPINE-Net$^+$	22.28/0.6065	26.60/0.7766	30.26/0.8756	34.87/0.9439	40.20/0.9790	30.84/0.8362
	CSNet$^+$	22.09/0.5853	25.87/0.7519	28.85/0.8605	32.65/0.9303	36.82/0.9668	29.26/0.8190
	AMP-Net	22.69/0.6196	26.64/0.7806	30.09/0.8775	34.79/0.9473	40.10/0.9807	30.86/0.8411
	MADUN	22.45/0.6184	26.54/0.7886	30.21/0.8870	34.96/0.9518	38.80/0.9769	30.49/0.8445
	DGUNet$^+$	22.58/0.6127	26.90/0.7878	30.67/0.8862	35.20/0.9499	40.70/0.9821	31.21/0.8437
	BRBCN	22.88/0.6180	27.35/0.7975	30.97/0.8911	**35.93/0.9549**	**41.63/0.9844**	31.75/0.8492
	BRBCN$^+$	**23.40/0.6433**	**27.47/0.8016**	**31.09/0.8933**	**35.93**/0.9548	41.42/0.9840	**31.86/0.8552**

Table 1: Comparison of average PSNR/SSIM results on gray-scale datasets Set14 and BSD68 and color dataset Waterloo. The best results are in bold while the second best results are marked with underline.

DGUNet$^+$(Mou, Wang, and Zhang 2022). The implementation codes are downloaded from the author's websites and run with the defaulting settings.

Quantitative analysis. Tab. 1 reports the comparison results on two gray-scale datasets. The results show that BRBCN and BRBCN$^+$ outperform all methods at each sampling ratio and gain both highest average PSNR and SSIM values. Our BRBCN performs best in every sampling ratio and is 0.36dB and 0.21dB higher than the second method as to the average PSNR value on grey-scale datasets Set14 and BSD68, respectively. Although BRBCN is block-based as ISTA-Net$^+$, the information of blocks' communication improves the quality of block reconstruction and successfully beats image-based methods including CSNet$^+$, MADUN$^+$, and MR-CCSNet$^+$. Although AMP-Net imported image-level filters as well, the mere local relation without global information can not provide sufficient blocks' communication, hence falling behind BRBCN. Moreover, BRBCN also performs best on the color dataset Waterloo, achieving 0.19dB, 0.45dB, 0.30dB, 0.73dB, and 0.93dB improvement over the second-best method at each sampling ratio with respect to the PSNR value. Furthermore, focusing on global communication while paying additional attention to the relation among adjacent blocks, BRBCN$^+$ achieves better performance especially at 0.01 sampling ratio, for example, 0.52dB, 0.53dB and 0.71dB higher PSNR values on Set14, BSD68 and Waterloo, respectively. An interesting finding is that BRBCN$^+$ performs weaker than BRBCN at 0.5 sampling ratio. This is because the additional IBR pursues the harmony among adjacent blocks and would hurt the fine-grained inner block details. That is the reason why we choose to conduct the block-based reconstruction, rather than choosing the image-based way to bypass the communication problem at the cost of losing fine-grained reconstruction within blocks.

Quality analysis. Figs. 3 shows the visualization results of our methods and state-of-the-art CSI methods on gray images and color images. As can be observed from Figs. 3, when the sapling ratio is low, ISTA-Net$^+$ suffers from blocking artifacts due to the lack of blocks' communication. Although BRBCN is a block-based method without local filtering like ISTA-Net$^+$, the blocks' communication helps relieve the block artifacts problem. Meanwhile, the image-based methods, like CSNet$^+$ and OPINE-Net$^+$, fail to reconstruct the fine details (e.g., words) without careful local concern. On the other hand, our BRBCN can well ha Besides, with the help of extra IBR, the interaction between adjacent blocks is further strengthened and BRBCN$^+$ has better performance, especially at low sampling ratios.

Ablation Study

In this section, we verify the validation of blocks' communication, the way of fusion, and the proposed modules, and discuss the hyperparameter setting and the inference time. To avoid the image-level IBR affecting the comparison, we conduct all the ablation experiments based on BRBCN.

