Enhanced Compressed Sensing Based Deep Learning Neural Network for Single Image Super Resolution of COVID-19 Using X-Ray Images

Hossam M. Kasem[1,2], Samar Atef [1], Mohamed E. Nasr[1]

[1] *Electronics and Electrical communications Dept. Faculty of Engineering, Tanta University, Egypt*
[2] *Faculty of Engineering, Horus University in Egypt (HUE)*

Correspondence: Samar Atef; Email: Samaratef307@gmail.com

Compressed sensing (CS) represents an efficient framework to simultaneously acquire and compress images/signals while reducing acquisition time and memory requirements to process or transmit them. Specifically, CS is able to recover an image from random measurements. Recently, deep neural networks (DNNs) are exploited not only to acquire and compress but also for recovering signals/images from a highly incomplete set of measurements. Super-resolution (SR) algorithms attempt to generate a single high resolution (HR) image from one or more low resolution (LR) images of the same scene. Despite the success of the existing SR networks to recover HR images with better visual quality, there are still some challenges that need to be addressed. This paper designs a deep neural network that generates HR images from LR Xray COVID-19 images. To address this problem, we propose a novel robust deep CS framework that is able to mitigate the geometric transformation and recover HR images. Specifically, the proposed framework is able to perform two tasks. First, it is able to compress the transformed image with the help of an optimized generated measurement matrix. Second, the proposed framework is able not only to recover the original image from the compressed version but also to mitigate the transformation effects. The simulation results reported in this article show that the proposed framework is able to achieve a high level of robustness against different geometric transformations in terms of peak signal-to- noise ratio (PSNR) and similar structure index measurements (SSIM).

Keywords: Compressed Sensing, Covid-19, Deep Neural Networks, Super Resolution

1. Introduction

During last 2 years, the world faces another pandemic which is novel coronavirus, namely SARS-CoV-2. It is discovered for the first time in in Wuhan, China, around December 2019. After that by short time, he coronavirus became known as COVID- 19 and had spread across nearby countries within a short period, and soon after, around March 2020. The World Health Organization (WHO) declared a global pandemic [1]..Due to high virus infection, it is essential to control the disease, including rapid diagnosis and timely quarantine. Coronaviruses are a broad family of four types of viruses, including alpha- coronavirus, beta-coronavirus, coronavirus delta and gamma- coronavirus [1]. Up to now, seven out of 40 different species have been found capable of spread to humans in the Coronavirus family resulting in diseases such as common cold [1]. Past experiments have shown that both SARS and MERS viruses are passed on to humans from cats and camels. COVID- 19 is believed to have been transmitted to humans from bats and anteaters [1].

According to the latest guidelines published by the World Health Organization (WHO), reverse transcription-polymerase chain reaction (RT-PCR) or gene sequence for respiratory or blood samples must validate the diagnosis of coronavirus as the main indicator for hospitalization [2]. But it takes a long time for the relevant kits to detect a virus, in addition to the high time required for diagnosis, they are also less sensitive to the detection of the virus [2]. In addition, previous studies have shown that most doctors and nurses have been infected with the COVID-19 during salvia sampling based on RT-PCR kits [2]. These kits are also limited in the world and require transport costs. According to the content, the diagnosis based on these RT-PCR kits exposes doctors and nurses to the COVID-19 virus and is ultimately not cost-effective. Therefore, rapid diagnosis of COVID-19 is necessary to treat and control the disease [2].

The chest scan is another method used during the treatment of this disease [10]. Computerized tomography (CT) and Digital Radiography (or standard 2D X-ray) scans are among the methods of chest imaging. DR is used to scan the body for the diagnosis of fractures, lung infections, pneumonia and tumors. CT is an advanced DR that provides clearer images of organs, bones, and tissue. However, CT scans are not available in all medical centers and are not as affordable as DR. For this reason, today physicians usually use DR in the first step of

diagnosis. Using X-ray is a faster, easier, more affordable and more harmful method than CT [2]. After taking chest X-ray images, the doctor should visually diagnose bacterial, viral, COVID-19 and etc. infections. Due to exhaustion and the need for expert personnel, visual examination-based diagnosis is unpleasant, time-consuming and incorrect often results in low diagnosis accuracy. Also, the image consistency of chest X-ray has certain flaws, such as low contrast, over- lapping organs and blurred boundary, which has a significant impact on the identification of chest X-ray pneumonia [2]. Based on the above-mentioned facts, automatic detection of the virus type (including COVID-19) based on chest images has gained a lot of attention in recent researches. Automatic detection of COVID-19 not only causes a fast diagnosis but also reduces the workload of doctors and is useful for timely treatment and patient mortality reduction.

Recently, various computational methods based on deep learning have been developed for the observation and analysis of the automatic detection of COVID-19 used X-Ray images. Consequently, this paper proposes a novel robust deep CS framework that is able to mitigate the geometric transformation and recover HR images. Specifically, the proposed framework is able to perform two tasks. First, it is able to compress the transformed image with the help of an optimized generated measurement matrix. Second, the proposed framework is able not only to recover the original image from the compressed version but also to mitigate the transformation effects.

The traditional image acquisition system typically follows the Nyquist-Shannon sampling theorem that yields acquiring an intensive number of samples. The Nyquist-Shannon sampling theorem states that the sampling frequency should be equal or greater than twice the bandwidth of the signal. So, for efficient storage and/or transmission, compression of the signal is needed to remove redundancy by a computationally complex compression method. The data acquisition devices, in some applications, are required to be simple. Additionally, the over-sampling can damage the captured object as in medical imaging. Consequently, these kinds of image acquisition systems maybe not suitable for such kind of applications.

The developing technology of compressed sensing (CS) presents a new technique not only for image acquisition but also for image reconstruction that can simultaneously perform the sampling and compression processes. Specifically, ac- cording to [3, 4], CS is able to recover the signal from much fewer measurements when the signal is sparse in some do- main compared with the number of measurements required by Nyquist-Shannon sampling theorem. It is well known that the images can be sparsely represented as images have much amount of redundant information. In this way, according to CS theory, the images can be compressed and reconstructed efficiently. To this end, there are basic challenges that need to be solved including the design of the sampling matrix and the development of the reconstruction method.

Single image super-resolution (SISR) can be considered as one of the most famous computer vision problems that recently attracts researchers attention. SISR is an image processing task that aims to generate a high-resolution (HR) image from its low-resolution (LR) version of images. The difficulty of achieving SISR lies in computing the HR image from its LR version as it is a many-to-one mapping problem. How- ever, numerous image super-resolution (SR) methods have been reported to deal with this non-trivial problem [5,6]. Over the last decade, a various traditional non deep-learning (DL) based approaches have been utilized in SISR computer vision task, including prediction-based methods [7–9], edge- based methods [10, 11], statistical methods [12, 13], patch- based methods [14, 15], missing data reconstruction in re- mote sensing image [16–18],and sparse representation methods [12, 19].

Recently, with the massive progress achieved in the DL approaches, DL-based SR models are widely proposed and often achieve better performance compared to various traditional super-resolution methods. Specifically, convolutional neural network (CNN) methods have achieved superior performance in SISR tasks. Consequently, CNN has been presented to leverage SISR tasks [6, 20–37]. The correlation be- tween the LR and HR images can be easily learned by DL- based techniques and better performance compared to conventional methods can be achieved. Additionally, these techniques are able to generate a high resolved image. However, there are still several challenges and limitations of the existing algorithms [38–41].

2. Material and methods

2. Challenges And State-Of-The-Art

In practical scenarios, our proposed framework is able to overcome the following challenges:

2.1 Measurement Matrix Design

The restricted isometry property (RIP) must be satisfied in the measurement matrix design in order to guarantee the performance of CS [3, 4]. Therefore, the design of the measurement matrix is crucial for spike recovery. In the last few years, various methods have been developed to design the sampling/measurement matrices. These methods are built based on random, binary, [42, 43], and structural matrices [44, 45]. However, these sampling matrices are all signal independent as well as non-optimal, due to the fact that they are unaware with the characteristics of the signal. In order to circumvent this issue, we propose a DL-based framework to design optimized sampling matrices for sensing and compressing tasks.

2.2 Efficient Recovery Solver

In addition, CS recovery methods should be able to recover the original image with better visual quality from its com- pressed version. Recently, various sparsity-regularized-based methods have been presented (e.g. [46–48]), the greedy algorithms (e.g. [49, 50]), and the iterative thresholding algorithms (e.g. [51]). However, these algorithms are not suit- able for real-time transmission due to the high computational complexity of these algorithms. Consequently, the challenge for developing recovery algorithms for real-time applications needs to be solved. In our framework, we propose to use the DL-based framework to recover the original image.

