wganBCS : Block-wise image compressive sensing and reconstruction model using WGAN-gp network to eliminate block effects

Chen, Boyan

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
17
(開始ページ / Start Page)
1
(終了ページ / End Page)
6
(発行年 / Year)
2022-03-24
(URL)
https://doi.org/10.15002/00025253

wganBCS: Block-wise image compressive sensing and reconstruction model using WGAN-gp network to eliminate block effects

Boyan Chen,
Graduate School of Computer and Information Sciences, Hosei University, Japan

***Abstract*—In the past decade, compressive sensing has been successfully applied to many image processing tasks based on an empirical observation that most image signals are sparse in a certain domain. Several image compressive sensing methods have been proposed these years including the block compressive sensing (BCS) which uses a relatively smaller sensing matrix to raster-scan the entire image to acquire the signal block by block, which can greatly reduce the memory consumption. However, BCS still suffers from two issues. One is that block-wised sensing causes heavy block effect on the reconstructed image, which leads to degradation in the image quality metrics. Besides, the sensing matrix usually needs to be handcrafted, which requires a huge amount of work. In this paper, we propose a WGAN-gp based image compressive sensing and reconstruction model using generative adversarial networks which have achieved a great success these years on dealing with computer vision tasks, e.g. super-resolution and deblurring. More precisely, we train a plain convolutional neural networks to perform the block-wised sampling and initial reconstruction. Then we use a generative adversarial network which provides fine reconstruction on the initial reconstructed images to eliminate the block effect to improve the final reconstruction quality. Experimental result shows that our model is superior both in visual quality and in the image quality metrics compared to the traditional image compressive sensing methods.**

***Index Terms*—image compressed sensing, generative adversarial networks, neural network**

I. Introduction

A. motivation

Compressive sensing proposed by Candes and Donoho [1] is a novel theory in the signal processing field. The key idea of compressive sensing is to recover a sparse signal from sub-Nyquist samplings by convex optimization. It suggests that most of the signals are sparse signals that we can use a $a*A$ sampling matrix to sub-Nyquist sample the target signal

$$s = \Phi o \tag{1}$$

where o is the $A*1$ original signal and s is the compressive sampling result of $a*1$ where $a << A$, notice that o has to be sparse and the sensing matrix Φ must meets restricted isometry property (RIP):

$$(1-\delta_k)\,\|c\|^2 \leqslant \|\Phi_T c\|^2 \leq (1+\delta_k))\,\|c\|^2 \tag{2}$$

let k denotes the sparsity of the signal o meaning that it only contains very few non-zero coefficients

$$\|o\|_0 \leqslant k \tag{3}$$

and the columns of the sensing matrix Φ is a finite collection of vectors, then δ_k is the k-restricted isometry constants [2].

Generally speaking, we can use a sampling signal which much lower than the Nyquist frequency to sample the signal meanwhile enables data hardware compression during the sampling process and exactly reconstruct the target signal by solving a reverse ill-posed problem:

$$r = \Phi^{-1} s$$

where r is the recovery from the s, and Φ^{-1} is the pseudo inverse of sampling matrix.

To solve this ill-posed problem, many non-linear algorithms have been developed which basically can be categorized into three kinds: one kind is convex optimization methods, which translate the non-convex problem into a convex one then replace the L0 norm constraint with L1 norm and solve a linear programming problem to get the approximate recovery result r [3]. Another kind is greedy algorithms including orthogonal matching pursuit (OMP), and Simultaneous Orthogonal Matching Pursuit (SOMP). For images processing, methods like Total Variation [4] also have been proposed. Before the Block Compressed Sensing (BCS) [5] was proposed by Lu Gan et al, most image compressive sensing algorithms were not practiced in real word application, since the sensing matrix Φ is a $a*A$ matrix thus we need to flatten the whole input image into a long A vector which causes a huge memory consumption.

The third kind is Deep Learning based methods, that have come to be widely used for image reconstruction tasks in the last few years. Stacked denoising autoencode (SDA) [6] proposed by Mousavi et al was the first deep learning based network specially designed for reconstructing sampled images. Wuzhen Shi et al in their work [7] successfully establish a relationship between the Deep learning and the compressive sensing. Compared to those non-Deep learning iteration reconstruction algorithm their work not only achieved good reconstruction quality but also a way faster speed, and since the sensing matrix can be automatically learned during the training process, it no longer needs to be handcrafted.