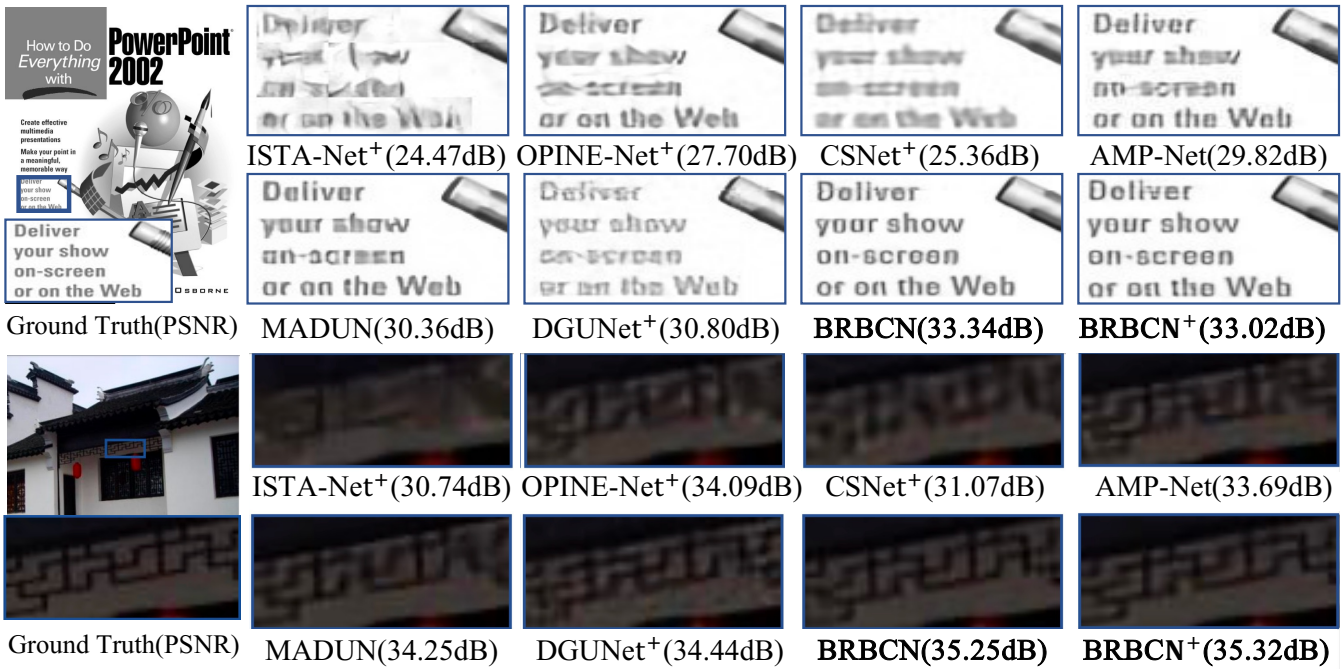

Figure 3: Visualization comparison on reconstructing images at 0.1 sampling ratio: "PPT3" image from Set14 dataset (upper) and an image from Waterloo dataset (lower) .

Global Info.	Set14		BSD68	
	0.01	0.1	0.01	0.1
w/o	20.67/0.4726	29.31/0.8414	21.46/0.4728	28.07/0.8137
w/	**21.87/0.5332**	**29.51/0.8460**	**22.54/0.5267**	**28.22/0.8178**

Table 2: Comparison of w/o and w/ global information. Average PSNR/SSIM at ratio=0.01,0.1 on Set14 and BSD68.

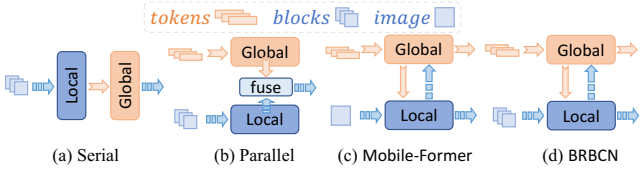

Figure 4: The visualization of different ways of fusion.

Fusion	Set14		BSD68	
	0.01	0.1	0.01	0.1
①	21.42/0.5073	28.66/0.8303	22.04/0.5104	27.70/0.8046
②	20.45/0.4563	22.71/0.6559	21.00/0.4697	22.63/0.6367
③	**21.56/0.5143**	28.13/0.8234	**22.23/0.5122**	27.47/0.8026
④	20.23/0.4435	26.51/0.8060	21.82/0.4833	26.80/0.7864
⑤	**21.87/0.5332**	**29.51/0.8460**	**22.54/0.5267**	**28.22/0.8178**

Table 3: The comparison of different fusion ways. ① fusion-free; ② serial; ③ parallel; ④ Mobile-Former; ⑤ BRBCN. Average PSNR/SSIM at ratio=0.01, 0.1 on Set14 and BSD68 datasets. The results better than fusion-free (CNN-only) are in bold.