2.3 Robustness Against Geometric Transformation

The images can be affected by one or many geometric trans- formations in many real-life applications such as translation, rotation, and scaling, etc. To establish a better relationship between the input and target output images, the spatial trans- formations can be utilized [52] including perspective or affine transformations to overcome the geometric transformations in all applications. Thus, the design of accurate and robust DL networks must be invariant to input spatial transformations. Despite the good performance achieved by the existing SR networks to generate the HR images, it can not alleviate the effect of the geometric transformations. This is mainly due to the limited receptive field of CNNs which makes these net- work unable to be spatially invariant to the position of features [53].

The spatial transformer network (STN) [52] has been presented to achieve the spatial invariance. The STN is characterized by the capability to be merged into existing CNN to provide the ability of mitigating the geometric transformation effect. Additionally, the STN network can provide a dynamic mechanism for performing a suitable transformation for each input sample [54, 55]. The spatial invariance of the SR net- work may not be typically non-adequately expanded by inserting the STN into the existing SR image networks. As known, the bilinear interpolation in the pixel domain is utilized by STN to estimate the transformation parameters and re-sample the input image. Thus, its output often becomes blurred compared with its original input image [52]. Consequently, we propose a robust DL-based HR image recovering framework to simultaneously address the geometric corrections and recover HR image from its corrupted version.

2.4 Contributions And Paper Organization

In summary, our main contributions are as follows:

•   We propose a robust CS-SR based network that is able to simultaneously perform various tasks including; 1) compressing the input image according to the CS theory; 2) geometric transformation corrections;3) recovering HR images.

•   We propose to utilize the CS algorithm to compress the input LR images. Consequently, the processing time required to recover the HR images is reduced.

•   We compare our proposed framework with existing SR image methods in terms of the training cost and learning complexity.

•   The superiority of our proposed framework compared with the existing benchmarks is supported and verified in terms of both peak signal-to-noise-ratio (PSNR) and similar structure index measurements (SSIM) by extensive experiments.

The remainder of this article is structured as follows. Section II presents a survey on SR related work besides the basics of CS, and spatial transformations. Our proposed robust CS network is described in Section III. Section IV presents experimental results and their analysis. Finally, section V presents the conclusion of this article.

3. Related Work

over the past decade, significant progress has been achieved in several applications for real-world image processing, including SISR [5], image classification [56], object detection [57, 58]. These achievements are related to the huge advances in CNNs [56, 59], computational power, and the availability of enormous amount of data [60]. This section begins by ex- plaining related work on SISR, spatial transformations, followed by a review of CS concepts.

3.1 Single Image Super-resolution

Single image super-resolution (SISR) reconstruction is to recover a corresponding high-resolution (HR) image from a low-resolution (LR) image. As a result, this problem can be formulated as :

$$\hat{x}_{HR} = argmin_{x_{HR}} \left|\left| x_{LR} - DH_{x_{HR}} \right|\right|_2 \tag{1}$$

where x_{LR} , x_{HR} are LR and HR images, respectively. The geometric transformations, and a down-sampling operator can be represented by degradation matrices H, and D, re- spectively. The down-sampling operator, D, is firstly applied to the HR images, (i.e., x_{HR}), to generate the observed LR

images (i.e., x_{LR}). For this problem to be solved, an effective prior is needed to transform this problem into a deterministic one. In the last decades [5, 61], several solutions have been proposed to solve this ill-positioned

problem. Recently a large number of DL studies have been conducted to solve the question of SR image. As a result, the existing DL-based SR techniques related to our work are evaluated as follows.

Super-resolution deep convolutionary networks (SRCNN)

[5] is one of the ground-breaking research which is the first attempt to apply CNN for SISR problem. The main idea be- hind SRCNN is mapping the bicubic up-sampled LR space to HR space. Specifically, SRCNN utilizes the bicubic interpolation as a pre-processing step. Further, SRCNN uses deep convolutional layers to extract the features of the overlapping patches. Finally, the reconstructed HR images are generated by non-linearly mapping of the extracted feature vectors to each other and then aggregated in the form of patches. The SRCNN is characterized by its simple structure which utilizes only convolution layers. Thus, the input image with various sizes can be passed through the SRCNN in one path. De- spite the straightforward structure of the SRCNN, it still has a number of limitations. The most critical limitation is related to the slow convergence of the network, and the network only operates on a single scale.

The authors of [62] have made great efforts to improve SRCNN's performance through proposing FSRCNN. Specifically, the pre-processing Bicubic interpolation in SRCNN is replaced by a deconvolution post-processing activity [62]. Structurally, the FSRCNN performs extraction, shrinking, mapping, and expansion of the input by utilizing four layers of convolution. Sequentially, the FSRCNN performs the mapping step on the input followed by shrinking dimensions of the features. Finally, the expansion step is performed back at the latter point. Both SRCNN, and FSRCNN utilize the mean square errors (MSE) functions as the loss functions in network training.

Recently, the researchers tried to answer a crucial question that has appeared due to the huge utilization of CNN in SISR. The question is "Should a deeper network be implemented to optimize the SR performance?". Among all the answers, the authors in [21] have presented an answer to this question by designing a very deep super-resolution network (VDSR). The design of VDSR is inspired by a very deep VGG-network that is used for ImageNet classification. Structurally, VDSR is a cascaded network and consists of 20 layers in depth. Furthermore, all the filters employed have a size of 3-by-3 [21]. Inspired by [56], the authors of VDSR [21] utilize the residual learning to train the VDSR network. Additionally, the weakness of SRCNN is addressed by expanding SR with a single network model to multi-scale. Despite the success of CNN based SR models, but these networks are not able to mitigate the effect of the geometric transformation. Motivated by that, we propose a deep robust network that is not only able to overcome the geometric transformation effects but also recover the HR image from its corrupted version.

The author of [23], proposed deepens the network by stacking simplified residual units on the basis of global residual, namely EDSR. This network uses global residual learning, end-to-end image super-resolution. The simplified residual unit includes only two convolution layers and one ReLU activation layer

3.2 Spatial Transformer

Invariance against geometric transformations is often a desirable property to any computer-vision model, which is also highly demanded in many practical applications across areas of both computer vision and multimedia processing. Traditionally, the researchers have proposed various methods to design a model that is invariant to affine transformations including translations, rotations, scaling, etc. Some examples of these methods that have achieved a robustness level to various transformations are hand-crafted features such as HOG [63], SIFT [64] and SCIRD [65]. Despite that the convolutional layers are able to learn the filters in a translation-invariant manner, the filter response is still not invariant. Addition- ally, the Max-pooling methods provide a way to be invariant to affine transformations. However, this way is not suitable for large translations. This is mainly due to the pooling in practice is performed over a small region (e.g., 2   2 or 3 3). Thus, spatial invariance is provided only up to a few pixels by each pooling. By applying filters at multiple scales and locations followed by max-pooling, a locally scale-invariant representations are obtained [66].

The spatial transformer network (STN) [52] is also closely related to our proposed work. STN is able to permit for significantly-larger (parameterized) transformations. STN can be inserted into CNN and provide an end-to-end learning mechanism. In this way, STN is able to provide a way for achieving the spatial invariance for the model. However, STN utilizes the bilinear interpolation method to sample the input images in the pixel domain. Consequently, the input images are passed through the bilinear interpolation. Thus, the quality of the transformed image deteriorates.
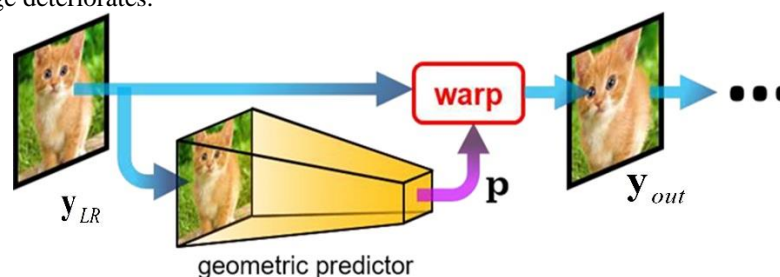

Fig. 1: Structural overview of the existing STN [52]

Fig.1 shows a structural overview of STN [52], where three modules can be identified, including: (i) a localization network, which takes the input and estimates the transformation parameters θ; (ii) a grid generator, which creates a sampling grid for reconstruction of output images; and (iii) a sampler, which computes the output by populating the sampling grid.