These years, many Generative Adversarial Network (GAN)

* Supervisor: Prof. Kaoru Uchida

[8] based models have been proposed and achieved superior performance in image processing tasks such as SRGAN for super-resolution image reconstruction and DeblurGAN [9] for image helpful way to deal with the block-effects and improve the reconstruction quality.

B. proposed architecture

In this paper, we propose a new block-based image compressive sensing model called wganBCS which uses Wasserstein GAN to solve the block-effects problems. Fig. 1 shows the overview architecture of our model; wganBCS is composed of four parts. First, a BCS layer L_{BCS} and the initial networks N_I which performs the block-wised sensing on the original input images D_T and initial reconstruct the sampled results into initial reconstructed images D_L. Then a multi-layer convolutional neural network N_G receives the D_L which are suffered from heavy block effects and outputs the fine reconstructed images marked D_H. Finally, both original images D_T and fine reconstructed images D_H are sent into the discriminator N_D as a pair to calculate a distance loss between the real ground truth distribution P_t and generated distribution P_f. Experimental result shows that our model is superior both in visual quality and the image quality metrics compared to traditional compressive sensing methods.

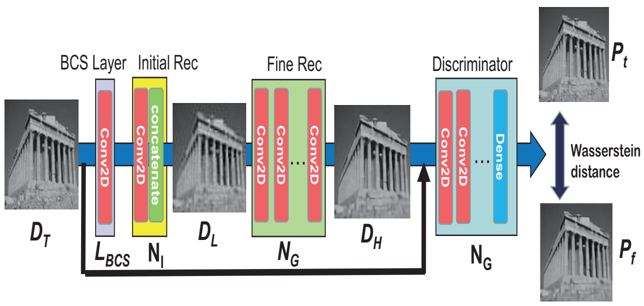

Fig. 1: Overview architecture of wganBCS. wganBCS is composed of four parts: a BCS sensing layer L_{BCS}, a initial reconstruction network N_I, a fine reconstruction network N_G and a discriminator network N_D.

C. main contribution

In short, the main contributions of our work are summarized in three aspects:

- We explore a relationship between BCS sampling paradigm and the GAN networks.
- We propose an end to end image compressive sensing and reconstruction model with generative adversarial network, and prove that it is superior both in reconstruction time and the image quality metrics compared to traditional compressive sensing methods.
- We show that the trained generator network can be applied to other block-wise image compressive sensing to eliminate block effects.

The rest of this paper is organized as follows: In Section II, first we review image compressive sensing methods. Section III presents the details of our proposed model. In Section IV, we give the detailed parameters for the training. Experimental results are given in Section V. In Section VI, based on the experimental results we discuss the advantages of our model and also the relationship between image entropy and reconstruction quality. Finally, in Section VII we summarize the paper.

II. Related Work

In this Section, we briefly review the block-based compressive sensing paradigm; Then we briefly introduce the Wasserstein GAN (WGAN).

A. Block-based CS network

Block-wised compressive sensing was proposed by Gan Lu et al. in [5]. It generally consists of two stages, block-wise sampling and iterative reconstruction. First the image would be divided into $B \times B$ blocks and use a sensing matrix S to perform the matrix multiplication with each block in a raster scanning manner. Let y_i denotes the sampled result of the ith block x_i. This procedure can be described as Eq.4

$$y_i = Sx_i \tag{4}$$

The second stage is to reconstruct the image X from y_i through a iterative algorithm. In [5] a two-stage iterative hard thresholding linear estimation algorithm has been proposed. To be specific, to reconstruct natural image X by solving Eq.5 through minimum mean square error (MMSE)

$$\hat{x}_i = \hat{S}y_i \tag{5}$$

where $\hat{S}$ is

$$\hat{S} = R_{xx}S^T(SR_{xx}S^T)^{-1} \tag{6}$$

R_{xx} denotes a autoregressive model AR(1).

