Image Compressed Sensing Using Deep Learning

Saksham Gera[1], Shiva Gupta[2], Paras HS[3], Rohan Pillai[4], Madan Mohan Rayguru[5]

Electrical Engineering Department,
Delhi Technological University
Shahbad Daulatpur Village, Rohini, New Delhi,
Delhi 110042, India

[1]sakshamgr3@gmail.com
[2]Shiva gupta email
[3]Paras HS email
[4]rohan92@dtu.ac.in
[5]mmrayguru@dtu.ac.in

***Abstract*—In this paper, techniques used to compress data, especially images are discussed. A deep learning model based on convolution neural networks is implemented and the results in the form of PSNR (Peak Signal to Noise Ratio) are compared. The main hurdles in compressed sensing of images are the sampling matrix design and the reconstruction method. To overcome these we propose a CNN framework called CS Net, which consists of a joined sampling network and reconstruction network. The sampling network adapts the sampling matrix during the training of data, which enables the compressed sensing measurements to maintain additional structural data for better recreation of images. The reconstruction network is divided into two parts, an elementary linear network, and a deep reconstruction non-linear network; it adapts a mapping between the endpoints of structural measurements and reconstructed images. In this project, we used a different number of blocks in the deep reconstruction and compared the results. The PSNR values of the reconstructed images gradually increase as we increase the number of residual blocks, but after a certain point, the values start decreasing[1].**

The results obtained show that CSNet produces high-quality images like those obtained from methods like FFT and Nyquist Shanon but with comparatively low computation costs[2][3].

***Keywords*— Compressed Sensing, Convolution Neural Networks, CSNet, Sampling matrix, Image reconstruction.**

I. INTRODUCTION

Image compression is a signal compression technique in which the real picture is reproduced using a limited amount of memory. Picture compaction aims to reduce image duplication and make data storage and transit extra affordable. The fundamental purpose of such a system is to decrease the maximum quantity of data that may be stored so that the decoded image is as close to the original as possible.

A. Overview

The classical image capturing system typically obtains a high volume of samples focused on the Nyquist-Shannon sampling theorem, with a sampling ratio of at least two times the signal's bandwidth, and then compresses the signal for storage and retrieval using a computationally complex compression method to remove redundancy. However, in certain image processing applications where data gathering devices must be simple, this type of image capture system may not be preferred.

Compressed Sensing (CS) is an emerging technique that illustrates a paradigm shift for picture storage and recovery that incorporates skimming and contraction. When a signal is sporadic in a domain, the compressed sensing theory indicates that a signal can be retrieved with many lesser observations than the Nyquist-Shannon sampling theorem suggests. It is commonly known that images contain a lot of unnecessary data and may be expressed sparsely. As a result, according to CS theory, an image may be compressed and rebuilt effectively.

Since the core CS study indicated that when a signal is sparse in a particular domain, it may be recreated with excellent clarity at a lower sampling ratio, there's been a great deal of interest in CS for picture capture. Image acquisition devices based on CS have been developed. The so-called single-pixel camera is the most well-known CS gadget among them. Other steps have been taken to improve the imaging versatility of CS-based cameras and to investigate their potential use in cellular devices and other portable devices.

The creation of the sample matrix and the design of the reconstruction method are the two primary issues in the research of CS. In the literature, several solutions for dealing with these two issues have been offered. To address the very first difficulty, many sample matrices, such as the random matrix, binary matrix, and structural matrix, have been created. Furthermore, all of these matrices are input signal independent sampling matrices, meaning they disregard the signal's properties. Many sparsity-regularized-based solutions have been developed for the second issue, such as convex-optimization algorithms, greedy algorithms, and iterative thresholding algorithms. Certain efforts in image CS explored picture priors to create more complicated models; others, on the other hand, included more treatment efficiency in the repeated process of compression.

In our research paper, new variations on the CS framework based on convolution neural networks(namely CSNet) are tried and the results have been compared to find the most optimum number of residual blocks in the advanced reconstruction network.

We have used a Convolutional Neural Network (CSNet) to create a CS framework that contains a sampling network & a reconstruction network that are optimized together. The floating-point sampling matrix is foremost learned from the training pictures by the sampling network, allowing the CS measurements to preserve further image contextual data for improved recovery.