3.3 Compressed sensing (CS)

The CS theory states that the vector x can be accurately recovered from M random measurements over the measurement matrix [3, 4]. This can be formulated as:

$$y_{CS} = \Phi_x \qquad (2)$$

where x is a signal with length N (i.e. x RN). The measurement vector y_{CS} length is equal to M and M << N . Furthermore, there is a basis function Ψ in which x is sparse. Consequently, the signal x can be represented as x = Ψf . Then, the compressed vector can be expressed as:

$$y_{CS} = \Phi \Psi f = Af \, . \qquad (3)$$

The sensing matrix, A, is typically full rank and should satisfy the RIP [67].

The main challenge of CS is reconstructing the original signal x from the measurement vector y_{CS} according to (2), which is an under-determined problem. Linear inverse problems attract a lot of attention throughout engineering and mathematical sciences. In most applications, these problems are under-determined, so one must apply additional regularizing constraints in order to obtain interesting or useful solutions. Sparsity constraints have emerged as a fundamental type of regularization [68].

The CS recovery problem can be modeled as in (4). The objective is to recover f by the knowledge of y_{CS} by satisfying the following condition:

$$\hat{f} = argmin\big|\big|f\big|\big|_0, s.t \; y_{CS} = Af \qquad (4)$$

As well known, this problem is NP-hard [68, 69] and needs a combinatorial search. Thus, to solve this problem, substituting the l0-norm by the closest convex norm, which is the l1 norm is proposed by Chen, Donoho, and Saunders in [70]. This solution can be formulated as:

$$\hat{f} = argmin\big|\big|f\big|\big|_1, s.t \; y_{CS} = Af \qquad (4)$$

During the last decades, various types of recovery algorithms were suggested including the convex optimization and greedy algorithms. The convex optimization algorithms try to convert the non-convex problem into a convex one and then an approximate solution is obtained [46–48,71]. Despite the success of convex optimization to solve the CS recovery problem, it still requires a very high computational complexity. Alternatively, greedy algorithms have been proposed to reduce the computational complexity of convex optimization algorithms. The greedy algorithms, for example, include matching pursuit [49], orthogonal matching pursuit [50]. Despite the low computational complexity of the greedy algorithms compared to the convex optimization but the reconstruction quality is low. The authors of [72] proposed a non-local tensor sparse and low-rank regularization (NTSRLR) approach, which can encode essential structured sparsity of an HSI and explore its advantages for HSI-CSR task. In addition, the authors of [73] proposed the first effort to characterize the spatial and spectral knowledge using the structure-based sparsity prior. Specifically, they introduce the non-local low-rank matrix recovery model and the hyper-Laplacian prior to encode the spatial and spectral structured sparsity, respectively.

With the huge success of DL, recently, there are numerous DL-based methods that have been proposed for image CS reconstruction [74–78]. The authors in [77] have presented a stacked denoising autoencoder (SDA) which is able to learn the statistical dependencies between the different elements of certain signals. In this way, the signal recovery performance is improved. One of the drawbacks of the proposed SDA is that SDA suffers from the high computational complexity with increasing the signal dimension. This is mainly due to its architecture that includes a full connection between any two successive layers. By utilizing the weight sharing method, the authors in [78] presented a CNN-based reconstruction method (ReconNet). This network is able to reduce the computational complexity. Inspired by the iterative shrinkage-thresholding, the authors in [76] proposed CNN (ISTA-Net) for CS reconstruction. To this end, we can conclude that the DL-based methods run faster than the traditional image CS methods. Motivated by that, in this article, we focus on utilizing the DL-based methods to compress and recover the original images.

4. The Proposed Deep Feature

The main goal of our proposed robust CS- SR network is to compress the original image and mitigate the effect of geo- metric transformation, simultaneously. Furthermore, our pro- posed framework is able to generate HR images that are similar to the original images from its compressed transformed LR images. Fig.2 shows the flow chart of our proposed framework. Our proposed framework is able to achieve this task through the mitigation of the spatial transformer effect for distorted LR images. Specifically, the proposed frame- work is able to perform three tasks including (1) generating an optimized measurement matrix, (2) mitigating the geo- metric transformation effects, and (3) recovering HR images from its compressed version. As seen in Fig.2, our proposed framework includes three stages starting by feeding the trans formed input to the compression network which utilizes CS to compress the input. Then, the output of the compression stage is passed through

the feature transformer network (FTN). The FTN is able to mitigate the geometric transformation effects by estimating the transformation parameters efficiently. Finally, the output images from the FTN are passed through an SR/CS recovery network. This network is able to generate HR images from its compressed LR version similar to the original images. To this end, an overview of our proposed framework is provided. The remaining part of this section includes more specific details of the compression network, the feature transform network, and the SR/CR recovery network.
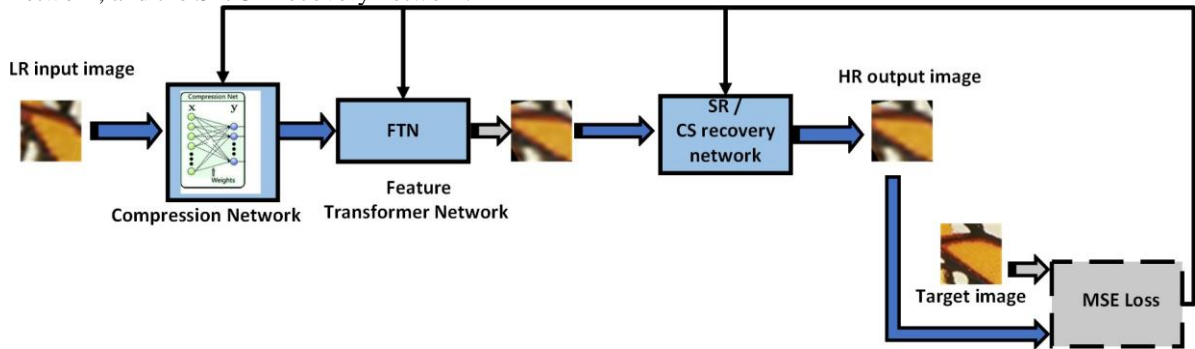

Fig. 2: Flow chart of our proposed framework

4.1 Compressed Sensing Network

The structure of our proposed framework is shown in Fig. 3. As shown in Fig.3, the first stage of the proposed framework is the compression network. In other words, the sampling matrix, which is used for acquiring CS measurements, is learned and optimized by the first stage. In the compression network, the image is first converted into one-dimensional vector with size (W* H *C) x1, where W , H, and C are the width, height, and channel of the image. Then, a fully connected (FC) layer with M neurons is adopted to compress the image vector into a lower dimension. The weights of the FC layer are considered as the CS sampling matrix Φ. During the training step, the sampling network with the training images learns the sampling matrix. The characteristics of images can be utilized by the learned sampling matrices, and hence more image structural information can be represented in the CS measurements. Thus, the quality of the reconstructed images is in- creased. Furthermore, this learned sampling matrices can be employed to generate CS measurements. Finally, the com- pressed image vector is converted into a 3D compressed image.

4.2 Feature Transformer Network

The goal of our proposed framework is to overcome the spatial transformation effects. This goal can be achieved through the mitigation of the transformation effects and then estimating the geometric parameters efficiently. To estimate the geo- metric parameters, we propose a feature transformer network (FTN) as shown in Fig.3. Our proposed FTN includes three main modules including localization network, grid generator, and sampling grid. Our proposed FTN preserves the essential spatial transformer structure and the working principle. How- ever, we have redesigned the localization network to improve the estimation of the geometric transformation parameters. In addition, our proposed FTN estimates these parameters with deep features rather than pixels.

The main contribution of our proposed framework is adding the proposed FTN as shown in Fig.3. The FTN is considered a multi-scale deep feature mapping. In addition, to refine the features, the feature refinement unit (FRU) is added into the localization network inside the FTN. Further, we have redesigned our proposed framework to work in the feature domain rather than the pixel domain. To ex- tract the deep features from the input image, we propose firstly passing the images through a VGG19 network. Then, to refine the extracted features, they are passed through the FRU. Structurally, FRU consists of 18 layers to provide a multi-channel feature maps. Thus, in this way, the generated multi-channel feature maps have sufficient details for estimating the optimized transformation parameters in feature domain. Therefore, we ensure that our proposed framework performs the spatial transformations in deep features rather than pixels.
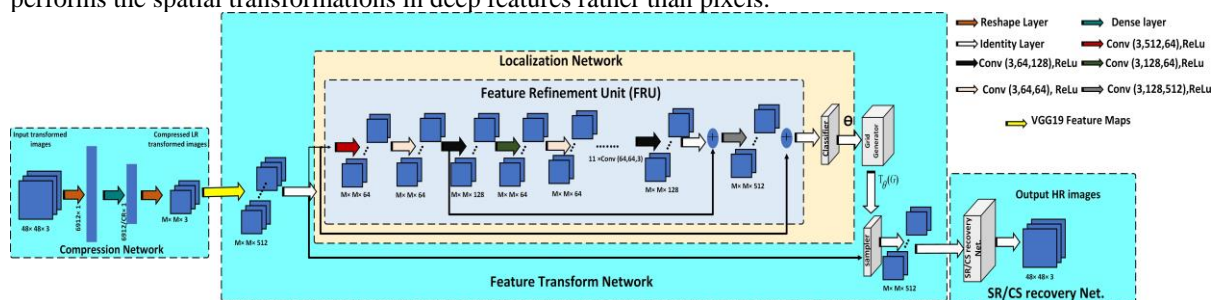
Fig. 3: Our proposed framework architecture: Conv(c_i, n_i ; k_i), where the variables k_i, n_i , c_i represent the number of feature maps, the number of filters and filter size, respectively. M , CR are the compressed image size, and the compression ratio, respectively.