B. Generative Adversarial Network

WGAN proposed by Arjovsky et al [10] try to solve this problem by replace the Jensen-Shannon divergence with Wasserstein distance, specifically speaking WGAN remove the log from the original GAN loss function and add a weight clip letting the discriminator to fit the Lipschitz function

$$\|f(x_1) - f(x_2)\| \le k\|x_1 - x_2\| \tag{7}$$

where $f(x)$ denote a neural network and when k=1 it is 1-Lipschitz. The target function of WGAN is

$$\max_{Dis} E_{X\sim P_t}[Dis(X)] - E_{X_r\sim P_f}[Dis(Gen(X_r))] \tag{8}$$

Based on WGAN, WGAN-gp proposed by Gulrajani et al. [11] made an further improvement on the training stability by using gradient penalty instead of weight clipping. Experiment result [11] shows that WGAN-gp is more faster than WGAN in model training speed and avoids gradient binaryzation, gradient vanishing and gradient exploding which are caused by weight clipping.

III. Proposed wganBCS

In this section, we propose an end to end fully data-driven block-based image compressive sensing and reconstruction model called wganBCS, which use a WGAN-gp network to eliminate the block effects that are caused by the BCS sensing. Our model consists of mainly four parts: a BCS layer and a initial reconstruction network, a fine reconstruction network, and a discriminator network. We put the BCS sensing and the initial reconstruction into a same sub-network and share a same loss function. And the rest sub networks are trained as a GAN network. More details are given in the rest sections of this section.

A. BCS layer and Initial Reconstruction Network

We use a convolutional layer to imitate the block-wise sensing same as Wuzhen Shi did in their work [7]. Since we are sensing the grayscale image, the size of kernel H of this layer is set to 32×32, the strides s of this kernel are also set to 32×32. The mechanism of the convolution lets the kernel scan though the image, for each calculation a filter outputs a value y_{ij}, i denotes the ith filter and j denoted the times of the calculation. This layer outputs a $102 \times 4 \times 4$ feature map (assuming the input image is 128×128), which represents the block-wise sampling results. Same as in [7] we use another convolutional layer to perform the initial reconstruction which outputs a $32 \times 32 \times 4 \times 4$ feature map Y. Finally, a custom layer reshapes and concatenates blocks of Y back to image. We train

(a) (b) (c)

Fig. 2: The ground truth input image (a) and its initial reconstruction result (b), which suffers from heavy block effects. (c) is the result after fine reconstruction that block effect has been reduced. The shape of input image is 128×128 and the sampling rate is 10%. We can also find that the leopard print in (c) is much more clear than it in (b).

the BCS and the initial reconstruction layer as a standalone end to end sub-network and the loss function is Mean Square Error (MSE). Therefore for an input image X_i, the target of optimizer is to minimize

$$\frac{1}{m}\sum_{i=1}^{m}(X_i - Reshape(InitialRecon(BCS(X_i))))^2 \tag{9}$$

where i denotes the ith image in a m size batch.

B. Fine Reconstruction Network

The initial reconstruction results are suffered from heavy block effect, to eliminate the block effect we propose a fine

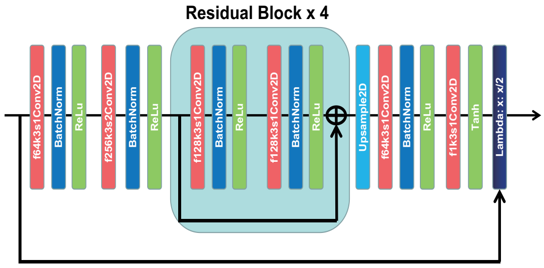

Fig. 3: Architecture of fine reconstruction network. This figure shows the structure of the fine reconstruction network N_G, which works as the generator in the adversarial training. The basic structure is from Unet, we cut off some layers in order to make the training easier. For the baseline model we use four residual blocks and add a skip connection between the input and the output.

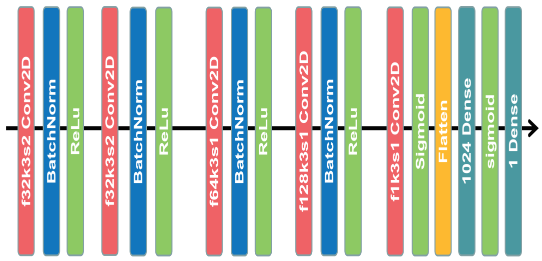

Fig. 4: Architecture of discriminator network. This figure shows the architecture of discriminator, the output of the last layer are not binary labels as original GAN did, instead it outputs a tensor vector called Wasserstein distance which measuring the divergence between the ground truth images and generated images.

reconstruction network. The architecture of this fine reconstruction network are shown in Fig. 3. This network is variant from the famous U-net [12] which is widely used in image segmentation tasks. To keep the model from collapsing, we also train the fine reconstruction network with the MSE loss before the adversarial training at each epoch, then the loss function is Eq.10

$$\frac{1}{m}\sum_{i=1}^{m}(X_i - FineRecon(X_L)^2 \tag{10}$$

where i denotes the ith image in a m size batch. Notice the BCS sensing layer and the initial reconstruction network are trained together but separated from the fine reconstruction network, which is different with the CSnet.