Among the readings of CS and the rebuilt pictures, the reconstruction network adapts an end-to-end point mapping. It comprises a remnant training-based non-linear deep reconstruction structure and an initial reconstruction network. Inter-block information may be successfully used by the reconstruction network, and blocking artifacts can be avoided.

Unlike classic CS approaches, the learned model does not require the sampling matrix to be transferred from encoder to decoder. Traditional image CS reconstruction algorithms can also benefit from the learned sampling matrix.

B. Principal of Image Compression

The fact that nearby pixels are connected, therefore, they contain duplicate information. It is a frequent feature of most images. The first and foremost step is to find a representation of the image that is less coupled. Redundancy and the reduction of irrelevant features are two key components of compression. Irrelevancy reduction removes sections of the signal that the receiver will not notice, such as the Human Visual System (HVS).

1) *Coding Redundancy*: The representation of information is linked to coding redundancy. Source codes serve as a representation of the information. When an image's grey levels are processed in a way that employs more coded symbols to represent each grey level than what is strictly necessary, the resulting picture is said to have coding redundancy.

2) *Spatial/Temporal Redundancy:* As maximum 2-D intensity array pixels are connected spatially; content is duplicated in the representations of the associated pixels. Sequentially connected pixels in a video clip also replicate information.

3) *Irrelevant Information:* Many 2-D intensity arrays carry data that the human supervision ignores and is irrelevant to the image's intended use. In the perspective that it isn't used, it is redundant.

Image compression research tries to reduce the bits needed to represent an image as much as possible by eliminating spatial and spectral redundancy.

C. Motivation

In recent years, the transmission and storage of data have increased by many folds. Data compressions techniques are used to reduce the storage size of data and allow faster transmission of larger data over the internet. Images are one of the most popular data types. Image compression deals with reducing the size of an image by eliminating redundant and unnecessary signals from its sparse matrix. State-of-the-art methods like Nyquist-Shannon or Fast Fourier Transform allows high-quality reconstruction of images but at the cost of computational power. Higher computational power is not available throughout and also slows the speed of the process. Therefore, the compressed sensing technique offers compression of data below Nyquist rates making it an attractive solution in facial recognition systems and medical imaging. Compressed sensing produces images of similar quality but at lesser computation time. Hence, the need for a proper deep learning model for compressed sensing.

II. BACKGROUND

In classical picture acquisition, the analog signal is captured by collecting a large number of data samples based on the Nyquist-Shannon sampling theorem.

Following the acquisition, the image (dense image) is converted over a basis Ψ (e.g., Fast Fourier transform of wavelet basis) with tiny coefficients for the majority of the places in the spatial domain of the image.

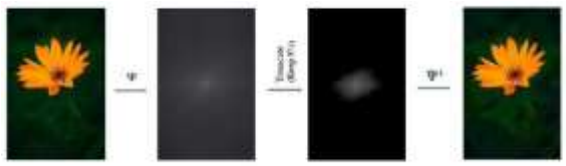

Figure 1An example showing traditional image compression and decompression using a basis function Ψ over a high-resolution image.

The pixels in the modified image are then shortened, removing a significant amount of data (95% in Fig. 1). The obtained picture (sparse image representation) may be easily stored as a dictionary of key-value pairs, with the key being the pixel's spatial orientation and the value being the transform coefficient. Compression is the term for this procedure.

Image data in the form of a dictionary is very straightforward to send over computer networks and consumes a smaller amount of storage.

The compressed sparse image is subjected to the inverse basis transform in order to rebuild the image. Decompression is the term for this procedure.

A. *Compressed Sensing Background*

Compressed sensing is a type of data collection that allows us to rebuild high-resolution pictures by sampling a significantly less amount of data than the usual image capture approach.

When a signal is sparse in some domains, the CS theory allows flat representation of a wide-dimensional signal into a dimension considerably tiny in contrast to the raw signal while permitting retrieval of the data from projections with precision.

