Adaptive Sampling for Image Compressed Sensing Based on Deep Learning

Liqun Zhong\(^a\), Shuai Wan\(^b\) and Leyi Xie

School of Electronics and Information, Northwestern Polytechnical University, Xi’an, China.

E-mail: \(^a\) zhongliqun0314@mail.nwpu.edu.cn, \(^b\) swan@nwpu.edu.cn

Abstract. The compressed sensing (CS) theory has been applied to image compression successfully as most image signals are sparse in a certain domain. In this paper, we focus on how to improve the sampling efficiency for network-based image compressed sensing by using our proposed adaptive sampling algorithm. We conduct content adaptive sampling to achieve a significant improvement. Experiments results indicate that our proposed framework outperforms the state-of-the-arts both in subjective and objective quality. An average of 1-6 dB improvement in peak signal to noise ratio (PSNR) is observed. Moreover, the proposed work reconstructs images with more details and less image blocking effects, leading to apparent visual improvement.

1. Introduction
Since the advent of compressed sensing [1], [2], [3], many breakthroughs have been made in this filed. Compressed sensing makes full use of the characteristics of signal sparsity and combines signal sampling and compression together, which is one of the main reason why it is different from traditional signal compression. Regarding images, block-based CS (BCS) approaches are more feasible, as proposed in [4] which divides an image into small patches. Based on BCS, various efforts have been made, like [5] which combines the BCS with smoothed projected landweber reconstruction (BCS-SPL). In addition to non-adaptive CS, there are some other ways trying to conduct adaptive sampling allocation based on image content to make an improvement. Futhermore, [6] conducts compressed sensing by learning a gaussian mixture model from measurements.

In the research of CS, another method is based on deep neural network, which is introduced in literature. The key advantage of such method is that it allows simultaneous optimization of linear sampling and non-linear reconstruction operations in the training process. Therefore, the recovery quality and time complexity of CS based on network far outperform over those of traditional algorithms. Generally speaking, its performance is still limited, especially at very low sampling rates. For example, obvious blocking artifacts and some loss of image details are always observed when sampling rate is relatively low, seriously affecting subjective quality.

In non-adaptive CS, each image patch is generally assigned with the same number of samples. In practice, the human eye’s sensitivity to distortion varies from patches to patches with various contents. Inspired by this, This paper proposes an adaptive sampling allocation mechanism for BCS based on neural deep learning this paper proposes a way of content-adaptive image compressed sensing using deep learning, which can adaptively allocate appropriate sampling number for different image patches.
The rest of this paper is organized as follows. Section 2 provides a brief introduction on related work. Some adaptive sampling allocation methods are described in section 3. Our proposed algorithm is given in section 4. Experimental results will be exhibited in section 5. This paper concludes with a summary.

2. Related work

2.1 Block-based Compressed Sensing

In our work, we conduct research based on block-based compressed sensing (BCS). In BCS, an image is divided into overlapping patches of size $B \times B$. Each patch is sampled separately. Suppose that $x_j \in R^{B^2}$ is the vector representation of patch $j$, through a raster-scan fashion. Since $x_j$ can be expressed as $x_j = \Psi \theta_j$, in which $\theta_j$ is nearly sparse, the corresponding measurement $y_j$ can be expressed as:

$$ y_j = \Phi_B x_j = \Phi_B \Psi \theta_j = A \theta_j, $$

(1)

here, $\Phi_B$ is an $M_B \times B^2$ orthonormal measurement matrix, and sampling rate $R = M_B B^{-2}$. $\Psi$ is a certain transformation matrix, and $A$ is called sensing matrix.

Compressed sensing theory indicates that in order to recovery $x_j$ exactly, the sensing matrix $A$ needs to satisfy the restricted isometry property (RIP) [7]. Sparsity of signals can be described by $L_0$ norm, whereas it turns out to be a NP hard problem. Regarding the situation that $L_1$ norm is equivalent to the $L_0$ norm under certain conditions, the $L_1$ norm can usually be used instead. Mathematically, CS reconstruction is equivalent to solving the optimization problem:

$$ \min ||\theta_j||_1 \quad \text{s.t.} \quad A \theta_j = y_j, $$

(2)

2.2 Compressed sensing based on neural network

References [8], [9], [10] have presented some deep neural networks-based solutions to the problem of CS image sampling and reconstruction, in which fully-connected or convolution network (CNN) are utilized to conduct CS sampling and reconstruction. The general CS structure based on network can be illustrated in Figure 1:

![Figure 1. Compressed sensing framework based on deep neural networks.](image)

The first layer in Figure 1, as a sampling layer, projects the input image patch of size $B \times B$ into a vector of dimension $RB^2$, which is also called measurements. The second and third layer followed by an activation function RELU [11] are mainly used to restore the image structure and image details respectively. CNN layers and fully-connected layers or other layers are available ways in this part. The last layer is the output layer, which outputs the reconstructed image patch of size $B \times B$. Considering the excellent performance of a network composed of four fully-connected network, we decide to adopt it in our following experiment.