C. Discriminator

Fine reconstruction result X_H and the ground truth image X are sent as a pair into the discriminator. We use WGAN-gp [11] as the basic framework for the adversarial training; Compared to the original Wasserstein GAN (WGAN)[10], WGAN-gp made an improvement on the training stability by adding the gradient penalty. First we calculate an interpolation image X_{inter} between the X and X_H following the Eq.11

$$X_{inter} = X + \alpha \times (X_H - X) \tag{11}$$

where α is a random number follows a standard normal distribution. And the discriminator loss with gradient penalty shall be

$$Dis(X_{inter}) - Dis(X) + \lambda(\|\nabla_{X_{inter}} Dis(X_{inter})\|_2 - 1)^2 \tag{12}$$

where λ denotes the weight of gradient penalty, by default it is set to 10. Fig. 4 shows the structure of our discriminator network. Compared to the original GAN, the discriminator in WGAN outputs a Wasserstein distance between the generated input and the ground truth instead of a binary result, therefore there is no more loss function like binary cross entropy at the last dense layer.

Compared to the former works which did not incorporate WGAN network, using Wasserstein distance instead of only mean square error (MSE) loss function helps the model to reconsturct images better both in visual quality and quantitative metrics. Also gradient penalty mechanism helps the generator to learn in a much smoother way.

IV. Model Training

In this section, first we give the details about the selection of training data, pre-processing augmentations and some parameters we set for training.

A. Training Data and Parameters Setting

Unlike in other works which uses BDS500 [13] as training dataset, we train our model on the ImageNet2012_Test database which contains 100000 different natural images, since it already has enough images we did not use any data augmentation methods, instead we just resize each image into 128×128 and then normalize them into [-1,1]. To reach the maximum training efficiency meanwhile reduce the memory usage we set the batchsize to 32 and train our model for 80 epochs on a Nvidia GTX1060 graphic card. We choose Adam as the common optimizer in this work [11] and use callback functions to periodically reduce the learning rate during the training.

V. Experimental results

In this section, we introduce the metrics we used for the testing. Then we show the comparison results between our model and other former works.

A. Metrics and Testing Datasets

For the testing metrics, we use Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) which are a commonly used standards in image reconstruction tasks. we test our model on SET11 [14] which contains 11 different 256×256 grayscale images, except $fingerprint$ and $flintstones$ are 512×512. Notice that the testing dataset is separate from the training dataset. Since the kernel size we used in BCS network is 32×32 without padding, which means we can not directly send images that can not be divided exactly by 32. Therefore we zero-padding the images, and after reconstruction we de-padding them before the testing.

B. Comparison with DL-based method

We compared our proposed model with the ISTAnet [15], CSnet and AMPnet [16] which are all DL-based. Codes of ISTAnet and AMPnet are open sourced and we use the well-trained model which are provided by the authors to perform the experiment, and we reimplement CSnet following the parameters setting and training methods in [7] since it is not open sourced. We choose AMPnet-2 from the [16] since it is the base-line model in their work. Detailed results are shown in Table.I. Fig. 6 shows the comparison of image parrot from SET11 at different sampling rates. Comparison results shows that on SET11 our wganBCS are better than ISTAnet, CSnet and the AMPnet-2 as shown in Table.I. On average PSNR of SET11 we improved around 2.53 dB, 0.92 dB and 0.501 dB respectively to ISTAnet, CSnet and the AMPnet-2. The visual quality comparison results are given in Fig. 5. ISTAnet has the wrose reconstruction quality which only got 23.66 dB on image *cameraman*, then followed by CSnet which got 25.729 dB. Our wganBCS reached 26.463 dB which are the best among these models. Fig. 6 shows the reconstructed image at different sampling rates, the higher sampling rate were used in the BCS sensing stage, the better reconstruction results can be obtained. At the 0.01 sampling rate, we can see that the details of the image *parrot* is totally messed up; At the 0.04 sampling rate, it is obviously much clearer but the strips around the bird head are still blurred. When increased the sampling rate into 0.1, most detail of the image can be correctly restored, however fine detail like the texture of the feather is still kind of blurred. When using higher sampling rate like 0.25 and 0.3, the visual quality is good enough compared to the original image.