Particularly, visualize that $x \in R^{N\times 1}$ is a real-valued signal (original image) and $\phi \in R^{M\times N}$ is a sampling matrix. ($M << N$). The CS measurement collection process is written as

$$y = \phi x \quad (1)$$

where $y \in R^{M\times 1}$ is the Compressed sensed metric (CS measurement). The CS theory indicates that successfully recuperating x is achievable because the signal x is scarce in a certain domain ψ.

The simplest way to describe CS reconstruction is to write it like this:

$$\min ||\psi_x||p \text{ , s.t. } y = \phi x \quad (2)$$

The suffix P is normally assigned to 1 or 0, describing the sparseness of the vector ψ_x, where x are the sparing factors w.r.t to scope. In previous literature, a wide variety of techniques for tackling this optimization issue have been offered. The convex optimization approach, for example, converts a nonconvex issue into a convex one to obtain an estimated solution. The most frequent convex optimization approach for CS restoration is basis pursuit. It solves a linear programming issue by replacing the L_0 norm constraint with the L_1 norm constraint. Such convex-programming approaches, on the other hand, have a large computational cost.

Convex optimization, gradient-descent techniques, greedy algorithms, and Land weber (PL) algorithms are some of the methodologies presented in the literature for tackling this optimization issue. Despite their efforts, classical approaches have a high computational cost due to the necessity of repetitive computing[9]. Reconstructing a high-quality image using these CS approaches might take anything from a few seconds to many minutes. In addition, the sampling matrix from the encoder to the decoder should be sent

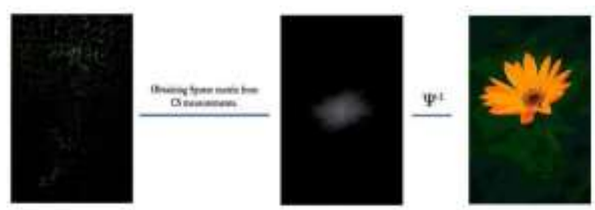

Figure 2An example showing recovering original image X by sampling a few pixels (Φx) while sensing and solving for the sparse representation over a basis Ψ.

A. *Sparsity*

The term sparsity, as the name implies, relates to the smallest quantity of data we can obtain and retain in order to reconstruct an original signal/document from it.

A sparse matrix is one with a large number of zero values. Sparse matrices are distinguished from dense matrices, which have a large number of non-zero values.

The fascination with sparsity stems from the fact that its use can result in significant computational savings, as well as the fact that many big matrix problems encountered in practice are sparse.

Sparsity should not be mistaken with missing data, because missing data indicates that a large number of data points are unknown. If the data is sparse, on the other hand, all of the data points are known, but the majority of them have little value[10].

A sparse matrix is a two-dimensional array of pixels with some pixels having zero intensity in terms of picture reduction. Sparsity is a technical term that refers to the percentage of cells in a database table that is empty.

Finally, all image compression techniques and methodologies aim to discover the optimum sparse matrix that can identify a particular image and compress it in the most optimal spatial and time complexity way possible.

B. *Nyquist-Shannon Theorem*

The Nyquist-Shannon theorem explains the sampling of a signal or waveform so that no information is lost[1][2].

Let's say we have a signal *X(t)*. On taking Fourier Transform of this signal, $X^{\Lambda}(f) = F\{x(t)\}$, there will be a f_{max} for which,

$$x|f| = 0 \forall |f| > f_{max} \quad (3)$$

such that beyond the maximum frequency f_{max} , there is no power in the signal. The Nyquist-Shannon theorem indicates that if we're to sample this signal, we'd require samples with a frequency greater than double the signal's peak frequency, that is,

$$f_{sample} > 2f_{max} \quad (4)$$

If that's the case, no information was sacrificed during the sampling procedure, and the source signal might be restored from the sampled signal.

This theorem is a stepping stone toward more advanced signal (image) processing, and it serves as a benchmark for various techniques and methodologies aimed at achieving the best image compression.

It also lays the groundwork for image compression algorithms that use a sampling frequency lower than Nyquist's.