3. Adaptive sampling for BCS

In order to investigate the relationship between sampling rates and image quality in BCS, some available indicators should be found to represent content of an image. In reference [12], it states that the entropy based on transformed image patch can be valid to evaluate the information contained. Generally, the higher the entropy is, the more corresponding patch information will be, so more samples should be used for reconstruction. In addition to information entropy, characteristics such as the number of significant coefficients of DCT and variation of DCT coefficients are both available to
measure the amount of content in an image. The results indicate that such method does improve constructed quality. However, the allocation is still too average, so that the image patches containing various contents cannot be distinguished to a large extent.

This problem above is upgraded in reference [13]. The authors propose a new content-adaptive sample rate allocation algorithm (CASRA) for network-based compressed sensing. This algorithm mainly employs posterior information of image patches, namely the reconstructed mean square error (MSE) curve, as the characteristic parameter. The slower the MSE declines, the smoother the corresponding image patch is, where human eyes are generally less sensitive to. Accordingly, few samples will be assigned to smooth areas and areas with more complex texture will be allocated with more samples, which are in line with human visual characteristics.

Experimental results demonstrate that the allocation through this method is more reasonable and less uniform. However, since this algorithm is completely based on posterior information, this method is relatively time-consuming. The time consumption of this algorithm is given in Table 1.

| rate | Lena  | Barbara | Peppers | Mandrill | Goldhill | Cameraman |
|------|-------|---------|---------|----------|----------|-----------|
| 0.05 | 21.15 | 21.75   | 21.61   | 21.88    | 21.97    | 22.02     |
| 0.1  | 36.44 | 36.17   | 37.26   | 37.49    | 37.24    | 37.20     |
| 0.2  | 66.09 | 68.14   | 67.46   | 67.48    | 67.84    | 67.57     |

4. Proposed work

Although the CASRA based on MSE descent curve has an excellent allocation effect, on the other hand, it has a high complexity and takes a lot of time. Therefore, we need to find a faster and more concise way to analyze image content, so as to make up for the defects of this algorithm.

4.1 Saliency map

Apart from prior information based on DCT coefficients, MSE curves mentioned above. We also consider to distinguish different patches to find focused ones through saliency map. In the pixel domain, the higher the saliency value is, the higher the significance of the corresponding pixel is, and the easier it is for human eyes to focus on the corresponding region. Therefore, it is reasonable to consider the pixel saliency as the allocation factor. This paper mainly introduces a paper published in 2007 [14], which proposes a simple calculation model for visual significance. This model is widely used because of its simple computation and low time complexity.

The core ideas of this paper include: from the perspective of information theory, information can be divided into redundant parts and variable parts. People’s vision is more sensitive to changing parts. A fundamental principle of the visual system is to suppress the response to frequently occurring features while remaining sensitive to unconventional features. Then the image is divided into the following two parts:

\[ H(\text{Image}) = H(\text{Innovation}) + H(\text{Prior Knowledge}), \]

in which, \( H(\text{Innovation}) \) denotes the novelty part, and \( H(\text{Prior Knowledge}) \) represents the redundant information that should be suppressed by a coding system. It is found that the amplitude \( A(f) \) of the averaged Fourier spectrum of images usually satisfy a certain rule shown in equation (4). So the spectral residual theories have been proposed to obtain a higher efficient saliency map.

\[ E[A(f)] \propto 1/f, \]

4.2 Sample Rate Allocation

In this section, on the basis of CASRA, we further add saliency map to instruct the allocation of sample numbers, in order to appropriately reduce allocation complexity and time consumption. Given a total number \( N \) of an image, the whole proposed framework of the measurement allocation algorithm is presented in Figure 2:
Saliency-based allocation CASRA Image Final SAM Initial SAM

**Figure 2.** Our proposed allocation framework.

Here SAM indicates the sampling allocation map, and CASRA represents the content-adaptive sample rate allocation algorithm. Our sampling allocation process is composed of two parts: in the first part, we allocate a certain number of samples according to the characteristics of saliency map to obtain an initial SAM. Then, based on the initial SAM, CASRA is utilized to further allocate the remaining samples until all the samples are allocated. This ensures that the given number of samples can be used as fully as possible.