The operation of FTN as shown in Fig.3 can be illustrated in various steps. First, the input image is transformed into the feature domain from the pixel domain by passing through the VGG feature extraction network. Specifically, in our pro- posed FTN, we propose to use the VGG 19 [79] network to extract the deep features. By utilizing several convolution layers together with max-pooling layers, the VGG 19 is able to extract the features and increase the input image features into 512 instead of 3 features. These 512 features are utilized by FRU to estimate the transformation parameters in the feature domain. Then, by utilizing the estimated parameters, our pro- posed FTN interpolates the input feature domain. The operation of FTN can be formulated as provided below:

$$F(x_{CS}) = VGG_{19}(x_{CS}, s) = F_{x_{CS,s}} \tag{6}$$

where the compressed distorted input can be represented by x_{cs}, V GG19 represents the VGG19 network to extract deep features, and s stands for the scale of the deep feature maps. To generate an HR image similar to its ground truth and mitigate the geometric transformation effects, we design the cor- responding loss function. This loss function can be illustrated as:

$$loss = argmin \left\|DF\left(F_{x_{CS}}, s\right) - x_{HR}\right\|_2^2 \tag{7}$$

where xHR is the ground truth, and the output of the deep feature transformation framework can be represented by DF (Fxcs,s). Then, the localization network in Fig.3 can be driven by Fxcs,s, the extracted deep feature maps and estimating the feature transformation parameters. This step can be represented as:

$$\hat{\theta}_F = L(F_{x_{CS}}, s) \tag{8}$$

where $\hat{\theta}_F$ stands for the estimated transformation parameters by using deep feature maps. Then, the localization network details with FRU are described as follows:

$$\hat{\theta}_F = L_{FRU}\left(F_{x_{CS}}, s\right) = FC\left(Conv\left(conv\left(\ldots\left(F_{x_{CS}}, s\right)\right)\right) + F_{x_{CS,s}}\right) \tag{9}$$

where Fxcs,s are the input features. To estimate the feature transformation parameters $\hat{\theta}_F$, the input features maps are refined by passing through convolutional layers to extract hierarchical levels of deep feature details. Then, the output of the last convolutional layer is added to the input according to the residual learning. The output of the summation step is fedinto the final fully-connected layer, i.e., classifier, to get $\hat{\theta}_F$.

Structurally, as shown in Fig.3, FRU consists of 18 convolutional layers. The first 3 convolutional layers are obtained from the pre-trained VGG layers and the other 15 layers are from ResNet [56] layers. The concept behind choosing some layers from VGG is ensuring a smooth transition from VGG- based deep feature extraction to feature refinement.

The estimated feature transformation parameters are fed into the transformation network TF, to obtain an estimate for the transformed feature maps Fxcs . To ensure restoration of the desired output dimension of the transformed feature map, the de-convolutional layers are added. Details of the operations are described as:

$$F_{x_{CS}} = T_F(F_{x_{CS}}, s , \hat{\theta}_F, s) \tag{10}$$

where s controls the deconvolutional layers

4.3 SR/CR Recovery Network

The CS theory states that if the image is well sparsely represented in a specific domain then the image can be correctly recovered from the CS measurements. So, we propose to re- cover the image by utilizing a CR recovery network. Our recovery network is able to achieve two tasks: 1) recovering the original image from its compressed version 2) generating HR image from its recovered version. In our proposed framework, we apply the output of our proposed FTN module to the image super-resolution module (VDSR), acting as a CS recovery network and a preprocessing unit to generate an HR image. Such a combination of FTN+VDSR not only pro- vides a robust single image super-resolution, but also converts the multi-channel deep feature maps back into pixel domain ridging the gap between the deep features and applications that require their output to be in pixel form.

The operation of the SR/CR recovery network can be explained as follows. First, the recovery network receives the output from the FTN, and the initial reconstructed image is obtained in the feature domain. Then, the VDSR network is used to generate HR images from its initial version. Correspondingly, our proposed deep feature transformation frame- work can be described via

$$\hat{x}_{HR} = Conv\left(conv\left(...\left(F_{x_{CS}}, s\right)\right) + F_{x_{CS}, s}\right) \tag{11}$$

where $\hat{x}_{HR}$ represents the output transformed images in pixel domain through the feature transformation

4.4 Loss Function

The main goal of our proposed framework is to create a robust compressed sensing SR network that is able to: 1) compress the original signal according to the CS theory. specifically, it is able to generate a compressed vector which is an optimized representation to the original signal; 2) alleviate the effect of spatial transforms for corrupted LR images by introducing the FTN module; 3) recovering and generating HR images from compressed LR transformed images simultaneously. Consequently, our proposed framework generates the optimized measurement matrix that is utilized for generating the optimized compressed vector y_{CS} . In addition, our pro- posed framework not only minimizes the effect of the geo- metric transformation effect but also minimizes the difference between the estimated HR image and the HR image itself. Specific details are described as follows.

$$L(\theta) = argmin\left\|\left\|DF\left(F_{x_{CS}}, s\right) - x_{HR}\right\|\right\|_2^2 \tag{7}$$

where L (θ) is a loss function (i.e., objective function), and θ represents the model parameters of the deep neural network. The loss function in 12 can be can be established via three operational steps. First, our proposed framework needs to generate the optimized measurement matrix A that is used to generate an optimized y_{CS} vector. second, our proposed framework needs to identify the affine transformation parameters, in order to capture any possible geometric transformation, through minimizing the errors between the HR and the distorted image. Finally, we generate an estimated HR image similar to the desired one through minimizing their corresponding MSEs. As a result, our loss function can be further formulated as follows:

$$L(\theta) = argmin_{\theta_A}\left\|\left|y_{CS} - y_{ocs}\right\|\right|_2^2 + argmin_{\theta_{FTN}}\left\|\left|x_{FTN_{out}} - \hat{x}_T\right\|\right|_2^2 + argmin_{\theta_{SR/CR}}\left\|\left|N(x_{CS}, \theta_{FTN}) - x_{HR}\right\|\right|_2^2 \tag{8}$$

where y_{ocs} is the optimized compressed vector, $\hat{x}_T$ is the out- put image after performing the spatial transformation, $\hat{x}_{HR} = N(x_{CS}, \theta_{FTN})$ is the estimated HR image, θF T N is the model parameters of the super-resolution neural network, and x^T is the output of a deep residual learning based spatial trans- form module. Additionally, θA, θF T N , and θSR/CR represent the compression network parameters, the estimated geometric/affine transformation parameters, and the SR/CR network parameters, respectively.

Specifically, the first part of 13 minimizes the error be- tween the compressed vector and its optimized version. correspondingly, this part is used to obtain an optimized measurement matrix. The second part of the equation is to minimize the error between the transformed LR image and the desired HR image for the super-resolution to estimate the affine trans- formation parameters and mitigate the transformation effects. The last part of the equation is to minimize the error between the output of the FTN module and the HR image to obtain an image similar to the desired HR image.

5. Evaluations and Experimental Result Analysis

Extensive experiments have been carried out to evaluate the performance of the proposed framework. In addition, we re- port our experimental results as well as their analyses. Firstly, our proposed FTN is compared with the existing STN over the number of modeling parameters. This comparison shows that the proposed FTN overwhelms the existing STN in terms of computing cost and learning complexity. Secondly, to validate the effectiveness of our proposed framework, we have carried out a number of experiments of computer vision ap- plications. To show the capability of our proposed framework to solve real-world problems, we apply our proposed framework to one popular computer vision task, i.e. single image super-resolution (SISR). Then, the following simulation results show that our proposed framework is a powerful learning tool, which is able to simultaneously handle geometric transformations of the compressed images and resolution enhancement for SISR applications.