III. Proposed Work

The formulation of the sampling matrix and the design of reconstruction algorithms are the two key problems in the research of compressed sensing. In the literature, several solutions for dealing with these two issues have been offered. To address the first difficulty, many sampling matrices have been devised, including the random matrix, the binary matrix, and the structural matrix. However, each one of these sampling matrices are independent of the signal, meaning they disregard the signal's properties.

Many sparsity regularized-based approaches, such as convex optimization algorithms, greedy algorithms, and iterative thresholding algorithms, have been presented in response to the second difficulty. Several image CS works investigated pictures, prior to constructing more intricate models, while others added additional optimal factors into the recurrent thresholding procedure[4]. Some present state-of-the-art CS reconstruction methods that take numerous minutes to recreate a solitary slightly elevated picture, whilst others have much worse reconstruction accuracy. Deep learning has lately been shown to be more effective in computer vision tasks. There are indeed some deep learning-based picture CS reconstruction techniques that employ a random floating-point sampling matrix that we are aware of. However, because block-by-block reconstruction approaches only employ Intra block information to recreate a block, blocking artifacts emerge, and postprocessing is frequently required. Because of the usage of an orthogonal matching pursuit algorithm, the postprocessing approach has are relatively high computational cost[7].

IV. Implementation

Fig 2. Shows the proposed architecture of CSNet. This architecture is basically divided into three parts, i.e. block-based compressed sampling, initial reconstruction, and non-linear signal reconstruction.

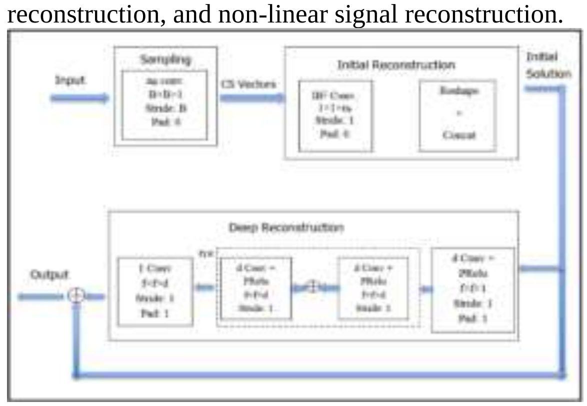

Figure 3 Flow Chart depicting the CSNet model architecture.

A. *Sampling Network*

In Block-based compressed sensing, the image is divided into $B \times B \times L$ where B is the height and width of the block and l denotes the number of channels. To make these blocks non-overlapping, we will be using a stride of Y.

The measurements are then obtained with a $n_b \times lB^2$ sampling matrix ϕ_B, where $n_B = \left[(M / N)^{lB^2}\right]$

$y_j = \phi_B X_j$ is the formula for this procedure.

If each row of the sampling matrix B is considered a filter, we may imitate this compressed sampling process using a convolutional layer.

Because each block of image is B × B × 1, each of the sampling network's filters is similarly B × B × 1, resulting in each filter producing one measurement.

To acquire n_B CS measurements, the sampling matrix B has n_B rows for the sampling ratio (M / N).

As a result, this layer has n_B filters of size B × B × 1. For non-overlapping sampling, this layer's Stride is B × B. Furthermore, each filter is free of prejudice. In addition, this layer has no activation[6].

The Block-based compressed sampling network (S) is conceptualized as

$$y = s(x) = ws * x$$

where * denotes convolutional operations, x denotes the input image, y represents the CS measurement, and $ws *$ denotes n_B filters that support B × B × 1. The result appears to be made up of n_B feature maps, with each column representing the n_B measurements of an image block.

B. Initial Reconstruction Network

The picture is rebuilt using a recovery network (abbreviated R), which consists of two networks. The first one is an Initial Reconstruction Network (abbreviated I) and the second one is a Deep Reconstruction Network (abbreviated D).

$$R(y) = D(I(y))$$

Provided the j^{th} block's CS measurement y_j, the first reconstruction yields $x_j\sim = \phi_B \sim yj$. Clearly, $\Phi^{\sim}_B$ is a matrix with dimensions of $1B^2 \times n_B$.

To implement the first reconstruction procedure, we employ a convolutional layer with a particular kernel size and stride. $\Phi^{\sim}_B$ is adaptively optimized in the network.