Blocks' communication. Our BRBCN introduces the global information of to help the reconstruction of each block. To analyze the influence of the global information, we compare two reconstruction strategies: 1. The blocks are fed into the network one by one, with only one block in the net and no blocks' communication; 2. All the blocks are directly fed into the network, with blocks' communication happening. Tab. 2 shows that the reconstruction quality with global information is always better than without it. Especially when the very limited block measurements, like 0.01 sampling ratio, are not sufficient for block reconstruction, the more significant the improvement the communication brings.

Way of fusion. We argue that the way to fuse different representations is the key point of whether the local and global interaction is helpful. As shown in Fig. 4, we compare our proposed method with three different fusion ways, including two direct ways, in serial and in parallel, and Mobile-Former (Chen et al. 2022b) which is designed for the image-level fusion. As illustrated in Tab. 3, BRBCN performs best on all datasets. Simple as fusion in serial, it performs even worse than without fusion, since the representations of CNN and Transformer are in different feature spaces. Although Mobile-Former adds local and global interaction between two modules, it is originally designed to process the whole image with two architectures and hence fails to fuse the block representation with global representation. The improper fusion way would not do a favor but become a hindrance, that is the reason why Mobile-Former even performs worse than simply combining in parallel. On the other hand, the $L2G$ and $G2L$ modules in BRBCN suc-

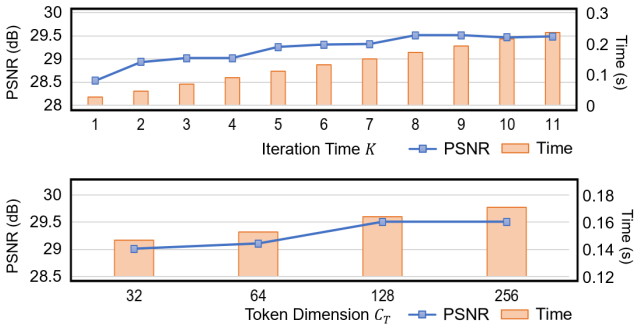

Figure 5: Average PSNR on Set14 at 0.1 sampling ratio. Upper: iteration time K. Lower: tokens dimension C_T.

Case	G2L	L2G	Set14		BSD68	
			0.01	0.1	0.01	0.1
	-	-	21.42	28.66	22.04	27.70
	✓	-	21.54	28.65	22.28	27.74
BRBCN	✓	✓	**21.87**	**29.51**	**22.54**	**28.22**

Table 4: Ablation study of the $G2L$ and $L2G$ modules. Average PSNR at ratio=0.01, 0.1 on Set14 and BSD68 datasets.

cessfully strengthen the interaction of two representations and improve the quality of reconstruction.

Hyperparameters analysis. Here we discuss the hyperparameters setting in BRBCN. The iteration time K is related to the reconstruction quality and inference time. As shown in Fig. 5, when $K \geq 8$, the reconstruction quality does not increase as the run speed increases. Thus, considering the tradeoff between model complexity and reconstruction, we set iteration time $K = 8$. Meanwhile, the token dimension C_T is related to the adequacy of inter-block communication. As shown in Fig. 5, the communication is insufficient when C_T is small and the loss curve is almost flat when $C_T \geq 128$, which means that $C_T = 128$ is sufficient for global communication.

Modules ablation. We conduct ablation studies on different modules to verify their effectiveness. Since our method focuses on introducing blocks' communication into block reconstruction, we regard the block-based reconstruction module IBR as the baseline and further verify $L2G$ and $G2L$ while IBC is involved in the interaction. As shown in Tab. 4, the one-way information passing from global tokens to local blocks with $G2L$ has finite improvement even side effects since the sole global communication, lacking inductive bias and pattern activation, is sufficient. As for the two-way interaction between local and global with $G2L$ and $L2G$, the local representation and global communication calculation are iteratively updated in a mutually promotional way.