5.1 Experiment Setup

5.1.1 Dataset For Training and Testing

For the training and testing of the proposed framework, we have utilized various datasets. To evaluate the proposed framework in solving the SISR computer vision problem, we have used images from the [80] and [81]. We augment the training data by rotation, scaling, and mirroring to in- crease the size of the training dataset. We have utilized the bi-cubic down-sampling and resized the images into the size of 48×48.

5.1.2 Implementation Details

We first simulate the effect of the geometric transformations as in [28,40,82] to test the proposed framework with different geometric transformation effects and under various compression ratios. Specifically, the

transformed LR training images are generated by four various transformations, including: (i) The rotation effect (R), where the original image is rotated clockwise by 20 degrees; (ii) the original image is scaled with a factor of 0.5, the scaling effect represented by S; (iii) the effect of both rotation and scaling represented by RS; (iv) translation represented by T, in which the LR images are translated by 5 pixels in both X and Y directions, and finally (v) combi- nation effect of rotation, scaling, and translation represented by RTS. Then, the transformed images are compressed by utilizing various compression ratios including 50%, 60%, 70%, and 80%.

5.1.3 Training Details

During the training, our proposed framework is optimized by utilizing the stochastic gradient descent (SGD) optimization algorithm. The SGD algorithm uses a learning rate of 0.001 with no learning rate decay. We train all experiments by set- ting the epochs value to be 20 and the batch size is set to 25. Furthermore, our proposed framework is trained in an end-to- end manner. All the training phase is performed on NVIDIA Tesla P100. The evaluation results of all experiments are presented in terms of two widely used metrics in the SR research community, which are peak-signal-to-noise-ratio (PSNR) and Structural Similarity Index Measurements (SSIM).

5.1.5 Benchmarks Comparisons

The performance of our proposed framework is validated by comparing it with the existing state-of-the-art benchmarks. Specifically, we compare our proposed framework with existing STN [52]. Structurally, VDSR is added with our pro- posed framework acting as a post-processing unit, to improve the quality of the output image.

5.2 Computational Complexity Evaluation

We follow the design of the STN presented in [83], where STN is used to mitigate the transformation effect for the traffic line signs. Specifically, the first convolutional layer is de- signed to generate 200 feature maps. Then, in the second convolutional layer, these feature maps are increased to 300. These feature maps are used to estimate the geometric trans- formation parameters. For our proposed framework, we utilize 16 convolutional layers to generate 64 feature maps, and two layers are used to generate 128 feature maps, and one layer to generate 512 feature maps. Table 1 shows the comparative results between our pro- posed framework and the existing STN in terms of the number of model parameters. We represent the convolutional layer by Conv(k_i; c_i; n_i) where the variables k_i, n_i , c_i represent the number of feature maps, the number of filters, and filter sizes, respectively. In addition, FC (m_i, o_i) represents the fully- connected layer where the variables m_i, o_i stand for the size of the input vector, and the size of the output vector, respectively.

The results in Table 1 show that our proposed framework is more powerful in structure, more cost-effective in learning, and more capable in capturing contextual information from input images, compared to the existing STN. Moreover, it requires less number of parameters to be tuned compared to STN.

Since the processing time is not only affected by the number of parameters but also dependent on the floating-point operations (FLOPs), we have compared the FLOPs number of our proposed framework with the existing STN. We have followed [84] for calculating the number of FLOPs. As well known, the convolutional and FC layers are the most computationally expensive parts in the network. These layers per- form huge numbers of multiplication and addition processes. Consequently, most of the FLOPs are consumed in these lay- ers. Specifically, the computational cost for the convolution operation is $k \times k \times n \times c \times W \times H$ where k, n, c, and W H denote the filter size, the number of the input fea- ture channels, the number of the convolutional filters, and the size of the input feature. Additionally, the computational cost of the FC layer is Nin Nout, where Nin and Nout are the numbers of the input and output neurons, respectively. There- fore, the number of FLOPs of the convolutional layers de- pends on the size of the input images. We have selected one of the compression ratios (e.g., CR=50%), and then we calculate the number of FLOPs of our proposed framework and the existing STN. Specifically, our proposed frameworks re- quire 616.21 million FLOPs at CR =50%. On the other side, the existing STN requires 4209.95 million FLOPs at the same CR. Therefore, it is clear that STN requires a higher number of FLOPs as its convolutional layer generates a huge number of feature maps. As noticed from Table 1, we can see that the STN localization network includes three convolutional layers that generate 200, 300, 200 features maps, respectively. As mentioned in Section 4.1.3 that we are training our model on a powerful GPU, namely Tesla P 100 for PCIe, it is able to perform a few teraFlops per second. Therefore, our training process can be performed easily. In general, fortunately, the neural networks can be first off-line trained using powerful GPUs and then online fine-tuned, which can be accelerated by techniques, e.g., meta-learning.

5.3 Experiments on Effectiveness and Robustness of Our Proposed Network

To validate the performance of the proposed framework, ex- tensive experiments have carried out . First, we compare the performance of our proposed framework with the VDSR [21]. Second, we have carried out

various experiments to compare the performance of our proposed framework with the existing spatial transformer network (STN).

The robustness of our proposed framework is tested against various transformation effects and different compression ratios. We transformed the original images with simulated geometric transformations that are mentioned previously in Sec.4.1.2, including rotation (R), scaling (S), rotation and scaling (RS), translation (T), and combination of rotation, scaling, and translation (RTS). In addition, we have compressed the original images using various compression ratios, e.g., 50%,60%,70%, and 80%. To this end, we have evaluated the performance of our proposed framework in five experiments compared to the existing state-of-the-art STN networks [52]. Then, we show the ability of our proposed framework to recover HR images from its compressed LR images.

Table 1: Comparison between the proposed FTN and the existing STN in terms of the number of parameters.

Layers	STN [52]	Parameters	Proposed DFTN	Parameters
Preprocessing	Max Pooling (2,2)	/	Max Pooling (2,2)	/
Feature Extraction	1×Conv(5,200,3) ReLU MaxPooling(2,2)	15200	2×Conv(3,512,64) ReLU	589952
	1×Conv(5,300,200) ReLU MaxPooling(2,2)	15180300	14×Conv(3,64,64) 1 ReLU	516992
			×Conv(3,64,128) ReLU	73856
			1×Conv(3,128,64) ReLU	73792
			1×Conv(3,128,512) ReLU	590336
Fully Connected layer	FC(2700,200)	540000	FC(294912,30)	8847360
Classifier	FC(200,6)	1200	FC(30,6)	1800
Total parameters		16276700		9344088

Table 2: The experimental results achieved by the existing STN compared with our proposed framework in terms of PSNR/S- SIM values under the Rotation effect.

	50%	60%	70%	80%
STN	12.96/0.6292	12.18/0.6279	15.07/0.5411	15.67/0.5651
Existing STN+VDSR	16.26/0.5549	16.64/0.5508	13.06/0.6256	16.46/0.5598
Proposed framwork	27.11/0.9129	f26.66/0.8955	26.56/0.8958	25.91/0.884

Table 3: The experimental results achieved by the existing STN compared with our proposed framework in terms of PSNR/S- SIM values under the Scaling effect.

	50%	60%	70%	80%
STN	16.30/0.5703	16.72/0.5932	12.11/0.6334	12.20/0.6318
Existing STN+VDSR	16.29/0.5630	16.66/0.5612	13.06/0.6256	16.34/0.5594
Proposed framework	26.26/0.8988	25.89/0.8876	25.33/0.8779	24.98/0.8729

Table 4: The experimental results achieved by the existing STN compared with our proposed framework in terms of PSNR/S- SIM values under the Rotation and scaling (RS) effects.

	50%	60%	70%	80%
STN	16.56/0.5803	16.27/0.5464	15.66/0.5532	15.62/0.5627
Existing STN+VDSR	16.08/0.5492	16.55/0.5485	13.32/0.6208	16.30/0.5547
Proposed framwork	26.12/0.8943	25.60/0.8845	25.44/0.8761	24.83/0.8653

Table 5: The experimental results achieved by the existing STN compared with our proposed framework in terms of PSNR/S- SIM values under the translation (T) effect.

	50%	60%	70%	80%
STN	16.40/0.5607	12.23/0.6094	11.86/0.6064	16.18/0.5255
Existing STN+VDSR	15.79/0.5265	16.34/0.5372	18.10/0.7146	15.95/0.5328
Proposed framework	26.92/0.9147	26.61/0.9034	26.19/0.8952	25.50/0.8838

Table 6: The experimental results achieved by the existing STN compared with our proposed framework in terms of PSNR/S- SIM values under the RTS effect.