The first image reconstruction can be described as the following procedure $I^{\sim}(y)$:

$$I^{\sim}(y) = W_{int} * y$$

The CS measurement is y, and the filters are W_{int}. The sampling network outputs a $1 \times 1 \times n_B$ vector for each image block and the size of the output of the sampling network and the size of each filter in the initial layer are the same.

As a result, W_{int} is equivalent to $1B^2$ filters with support of $1 \times 1 \times n_B$. To reconstruct each block, we have set the stride of this convolutional layer to 1 × 1. Bias isn't present either.

Each column of $I^{\sim}(y)$ appears to be of $1 \times 1 \times 1B^2$ type. The reassembled block, however, is still a vector. To get the initial linear reconstruction, the combinational layer has two functions, the first one is a reshape function and the second one is a concatenation function. This whole process has been denoted by *I(y):*

$$x^{\sim} = I(y) = k\,(\text{for all image blocks}\ \gamma(\text{Image block}))$$

where γ is the reshape function and k is the concatenate function.[8][11]

The reconstruction of the signal in this network is linear as we are not using any activation functions

C. Deep Reconstruction Network

We have used residual block in this network to accomplish the non-linear signal reconstruction. There are three operations in this network: Feature extraction, non-linear mapping and feature aggregation.

D. Why have we used Residual blocks?

In a word, ResNets handle the problem of deep neural network effectiveness degrading, as the system grows larger.

A leftover block is a group of tiers in which the result within one layer is retrieved and combined with the result of a level farther in the block. The fluctuation is then transferred to the main path by mixing it with the result of the appropriate layer. This by-pass connection is known as a shortcut or skip-connection[4].

1) Feature Extraction:

The high-dimensional feature is extracted from the local receptive field using a feature extraction method. It consists of a convolutional layer and an activation layer.

The convolutional layer has d filters of size $f \times f \times 1$ since it works with the initial reconstruction output. This operation is denoted by the symbol $De(x^{\sim})$:

$$De(x^{\sim}) = Activation(W_e * x^{\sim} + B_e)$$

The initial reconstructed result generated by the Initial reconstruction network is denoted by $x^{\sim}$.

W_e stand for d filters with sizes of $f \times f \times 1$, B_e for biases of size $d \times 1$, and *Activation()* for activation function. For the activation function, we used the Parametric Rectified Linear Unit (PRelu, $max(0,x) + a * min(a,x)$).

2) Non-Linear Mapping:

This network dynamically cascades residual block convolutional layer and activation layer after obtaining the high-dimensional image feature, increasing the net-work non-linearity.

This procedure is written as:

$$D^i_{m1}(x^{\sim}) = Act\,(D^{i-1}_{m2}(x^{\sim}) + W^i_{m1} * D^{i-1}_{m2}(x^{\sim}) + B^i_{m1})$$
$$D^i_{m2}(x^{\sim}) = Act\,(W^i_{m2} * D^i_{m1}(x^{\sim}) + B^i_{m2})$$
$$D^0_{m2}(x^{\sim}) = D_e(x^{\sim})$$

Where i ∈ { 1,2,....,n }. There is a short skip connection between the input and output in the residual block. W^{i}_{m1} and W^{i}_{m2} contain $f \times f \times d$ size d filters, B^{i}_{m1} and B^{i}_{m2} are $d \times 1$ biases, and $Act(.)$ is a PRelu activation function.

3) Feature Aggregation:
A feature aggregation operation is used to reconstruct the image from the high-dimensional feature to provide the final output. The operation is described as $D_a(\tilde{x})$:

$$D_a(\tilde{x}) = W_a * D^{n}_{m2}(\tilde{x}) + B_a$$

W_a denotes a filter of size $f \times f \times d$, while B_a denotes the bias of size 1 X 1.
In the diagram, we can see that we have also made a long link between the linearly reconstructed picture and the deep reconstruction network's output $D_a(\tilde{x})$ to speed up network convergence. As a result, the final reconstructed image looks like this:

$$D(\tilde{x}) = \tilde{x} + Da(\tilde{x})$$

V. Result Analysis

Table 1 Results obtained from the deep learning model.