Model complexity In this section, we compare the model size and inference time of different methods. The testing was implemented on an Nvidia RTX 3090 GPU at 0.01 sampling ratio. As shown in Tab. 5, although ISTA-Net$^+$ and CSNet$^+$ have the fastest running time and small model size, they perform worst due to the lack of blocks' communication or inner-block concern. Meanwhile, the local filter operation in MADUN and AMP-Net slows down the inference speed but improves the performance less than our method because it does not realize sufficient global communication. On the contrary, our BRBCN performs well with the help of sufficient interaction between local and global information at the cost of a small decrease in speed. The key factor influencing our model complexity is the maximum channel setting in IBR. Our model has high computational costs due to the big maximum channel setting. Using similar sizes as DGUNet$^+$, BRBCN_32 still achieves 0.32dB and 0.18db improvements on BSD68 and Waterloo respectively, which means it is the block-communication works not the parameter increment. When computing resources are rich, a bigger model BRBCN_256 could achieve better performance

Methods	Parameters (M)	PSNR (dB)/Running time (s)	
		BSD68	Waterloo
ISTA-Net	0.32	19.18/0.009	18.94/0.025
OPINE-Net	0.49	21.88/0.014	22.28/0.026
CSNet	0.41	22.05/0.010	22.09/0.075
AMP-Net	0.35	22.28/0.033	22.69/0.105
MADUN	2.89	22.20/0.102	22.45/0.472
DGUNet	6.51	22.13/0.041	22.58/0.260
BRBCN_32	5.34	22.45/0.056	22.76/0.220
BRBCN_64	8.11	22.52/0.064	22.82/0.407
BRBCN_256	43.35	22.54/0.159	22.88/0.736
BRBCN$^+$_32	7.95	22.80/0.060	23.25/0.249
BRBCN$^+$_64	10.72	22.81/0.065	23.27/0.252
BRBCN$^+$_256	45.96	22.81/0.161	23.40/0.747

Table 5: The comparison of different ways on parameter numbers (M), inference time (s) and PSNR (dB)

Conclusion

In this paper, we proposed BRBCN for block image CS which leverages the information interaction between local and global information to help the blocks reconstruction. BRBCN maximized the benefits of CNN's aptitude for local processing to reconstruct the blocks and harnessed the power of the Transformer's global relationship modeling to compute blocks' communication through a dual CNN-Transformer architecture design. Moreover, the proposed $L2G$ and $G2L$ modules successfully established a bi-directional interaction between local representation and global communication. In the experiments, we demonstrated how blocks' communication contributes to block reconstruction and we showcased the advantages of BRBCN via ablation experiments, particularly in scenarios with extremely low sampling ratios. In future work, the robustness against noisy inputs, as well as the generalization ability for multi-scale sampling and practicability for other applications like medical images need to be strengthened.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62106063, 62172126, in part by the Shenzhen Research Council

under Grant JCYJ20210324120202006, and in part by the Guangdong Natural Science Foundation under Grant 2022A1515010819.

References

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1–122.

Chen, Q.; Wu, Q.; Wang, J.; Hu, Q.; Hu, T.; Ding, E.; Cheng, J.; and Wang, J. 2022a. MixFormer: Mixing Features across Windows and Dimensions. In *Proc. of CVPR*, 5249–5259.

Chen, Y.; Dai, X.; Chen, D.; Liu, M.; Dong, X.; Yuan, L.; and Liu, Z. 2022b. Mobile-former: Bridging mobilenet and transformer. In *Proc. of CVPR*, 5270–5279.

Dong, W.; Shi, G.; Li, X.; Ma, Y.; and Huang, F. 2014. Compressive sensing via nonlocal low-rank regularization. *IEEE TIP*, 23(8): 3618–3632.

Donoho, D. L. 2006. Compressed sensing. *IEEE TIT*, 52(4): 1289–1306.

Donoho, D. L.; Maleki, A.; and Montanari, A. 2009. Message-passing algorithms for compressed sensing. *Proc. of NAS*, 106(45): 18914–18919.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proc. of ICLR*.

Duarte, M. F.; Davenport, M. A.; Takhar, D.; Laska, J. N.; Sun, T.; Kelly, K. F.; and Baraniuk, R. G. 2008. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2): 83–91.

Fan, Z.-E.; Lian, F.; and Quan, J.-N. 2022. Global Sensing and Measurements Reuse for Image Compressed Sensing. In *Proc. of CVPR*, 8954–8963.

Gan, L. 2007. Block compressed sensing of natural images. In *Proc. of ICDSP*, 403–406. IEEE.

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

Gregor, K.; and LeCun, Y. 2010. Learning fast approximations of sparse coding. In *Proc. of ICML*, 399–406.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proc. of CVPR*, 770–778.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. In *Proc. of NIPS*, 1097–1105.