	50%	60%	70%	80%
STN	14.97/0.6951	15.69/0.5405	15.22/0.5568	18.15/0.7032
Existing STN+VDSR	15.87/0.5439	16.54/0.5559	15.86/0.5439	12.54/0.6039
Proposed framework	25.56/0.8862	25.03/0.8689	24.87/0.8659	24.59/0.8587

Structurally, we add the VDSR with the existing STN to formulate a new benchmark, referred to as STN-VDSR. Tables 2, 3, 4, 5, and 6 show experimental for the existing STN- VDSR and our proposed framework under a number of trans- formation effects and various CRs. From the results shown in Table 2, we can see that the performance of our proposed framework outperforms the existing STN in terms of both PSNR and SSIM.

The simulation results show that the PSNR values of our proposed framework are higher than the existing framework by 7 dB in all test datasets and different CRs. This achievement is due to the fact that our proposed framework transfer the input image to the feature domain. Therefore, utilizing the feature extraction network facilitates the extraction of deep features from the input images. Furthermore, these features are used to estimate the geometric transformation parameters efficiently.
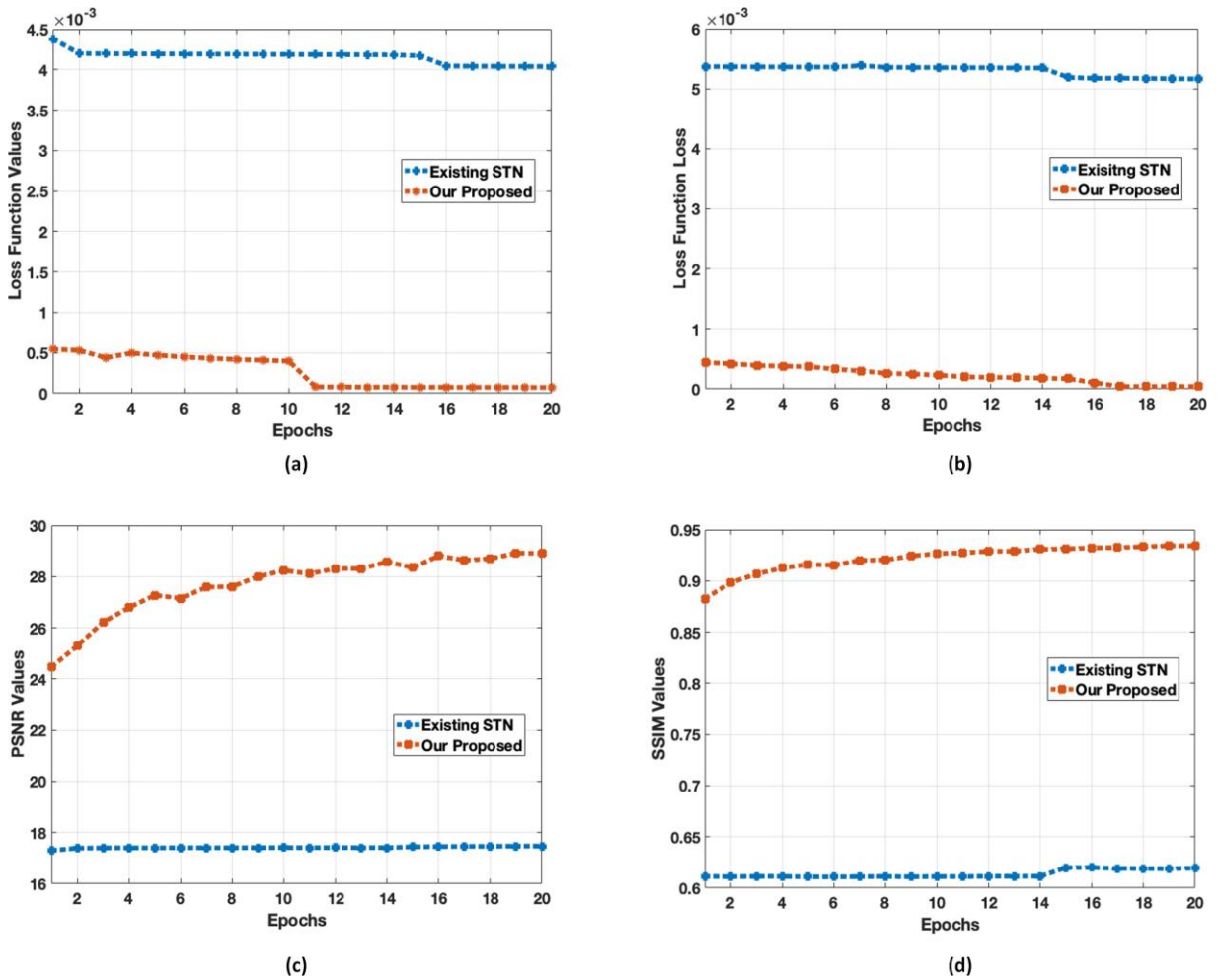
5.3.1 Experiment on Convergence

We also test the convergence of the proposed framework ac- cording to the loss function (7) compared to the STN-VDSR and the convergence curves during the training, and testing steps are presented. In addition, the performance of our pro- posed framework is tested by calculating the PSNR/SSIM values during each epoch to compare these values with the existing STN.

The convergence of our proposed framework under the rotation(R) effect and CR=50%, and the corresponding PSNR/SSIM values are presented in Fig. 4. The results in Fig.4 indicate that our proposed framework has lower loss function values compared to the existing STN. Furthermore, the PSNR/SSIM curves validate the superiority of our pro- posed VDST in terms of both PSNR and SSIM values. To provide a comprehensive assessment, the convergence of our proposed framework under a stronger transformation effect which is a combination of rotation, translation and scaling (RTS) with CR=50% is tested, and it is compared with the existing STN as shown in Fig.5. From the results shown in Fig.5, it can be confirmed that our proposed framework has lower loss function values and higher PSNR/SSIM compared with the existing STN.

6. Conclusion

In this paper, we have presented enhanced Compressed Sensing based Deep Neural Network for Single Image Super Resolution of COVID-19 using X-ray Images. Our proposed framework has shown immunity against the weaknesses of the widely-adopted STN. Our proposed framework owns a number of features over the existing STN, including 1) our pro- posed framework is able to extract the content-characterized deep features out of the compressed input LR in feature do- main to estimate the geometric transformation parameters and control the super-resolution process; 2) our proposed FRU is able to refine the multi-channel feature maps such that more details and their hierarchies can be revealed and provided to achieve better construction of the output images as well as a higher level of robustness; 3) our proposed framework is able to enhance the construction of the output images with higher level of robustness. Extensive experiments and comparative analysis have

proven the powerful capability of the proposed DFTN to tackle complex computer vision problems in comparison with the existing STN, such as robust single image super-resolution.