VI. Discussions

In this section, based on the experimental results given in Section VI we discuss the advantages of our model and also discuss the relationship between the image entropy and the reconstruction quality of our model. As we mentioned in Section I our model imports WGAN-gp into the BCS sampling paradigm, Figure.7 shows the final reconstruction quality with and without our WGAN based fine reconstruction network. At 10% sampling rate we improved 2.13 dB using the fine reconstruction network we proposed; Compared with the ISTAnet [15], CSnet and AMPnet [16] we improved 2.532 dB, 0.9 dB and 0.5 dB respectively. Image entropy are used as a measure of image information content, which can be interpreted as the average uncertainty of information source. The first dimension image entropy can be described as Eq.13

$$E = -\sum_{i=0}^{255} p_i \log p_i \tag{13}$$

where p_i denotes the probability of a pixel being i gray level. We implement the code with NumPy to calculate histogram then return the entropy of each image. The reuslts of SET11 are shown in Fig.8. X-axis denotes the entropy of images from SET11 which are from low to high, and Y-axis suggests the PSNR. Generally, image with higher entropy and entropy per pixel can get better reconstruction quality. Compared to other

Fig. 5: Comparison of reconstruction quality on image cameraman from SET11 at sampling rate 10%. ISTAnet has the worst reconstruction quality among four networks. Our model shows superiority on reconstructing edges of objects, notice there are many strips and blurry around tripod and the background building in images reconstructed from CSnet and AMPnet-2.

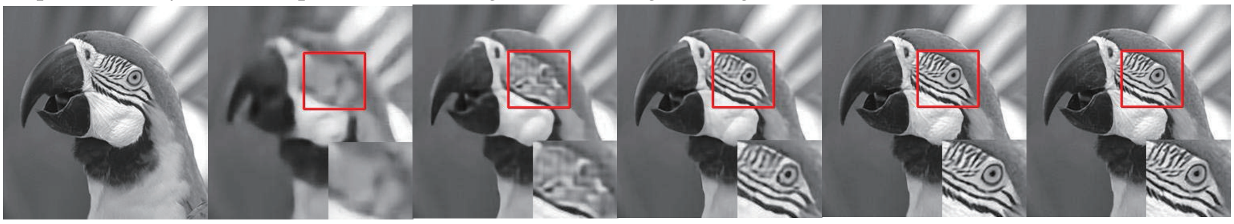

Fig. 6: Comparison of reconstruction quality on image parrot from SET11 at different sampling rates. Basically the higher sampling rate were used in the BCS sensing stage, the better reconstruction results can be obtained. When sampling rate reaches 30% the visual quality is good enough compared to the original image.

TABLE I: Comparison on SET11, On average PSNR of SET11 we improve around 2.53 dB, 0.92 dB and 0.501 dB respectively to ISTAnet, CSnet and the AMPnet-2.

Sample Rate	ISTA-Net	CSnet	AMP_Net_2_BM	AMP_Net_4_BM	wganBCS
0.01	0.431 17.42	0.552 20.943	0.554 20.412	0.556 20.358	0.567 21.215
0.04	0.612 21.32	0.745 24.800	0.756 24.724	0.768 25.070	0.763 25.304
0.1	0.805 26.64	0.857 28.270	0.865 28.671	0.873 29.045	0.874 29.172
0.25	0.921 32.59	0.935 32.947	0.942 33.917	0.945 34.380	0.957 34.053
0.3	0.926 33.05	0.948 34.177	0.953 35.121	0.956 35.665	0.955 35.136

TABLE II: Entropy of images from SET11. This table shows the entropy of each image in SET11; Image entropy are used as a measure of image information content, which can be interpreted as the average uncertainty of an information source.

barbara2	boats	cameraman	fingerprint
7.525	7.146	7.01	6.728
flintstones	foreman2	house	lena256
6.579	7.008	6.493	7.444
Monarch	Parrots	peppers256	
7.472	7.401	7.533	

methods, our wganBCS can reach a higher score on most entropy and had a close result with AMPnet-2 on 6.493, 7.444 and 7.472.