Kulkarni, K.; Lohit, S.; Turaga, P.; Kerviche, R.; and Ashok, A. 2016. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proc. of CVPR*, 449–458.

Li, C.; Yin, W.; Jiang, H.; and Zhang, Y. 2013. An efficient augmented Lagrangian method with applications to total variation minimization. *COMPUT OPTIM APPL*, 56(3): 507–530.

Li, C.; Zhang, Y.; and Xie, E. Y. 2019. When an attacker meets a cipher-image in 2018: A year in review. *JISA*, 48: 102361.

Lin, T.; Ma, Z.; Li, F.; He, D.; Li, X.; Ding, E.; Wang, N.; Li, J.; and Gao, X. 2021. Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer. In *Proc. of CVPR*, 5141–5150.

Ma, K.; Duanmu, Z.; Wu, Q.; Wang, Z.; Yong, H.; Li, H.; and Zhang, L. 2016. Waterloo exploration database: New challenges for image quality assessment models. *IEEE TIP*, 26(2): 1004–1016.

Mehta, S.; and Rastegari, M. 2021. MobileViT: Lightweight, General-purpose, and Mobile-friendly Vision Transformer. In *Proc. of ICLR*.

Meng, Z.; Yu, Z.; Xu, K.; and Yuan, X. 2021. Self-supervised neural networks for spectral snapshot compressive imaging. In *Proc. of ICCV*, 2622–2631.

Metzler, C. A.; Maleki, A.; and Baraniuk, R. G. 2016. From denoising to compressed sensing. *IEEE TIT*, 62(9): 5117–5144.

Michailovich, O.; Rathi, Y.; and Dolui, S. 2011. Spatially regularized compressed sensing for high angular resolution diffusion imaging. *IEEE TMI*, 30(5): 1100–1115.

Mou, C.; Wang, Q.; and Zhang, J. 2022. Deep Generalized Unfolding Networks for Image Restoration. In *Proc. of CVPR*, 17399–17410.

Mousavi, A.; and Baraniuk, R. G. 2017. Learning to invert: Signal recovery via deep convolutional networks. In *Proc. of ICASSP*, 2272–2276. IEEE.

Mousavi, A.; Patel, A. B.; and Baraniuk, R. G. 2015. A deep learning approach to structured signal recovery. In *Proc. of Allerton*, 1336–1343. IEEE.

Sapiro, G. 2008. Sparse Representation for Color Image Restoration. *IEEE TIP*.

Shi, W.; Jiang, F.; Liu, S.; and Zhao, D. 2019. Image compressed sensing using convolutional neural network. *IEEE TIP*, 29: 375–388.

Song, J.; Chen, B.; and Zhang, J. 2021. Memory-Augmented Deep Unfolding Network for Compressive Sensing. In *Proc. of ACM MM*, 4249–4258.

Srinivas, A.; Lin, T.-Y.; Parmar, N.; Shlens, J.; Abbeel, P.; and Vaswani, A. 2021. Bottleneck transformers for visual recognition. In *Proc. of CVPR*, 16519–16529.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Proc. of NIPS*, 30.

Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; and Zhang, L. 2021. Cvt: Introducing convolutions to vision transformers. In *Proc. of ICCV*, 22–31.

Wu, Z.; Zhang, J.; and Mou, C. 2021. Dense Deep Unfolding Network With 3D-CNN Prior for Snapshot Compressive Imaging. In *Proc. of ICCV*, 4892–4901.

Yang, Y.; Sun, J.; Li, H.; and Xu, Z. 2018. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE TPAMI*, 42(3): 521–538.

Zeyde, R.; Elad, M.; and Protter, M. 2010. On single image scale-up using sparse-representations. In *International Conference on Curves and Surfaces*, 711–730. Springer.

Zhang, J.; and Ghanem, B. 2018. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proc. of CVPR*, 1828–1837.

Zhang, J.; Zhao, C.; and Gao, W. 2020. Optimization-inspired compact deep compressive sensing. *IEEE JSTSP*, 14(4): 765–774.

Zhang, J.; Zhao, D.; and Gao, W. 2014. Group-based sparse representation for image restoration. *IEEE TIP*, 23(8): 3336–3351.

Zhang, Z.; Liu, Y.; Liu, J.; Wen, F.; and Zhu, C. 2020. AMP-Net: Denoising-based deep unfolding for compressive image sensing. *IEEE TIP*, 30: 1487–1500.