Fig. 4: Experimental results for convergence test and performance test under the rotation effect and CR=50% : (a) Convergence test for loss function values during the training phase; (b) Convergence test for loss function values during the validation phase; (c) Performance test for PSNR values; (d)performance test for SSIM values.
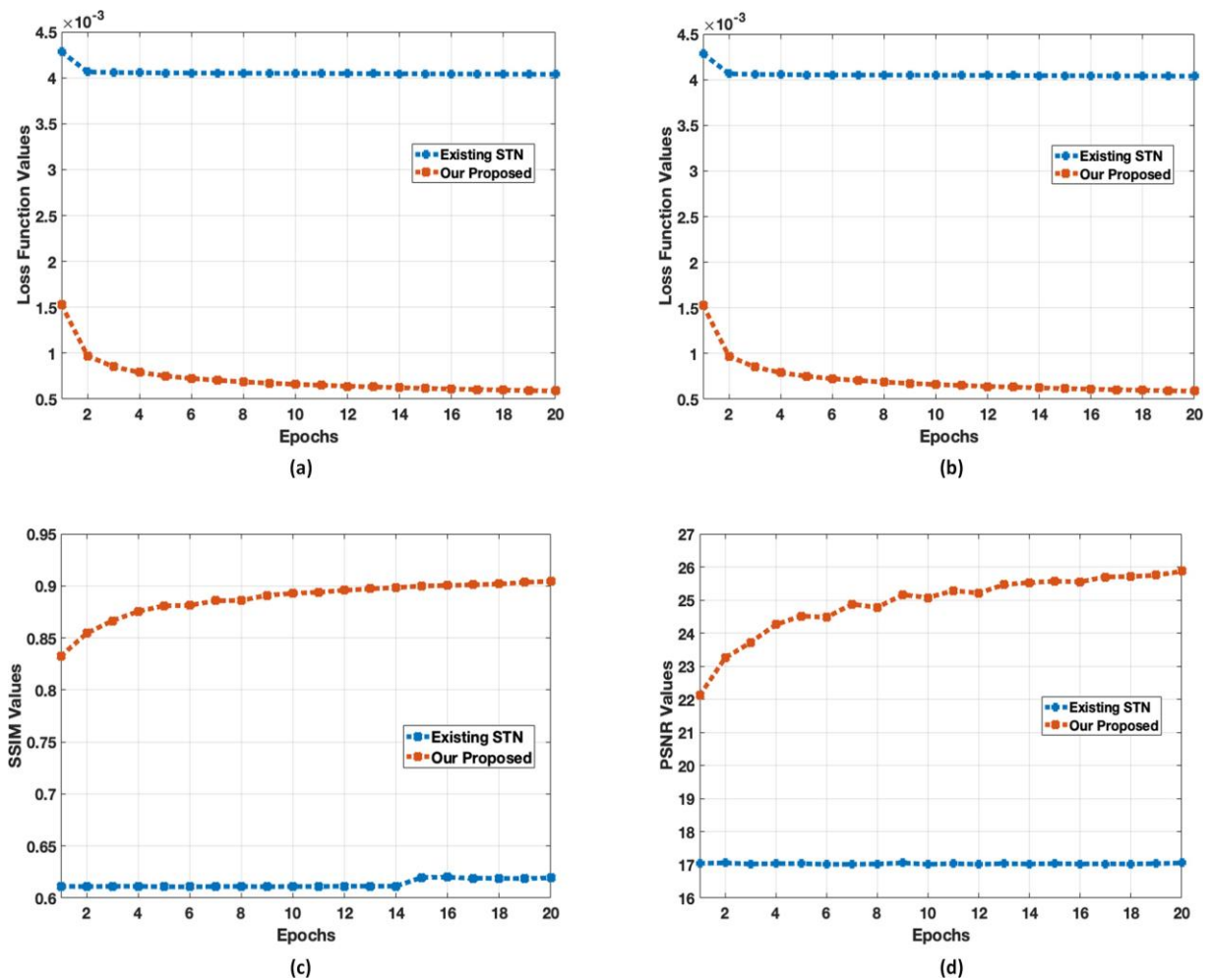
Fig. 5: Experimental results for convergence test and performance test under the RTS effect and CR=50% : (a) Convergence test for loss function values during the training phase; (b) Convergence test for loss function values during the validation phase; (c) Performance test for PSNR values; (d)performance test for SSIM values.

Disclosure

The author reports no conflicts of interest in this work.

References

[1]   David S Hui, Esam I Azhar, Tariq A Madani, Francine Ntoumi, Richard Kock, Osman Dar, Giuseppe Ippolito, Timothy D Mchugh, Ziad A Memish, Christian Drosten, et al., "The continuing 2019-ncov epidemic threat of novel coronaviruses to global healththe latest 2019 novel coronavirus outbreak in wuhan, china," International journal of infectious diseases, vol. 91, pp. 264–266, 2020.

[2]   Marco Cascella, Michael Rajnik, Abdul Aleem, Scott Dulebohn, and Raffaela Di Napoli, "Features, evaluation, and treatment of coronavirus (covid-19)," Stat- Pearls, 2022.

[3]   David L Donoho, "Compressed Sensing," IEEE Trans- actions on Information Theory, vol. 52, no. 4, pp. 1289– 1306, 2006.

[4]   Emmanuel J Candes and Terence Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies" IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5406–5425, 2006.

[5]   Chao Dong, Change Loy Chen, Kaiming He, and Xi- aoou Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE TPAMI, vol. 38, pp. 295– 307, 2016.

[6]   Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286–301.

[7]   Robert Keys, "Cubic Convolution Interpolation for Digital Image Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 6, pp. 1153–1160, 1981.

[8]   Claude E Duchon, "Lanczos Filtering in One and Two Dimensions," Journal of Applied Meteorology, vol. 18, no. 8, pp. 1016–1022, 1979.

[9]   Michal Irani and Shmuel Peleg, "Improving Resolution by Image Registration," CVGIP: Graphical Models and Image Processing, vol. 53, no. 3, pp. 231–239, 1991.

[10]   Gilad Freedman and Raanan Fattal, "Image and Video Upscaling from Local Self-examples," ACM Transactions on Graphics (ToG), vol. 30, no. 2, pp. 12, 2011.

[11]   Jian Sun, Jian Sun, Zongben Xu, and Heung Yeung Shum, "Image Super-Resolution Using Gradient Profile Prior," in IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[12]   Zhiwei Xiong, Xiaoyan Sun, and Feng Wu, "Robust Web Image/Video Super-Resolution," IEEE Transactions on Image Processing, vol. 19, no. 8, pp. 2017– 2028, 2010.

[13]   William T Freeman, Thouis R Jones, and Egon C Pasztor, "Example-Based Super-Resolution," IEEE Computer Graphics and Applications, no. 2, pp. 56–65, 2002.

[14]   Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super- Resolution Through Neighbor Embedding," in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, vol. 1, pp. I–I.

[15]   Daniel Glasner, Shai Bagon, and Michal Irani, "Super- Resolution From a Single Image," in 12th IEEE Inter- national Conference on Computer Vision. IEEE, 2009, pp. 349–356.

[16]   Qiang Zhang, Qiangqiang Yuan, Chao Zeng, Xinghua Li, and Yancong Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal– spectral deep convolutional neural network," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4274–4288, 2018.

[17]   Qiang Zhang, Qiangqiang Yuan, Jie Li, Xinxin Liu, Huanfeng Shen, and Liangpei Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 10, pp. 7317–7329, 2019.

[18]   Qiang Zhang, Qiangqiang Yuan, Jie Li, Zhiwei Li, Huanfeng Shen, and Liangpei Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatiotemporal patch group deep learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 162, pp. 148–160, 2020.

[19]   Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, "Image Super-Resolution as Sparse Representation of Raw Image Patches," in IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008, pp. 1–8.

[20]   Chao Dong, Change Loy Chen, Kaiming He, and Xiaoou Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," vol. 8692, pp. 184–199, 2014.

[21]   Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Ac- curate Image Super-Resolution Using Very Deep Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[22]   Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Deeply-Recursive Convolutional Network for Image

 Super-Resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.

[23]   Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced Deep Residual Net- works for Single Image Super-Resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.

[24]   Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[25]   Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 624–632.

[26]   Leon A Gatys, Alexander S Ecker, and Matthias Bethge, "Image style transfer using convolutional neural net- works," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2414–2423.

[27]   Justin Johnson, Alexandre Alahi, and Fei Fei Li, "Perceptual Losses for Real-Time Style Transfer and Super- Resolution," arXiv, vol. 1603.08155, pp. 694–711, 2016.

[28]   Jianmin Jiang, Hossam M Kasem, and Kwok-Wai Hung, "A Very Deep Spatial Transformer Towards Robust Single Image Super-Resolution," IEEE Access, vol. 7, pp. 45618–45631, 2019.

[29]   Alireza Esmaeilzehi, M Omair Ahmad, and MNS Swamy, "Compnet: A New Scheme for Single Image Super Resolution Based on Deep Convolutional Neural Network," IEEE Access, vol. 6, pp. 59963–59974, 2018.

[30]   Peng Liu, Ying Hong, and Yan Liu, "Deep Differential Convolutional Network for Single Image Super- Resolution," IEEE Access, vol. 7, pp. 37555–37564, 2019.

[31]   Feng Li, Huihui Bai, Lijun Zhao, and Yao Zhao, "Dual- Streams Edge Driven Encoder-Decoder Network for Image super-resolution," IEEE Access, vol. 6, pp. 33421– 33431, 2018.

[32]   Yifan Wang, Lijun Wang, Hongyu Wang, and Peihua Li, "End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks," IEEE Access, vol. 7, pp. 31959–31970, 2019.

[33]   Zhang Lu, Zhang Yu, Peng Ya-Li, Liu Shi-Gang, Wu Xiaojun, Lu Gang, and Rao Yuan, "Fast Single Image Super-Resolution via Dilated Residual Networks," IEEE Access, 2018.

[34]   Yang Zhao, Guoqing Li, Wenjun Xie, Wei Jia, Hai Min, and Xiaoping Liu, "GUN: Gradual Upsampling Net- work for Single Image Super-Resolution," IEEE Access, vol. 6, pp. 39363–39374, 2018.

[35]   Jae Sung Park, Jae Woong Soh, and Nam Ik Cho, "High Dynamic Range and Super-Resolution Imaging from a Single Image," IEEE Access, vol. 6, pp. 10966–10978, 2018.