Then how to allocate the sampling numbers by virtue of saliency map? And how many samples need to be allocated to form the initial SAM? Here we define a weight factor of a patch \( x_i \) related to saliency map as follows:

\[
weight_i = \sum_{h=1}^{H} \sum_{w=1}^{W} (s_{h, w})^p \left( \sum_{h=1}^{H} \sum_{w=1}^{W} (s_{h, w})^p \right)^{-1}
\]

where \( h, w \) respectively represents horizontal and vertical position of \( x_i \) and the whole image. \( p \) is the exponential (in the following experiment, we set \( p=1 \)). \( s_{h, w} \) denotes corresponding saliency value. It can be seen that the weight factor is completely determined by saliency value. Along with CASRA, our proposed content adaptive allocation algorithm primarily consists of two components, which is represented as follows:

1. **Initial allocation:** Given the total sampling number \( N \) of a whole image and allocate \( n_i \) for the patch \( x_i \) as:
   
   **Step-1:** \( n_i = \lfloor \alpha \cdot N \cdot weight_i \rfloor \).
   
   **Step-2:** Then \( n_i \) is quantized in multiple of 5. Obtain the initial SAM. The first phase is completed.

2. **CASRA:** Given the initial SAM, continue allocating.
   
   **Step-1:** Find the best patch with the highest priority and add 5 samples to it. Then update the SAM.
   
   **Step-2:** If the sum of the SAM is smaller than the total number of samples \( N \), go through step-1 to continue assigning samples. Otherwise, go to step-3.
   
   **Step-3:** The allocation is completed, the current SAM is the final allocation.

   Here, \( \alpha (0 \leq \alpha \leq 1) \) is the proportion of \( N \) for initial allocation. When \( \alpha = 0 \), our proposed algorithm will degrade to the original CASRA. Otherwise, when \( \alpha = 1 \), our proposed algorithm will be utterly relevant to saliency value.

**Figure 3.** Mandrill and its SAM.

Because the saliency map can be obtained directly from original image and takes very little time. On the contrary, CASRA in the second phase needs to refer to reconstruction information, so it’s quite
time-consuming. That is, the major time consuming part of our algorithm is in the second stage. Figure 3 illustrates our sampling allocation for Mandrill.

5. Experimental results
We refer to the network-based CS. In our experiment, we train 51 networks, corresponding to 5, 10, 15, …, 225 samples. Moreover, These networks are composed of four fully-connected layers, one for sampling and the other three for reconstruction. We randomly select 20,000 images from the LabelMe dataset [15] and generate 5,000,000 patches of size 16×16 as the training set.

This section provides the testing results of six 512 × 512 images, i.e. Lena, Barbara, Peppers, et. al. Images are divided into non-overlapping patches of size 16×16, and then in-dependently sampled and reconstructed by BCS-SPL-DDWT, network-based CS, entropy based allocation, variation based allocation and our proposed allocation algorithm respectively. Table 2 tabulates the PSNR and results in the range of R=0.05, 0.1, 0.2. Here we set $\alpha = 0.5$. Note that all allocation algorithms are sampled and reconstructed based on the same network mentioned above.

In addition, we also record the time consumption of the allocation algorithm at different sampling rates in several sets of images. Compared with CASRA, the results show that our allocation algorithm can effectively reduce the time consumption at the coding side when the recovery quality is not greatly affected. It turns out to be an efficient attempt. This can be seen in Figure 4.

Table 2. PSNR for reconstructed image, in which rate is sampling rate (displayed as PSNR (dB)).

| Rate | Method       | Lena | Barbara | Peppers | Mandrill | Goldhill | Cameraman | Mean  |
|------|--------------|------|---------|---------|----------|----------|-----------|-------|
| 0.05 | BCS-SPL-DDWT | 24.62| 21.42   | 24.68   | 19.67    | 24.98    | 22.81     | 23.03 |
|      | Network-based| 29.74| 23.77   | 30.79   | 21.21    | 28.72    | 29.4      | 27.27 |
|      | Entropy-based| 30.2 | 23.63   | 30.67   | 21.1     | 28.54    | 29.96     | 27.35 |
|      | Variation-based| 31.37| 23.36   | 31.45   | 21.2     | 28.77    | 31.81     | 27.99 |
|      | Proposed     | 32.66| 24.79   | 33.29   | 21.70    | 29.74    | 33.73     | 29.32 |
| 0.1  | BCS-SPL-DDWT | 27.49| 22.61   | 28.39   | 20.53    | 26.76    | 25.48     | 25.21 |
|      | Network-based| 31.63| 24.24   | 33.49   | 22.04    | 30.28    | 32.11     | 28.97 |
|      | Entropy-based| 34.02| 24.53   | 34.33   | 22.34    | 30.94    | 34.71     | 30.15 |
|      | Variation-based| 34.87| 24.67   | 35.19   | 22.62    | 31.17    | 36.95     | 30.91 |
|      | Proposed     | 36.30| 27.26   | 36.47   | 23.25    | 32.10    | 38.59     | 32.33 |
| 0.2  | BCS-SPL-DDWT | 31.08| 23.81   | 32.55   | 21.75    | 29.02    | 29.45     | 27.94 |
|      | Network-based| 34.97| 25.11   | 36.65   | 23.89    | 32.93    | 36.8      | 31.73 |
|      | Entropy-based| 38.34| 26.16   | 37.92   | 24.29    | 33.73    | 40.74     | 33.53 |
|      | Variation-based| 39.81| 28.3    | 39.44   | 24.96    | 34.36    | 44.01     | 35.15 |
|      | Proposed     | 41.12| 32.28   | 40.27   | 25.63    | 35.41    | 45.37     | 36.68 |
Figure 4. Time consumption of different images in the two algorithms. Left: when sampling rate is 0.05; Right: when sampling rate is 0.1. (each point means an image is allocated by corresponding method)