Number of Residual Blocks	PSNR
1	28.997789713705338
2	29.022606572723685
3	29.218752044341894
4	29.440226792141814
5	29.170336279307907

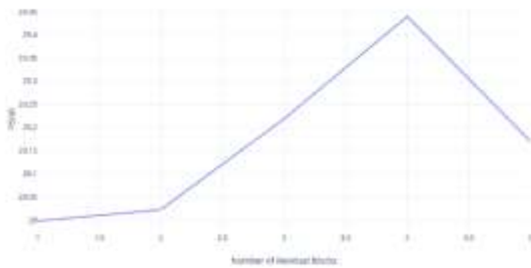

Figure 4 Plot of PSNR values vs number of residual blocks.

As shown in figure 4, initially the PSNR value rises with increase in the number of residual blocks, which implies that the quality of compressed images also increases. But after reaching a certain number of blocks (4 in this case), the PSNR value attains it's maximum value and decreases gradually thereafter. Therefore, 4 residual blocks is the most optimum setting to obtain high quality images from the CSNet framework.

VI. Conclusion & Scope

In this paper, a better approach to state-of-the-art image-compressed sensing techniques have been proposed. The traditional methods have many issues, like high computational power requirements and slow computational speeds.
Although there are methods that require less computation and produce faster results but at the expense of the quality of the output images. CSNet reconstructed images achieve high average PSNR values, and low computation times, depicting that the learned CS measurements retain more image structural information for better reconstruction, offering high quality compressed images at faster running speed. We found that the CSNet achieves peak results when four residual blocks were used.
Furthermore, there is a huge scope of research in the field of compressed data sensing. Like, applying the CSNet framework in real world applications such as facial recognition and MRI scans.

References

1. Wuzhen Shi, Feng Jiang, Shaohui Liu, Debin Zhao. "Image Compressed Sensing using Convolutional Neural Network", IEEE transactions on Image Processing, 2020.

2. P Hanumath, P Bhavana, Shreyanka Subbarayappa. " Application of deep learning and compressed sensing for reconstruction of images", Journal of Physics: Conference Series, 2020.

3. Yi Xu, Guiling Sun, Tianyu Geng, Ying Zhang. "A two-Stage Fusion Method of Reconstruction Algorithms for Compressed Sensing", International Journal of Wireless Information Networks, 2018.

4. Su H Bao,Q,Chen Z, ADMM Net " A Deep Learning Apporach for parameter Estimation of Chirp Signals under Sub Nyquist Sampling", IEEE Acess 2020,8,75714-75727.

5. Zheng, S.; Yang, X.; Yuan, X. Two-Stage is Enough: A Concise Deep Unfolding Reconstruction Network for Flexible Video Compressive Sensing. arXiv 2022, arXiv:2201.05810.

6 K Kulkarni, S Lohit,P Turaga, R Kerviche, A.Ashok " Recon Net Non iterative reconstruction of images from compressively sensed measurements", Proc IEEE Conf Vis Pattern Recongit PP 449-458 Jun 2016.

7. W. Shi, F. Jiang, S. Zhang and D. Zhao, "Deep networks for compressed image sensing", *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pp. 877-882, Jul. 2017.

8 Y. Wang, W. Yin and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization", *Comput. J. Sci. comput.*, pp. 1-35, 2018.

9 K. Zhang, W. Zuo, L. Zhang et al., "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising", *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142-3155, Jul. 2017.

10 D. L. Donoho, "Compressed sensing", *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

11 L. Zhu, X. Wu, Z. Sun, Z. Jin, E. Weiland et al., "Compressed-sensing accelerated 3-dimensional magnetic resonance cholangiopancreatography: Application in suspected pancreatic diseases", *Invest. Radiol.*, vol. 53, no. 3, pp. 150-157, 2018.

12 S. Beygi, S. Jalali, A. Maleki and U. Mitra, "An efficient algorithm for compression-based compressed sensing", 2017.

13. S. Ravishankar, B. E. Moore, R. R. Nadakuditi and J. A. Fessler, "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging", *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1116-1128, May 2017.