[36]   Xiaotong Zhao, Wei Li, Yifan Zhang, and Zhiyong Feng, "Residual Super-Resolution Single Shot Network for Low-Resolution Object Detection," IEEE Access, vol. 6, pp. 47780–47793, 2018.

[37]   Liling Zhao, Quansen Sun, and Zelin Zhang, "Single Image Super-Resolution Based on Deep Learning Features and Dictionary Model," IEEE Access, vol. 5, pp. 17126–17135, 2017.

[38]   Hossam Kasem and Maha Elsabrouty, "Perceptual com- pressed sensing and perceptual sparse fast fourier trans- form for audio signal compression," in 2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2014, pp. 444–448.

[39]   Hossam M Kasem, Kwok-Wai Hung, and Jianmin Jiang, "Revised spatial transformer network towards improved image super-resolutions," in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 2688–2692.

[40]   Jianmin Jiang, Hossam M Kasem, and Kwok-Wai Hung, "Robust image completion via deep feature transformations," IEEE Access, vol. 7, pp. 113916–113930, 2019.

[41]   Hossam M Kasem, Mahmoud M Selim, Ehab Mahmoud Mohamed, and Amr H Hussein, "Drcs-sr: Deep robust compressed sensing for single image super-resolution," IEEE Access, vol. 8, pp. 170618–170634, 2020.

[42]   Weizhi Lu, Tao Dai, and Shu-Tao Xia, "Binary matrices for compressed sensing," IEEE Transactions on Signal Processing, vol. 66, no. 1, pp. 77–85, 2017.

[43]   Arash Amini and Farokh Marvasti, "Deterministic construction of binary, bipolar, and ternary compressed sensing matrices," IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2360–2370, 2011.

[44]   Xinwei Gao, Jian Zhang, Wenbin Che, Xiaopeng Fan, and Debin Zhao, "Block-based compressive sensing coding of natural images by local structural measurement matrix," in Data Compression Conference. IEEE, 2015, pp. 133–142.

[45]   Khanh Quoc Dinh, Hiuk Jae Shim, and Byeungwoo Jeon, "Measurement coding for compressive imaging using a structural measuremnet matrix," in IEEE Inter- national Conference on Image Processing. IEEE, 2013, pp. 10–13.

[46]   Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Communications on Pure and Applied Mathematics: A Journal Is- sued by the Courant Institute of Mathematical Sciences, vol. 57, no. 11, pp. 1413–1457, 2004.

[47]   Stephen J Wright, Robert D Nowak, and Ma´rio AT Figueiredo, "Sparse reconstruction by separable approximation," IEEE Transactions on Signal Processing, vol. 57, no. 7, pp. 2479–2493, 2009.

[48]   Ma´rio AT Figueiredo, Robert D Nowak, and Stephen J Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 4, pp. 586–597, 2007.

[49]   Ste´phane G Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397– 3415, 1993.

[50]   Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Transactions on Information Theory, vol. 53, no. 12, pp. 4655–4666, 2007.

[51]   Jarvis Haupt and Robert Nowak, "Signal reconstruction from noisy random projections," IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 4036–4048, 2006.

[52]   Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al., "Spatial Transformer Networks," in Advances in Neural Information Processing Systems, 2015, pp. 2017–2025.

[53]   K Lenc and A Vedaldi, "Understanding Image Representations by Measuring Their Equivariance and Equivalence," in Computer Vision and Pattern Recognition, 2015, pp. 991–999.

[54]   Jonathan Huang and Kevin Murphy, "Efficient Inference in Occlusion-Aware Generative Models of Im- ages," arXiv preprint arXiv:1511.06362, 2015.

[55]   Danilo Jimenez Rezende, Shakir Mohamed, Ivo Dani- helka, Karol Gregor, and Daan Wierstra, "One-Shot Generalization in Deep Generative Models," arXiv preprint arXiv:1603.05106, 2016.

[56]   Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[57]   Ross Girshick, "Fast r-cnn," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[58]   Kaiming He, Georgia Gkioxari, Piotr Dolla´r, and Ross Girshick, "Mask r-cnn," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.

[59]   Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[60]   Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4873–4882.

[61]   Junjun Jiang, Xiang Ma, Chen Chen, Tao Lu, Zhongyuan Wang, and Jiayi Ma, "Single image super-resolution via locally regularized anchored neighbor- hood regression and nonlocal means," IEEE Transactions on Multimedia, vol. 19, no. 1, pp. 15–26, 2016.

[62]   Chao Dong, Change Loy Chen, and Xiaoou Tang, "Accelerating the super-resolution convolutional neural net- work," pp. 391–407, 2016.

[63]   Dalal, Navneet, Triggs, and Bill, "Histograms of oriented gradients for human detection," in IEEE Computer Society Conference on Computer Vision and Pat- tern Recognition, 2005, pp. 886–893.

[64]   David G Lowe, "Distinctive image features from scale- invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[65]   Roberto Annunziata and Emanuele Trucco, "Accelerating convolutional sparse coding for curvilinear structures segmentation by refining scird-ts filter banks," IEEE Transactions on Medical Imaging, vol. 35, no. 11, pp. 2381–2392, 2016.

[66]   Angjoo Kanazawa, Abhishek Sharma, and David W. Ja- cobs, "Locally scale-invariant convolutional neural net- works," CoRR, vol. abs/1412.5104, 2014.

[67]   Emmanuel J Candes et al., "The restricted isometry property and its implications for compressed sensing," Comptes Rendus Mathematique, vol. 346, no. 9-10, pp. 589–592, 2008.

[68]   Emmanuel J Candes and Terence Tao, "Decoding by linear programming," IEEE Transactions on Information Theory, vol. 51, no. 12, pp. 4203–4215, 2005.

[69]   John Skilling and SF Gull, "Algorithms and applications," in Maximum-Entropy and Bayesian Methods in Inverse Problems, pp. 83–132. Springer, 1985.

[70]   Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.

[71]   Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," SIAM Review, vol. 43, no. 1, pp. 129–159, 2001.

[72]   Jize Xue, Yongqiang Zhao, Wenzhi Liao, and Jonathan Cheung-Wai Chan, "Nonlocal tensor sparse representation and low-rank regularization for hyperspectral image compressive sensing reconstruction," Remote Sensing, vol. 11, no. 2, pp. 193, 2019.

[73]   Jize Xue, Yongqiang Zhao, Wenzhi Liao, and Jonathan Cheung-Wai Chan, "Hyper-laplacian regularized non- local low-rank matrix recovery for hyperspectral image compressive sensing reconstruction," Information Sciences, vol. 501, pp. 406–420, 2019.

[74]   Hancheng Lu and Lei Bo, "Wdlreconnet: Compressive sensing reconstruction with deep learning over wireless fading channels," IEEE Access, vol. 7, pp. 24440– 24451, 2019.

[75]   Lei Bo, Hancheng Lu, Yujiao Lu, Jianwen Meng, and Wenzhe Wang, "Fompnet: Compressive sensing re-construction with deep learning over wireless fading channels," in 9th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2017, pp. 1–6.

[76]   Jian Zhang and Bernard Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1828–1837.

[77]   Ali Mousavi, Ankit B Patel, and Richard G Baraniuk, "A deep learning approach to structured signal recovery," in 53rd annual alert on conference on communication, control, and computing (Allerton). IEEE, 2015, pp. 1336–1343.

[78]   Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 449–458.

[79]   Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recogni- tion," arXiv, p. 1409.1556, 2014.

[80]   Sobhan Sheykhivand, Zohreh Mousavi, Sina Mojtahedi, Tohid Yousefi Rezaii, Ali Farzamnia, Saeed Meshgini, and Ismail Saad, "Developing an efficient deep neural network for automatic detection of covid-19 using chest x-ray images," Alexandria Engineering Journal, vol. 60, no. 3, pp. 2885–2903, 2021.

[81]   Francis Jesmar P Montalbo, "Diagnosing covid-19 chest x-rays with a lightweight truncated densenet with partial layer freezing and feature fusion," Biomedical Signal Processing and Control, vol. 68, pp. 102583, 2021.

[82]   Hossam M Kasem, Kwok-Wai Hung, and Jianmin Jiang, "Spatial transformer generative adversarial network for robust image super-resolution," IEEE Access, 2019.

[83]   Ueli Meier and Jonathan Masci, "Multi-column deep neural network for traffic sign classification," Neural Networks, vol. 32, no. 1, pp. 333–338, 2012.

[84]   Jiajia Guo, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li, "Convolutional neural network-based

multiple-rate compressive sensing for massive mimo csi feedback: Design, simulation, and analysis," IEEE

Transactions on Wireless Communications, vol. 19, no. 4,