It is obvious that our proposed algorithm has prominent performance advantages. For example, when $R=0.2$, our proposed method gains an average of 2-6 dB in PSNR over other methods. Figure 5, Figure 6 illustrates subjective quality and some details. It can be seen that our proposed method improve the overall image quality by reconstructing more image details. Therefore, we can conclude that our method is with advanced performance in both subjective and objective quality of the reconstructed CS.

Figure 5. Reconstruction of ‘Barbara’ at sampling rate $R=0.2$, (a) Original image; (b) BCS-SPL-DDWT; (c) Network-based CS; (d) Entropy-based allocation; (e) Variation-based allocation; (f) Proposed.
Figure 6. Reconstruction of ‘Camerman’ at sampling rate R=0.05, (a) Original image; (b) BCS-SPL-DDWT; (c) Network-based CS; (d) Entropy-based allocation; (e) Variation-based allocation; (f) Proposed.

6. Conclusion
In this paper, an adaptive sampling scheme is proposed, aiming at allocating sampling number as reasonably as possible based on image content in the case of limited sampling resources, so as to make the reconstructed quality as high as possible. In this paper, a reasonable sampling allocation method is designed referring to characteristics of MSE descent curve and saliency. The results show that both objective and subjective quality are improved and time consumption is controlled to some extent.

Acknowledgements
This research was supported by the National Natural Science Foundation Research Program of China (Grant 61371089) and the Fundamental Research Funds for the Central Universities (Grant 3102016zy019).

References
[1] J. Romberg and T. Tao, “Exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.
[2] D. L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[3] E. J. Candes and J. Romberg, “Quantitative robust uncertainty principles and optimally sparse decompositions,” Foundations of Computational Mathematics, vol. 7, no. 4, pp. 529–531, 2007.
[4] J. E. Fowler, S. Mun, and E. W. Tramel, Block-Based Compressed Sensing of Images and Video, Now Publishers Inc., 2012.
[5] S. Mun and J. E. Fowler, “Block compressed sensing of images using directional transforms,” in International Conference on Image Processing, 2010, pp. 2985–2988.
[6] J. Yang, X. Liao, and et al., “Compressive sensing by learning a gaussian mixture model from measurements,” IEEE Transactions on Image Processing, vol. 24, no. 1, pp. 106, 2015.
[7] Candes and J. Emmanuel, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus - Mathematique, vol. 346, no. 9, pp. 589–592, 2008.
[8] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok, “Reconnet: Non-iterative reconstruction of images from compressively sensed measurements,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 449–458, 2016.

[9] A. Adler, D. Boublil, and Zibulevsky M., “Block-based compressed sensing of images via deep learning,” in International Workshop on Multimedia Signal Processing (MMSP), 2017, pp. 1–6.

[10] Xuemei Xie, Yuxiang Wang, Guangming Shi, Chenye Wang, Jiang Du, and Xiao Han, “Adaptive measurement network for cs image reconstruction,” in CCF Chinese Conference on Computer Vision, 2017, pp. 407–417.

[11] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in International Conference on Machine Learning, 2010, pp. 807–814.

[12] Zhu, Shuyuan , B. Zeng , and M. Gabbouj ." [IEEE 2014 IEEE International Conference on Multimedia and Expo (ICME)-Chengdu, China (2014.7.14-2014.7.18)] 2014 IEEE International Conference on Multimedia and Expo (ICME) - Adaptive sampling for compressed sensing based image compression." IEEE International Conference on Multimedia & Expo IEEE, 2014:1-6.

[13] Liqun Zhong, Shuai Wan, Leyi Xie and Shun Zhang. "Content-Adaptive Image Compressed Sensing Using Deep Learning." Asia-pacific Signal & Information Processing Association Summit & Conference.

[14] Hou X , Zhang L . Saliency Detection: A Spectral Residual Approach[C]/ 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2007.

[15] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “Labelme: A database and web-based tool for image annotation,” International Journal of Computer Vision, vol. 77, no. 1-3, pp. 157–173, 2008.