VII. Conclusions

In this paper, we proposed a block-based image compressive sensing and reconstruction model which we called wganBCS, it uses a generative adversarial network to eliminate block effects caused by the block-wise sampling. Our model mainly consist of four parts, a BCS sampling layer, a initial reconstruction network, a fine reconstruction network and a discriminator network. the network we used for the fine reconstruction is a simplified U-net and based on the WGAN-gp framework, we let the fine reconstruction network to work as the generator and use a discriminator network to form as a GAN network. After training, we compare with some former works on the SET11 and result shows that our model is better in SSIM and PSNR than most former works on natural images.

In the future study, we would like to add the attention mechanism like transformers or CBAM into the fine reconstrucion network. Further more, we also considered to use the conditional GAN mechanism to see if it can make the model training more stable. Last we would like to know if our model can be used on other compressive sensing tasks such as MRI image reconstruction.

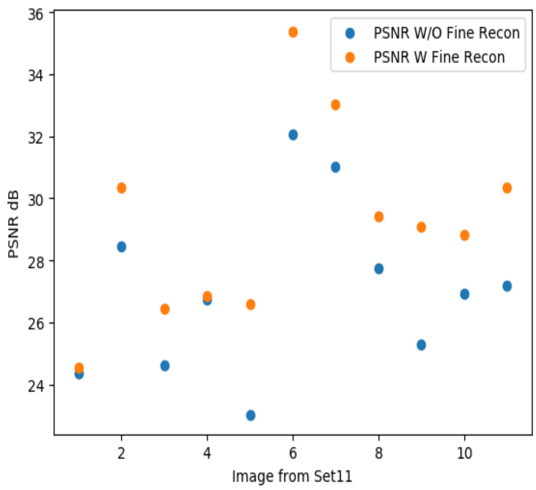

Fig. 7: Comparison of PSNR on SET11 at 10% sampling rate W and W/O fine reconstruction network. Our WGAN-gp based fine reconstruction network can reduce the block effect and improve the reconstruction quality; The average PSNR W/O and W fine recon is 27.04 dB and 29.17 dB. We improved 2.13 dB by our WGAN-gp based fine reconstruction network.

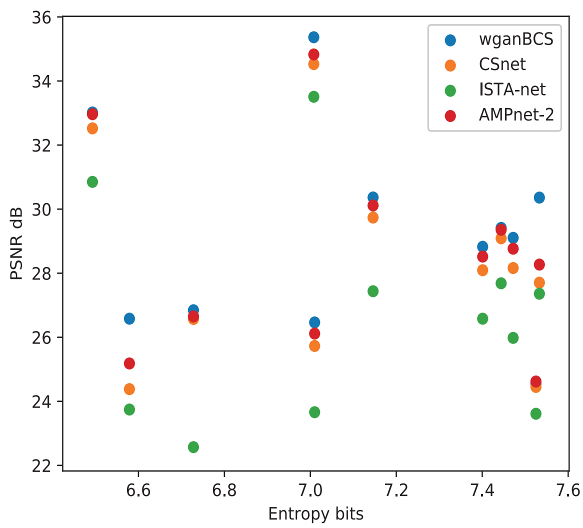

Fig. 8: PSNR, SSIM, Entropy and Entropy per pixel of each image from SET11 at sampling rate 10%. X-axis denotes the entropy of images from SET11 which are from low to high, and Y-axis suggests the PSNR. Generally, image with lower entropy and entropy per pixel can obtain better reconstruction quality. Compared to other methods, our wganBCS can reach a higher score at most entropy and had a close result with AMPnet-2.

References

[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[2] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[4] C. Li, W. Yin, and Y. Zhang, "User's guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.

[5] L. Gan, "Block compressed sensing of natural images," in *2007 15th International conference on digital signal processing*. IEEE, 2007, pp. 403–406.

[6] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," *IEEE*, 2016.

[7] W. Shi, F. Jiang, S. Zhang, and D. Zhao, "Deep networks for compressed image sensing," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 877–882.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[9] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[10] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," 2017.

[11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Springer, Cham*, 2015.

[13] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.

[14] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 449–458.

[15] J. Zhang and B. Ghanem, "Ista-net: Iterative shrinkage-thresholding algorithm inspired deep network for image compressive sensing," 2017.

[16] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "Amp-net: Denoising based deep unfolding for compressive image sensing," 2020.