



# Learning image compressed sensing with sub-pixel convolutional generative adversarial network

Yubao Sun, Jiwei Chen, Qingshan Liu\*, Guangcan Liu

Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China



## ARTICLE INFO

### Article history:

Received 9 January 2019

Revised 22 August 2019

Accepted 12 September 2019

Available online 13 September 2019

### Keywords:

Compressed sensing

Sub-pixel convolutional GAN

Compound loss

## ABSTRACT

Compressed sensing (CS) is a new technology to reconstruct image from randomized measurements, but the reconstruction procedure involves a time-consuming iterative optimization. In addition, the reconstruction quality becomes poor in low sampling rate. In order to alleviate these issues of the conventional CS image reconstruction, we propose a novel sub-pixel convolutional generative adversarial network (GAN) to learn compressed sensing reconstruction of images. The generator constructs the sub-pixel convolutional network to learn the explicit mapping from the low-dimensional measurement vector to the high-dimensional reconstruction, in which a compound loss, including reconstruction loss, measurement loss and adversarial loss, is designed to guide the network learning. By means of the adversarial training with discriminator, the generator can learn the inherent image distribution and improve the reconstruction quality. Moreover, the test image can be fast reconstructed by simply passing the low-dimensional measurement vector through the generator network. The proposed algorithm is tested on MNIST, F-MNIST and CelebA datasets, and the experimental results show that it is superior to some state-of-the-art deep learning based and iterative optimization based algorithms, in terms of both time complexity and reconstruction quality.

© 2019 Published by Elsevier Ltd.

## 1. Introduction

Natural images are full of various redundancies and enable the possibility of image compression and reconstruction, which is an important topic in many areas such as communication, data storage, computation etc. In terms of classical digital image sampling technology, an image is first acquired at Nyquist rate and then compressed in order to save storage space and reduce communication cost. Instead of this seemingly contradictory process, compressed sensing offers us an alternative to directly capture the compressed image with sampling rate below the Nyquist criterion [1]. CS technology has been applied in many practical applications [2,3], including single pixel camera, compressive magnetic resonance imaging (MRI), snapshot hyperspectral imaging and so on.

The theory of compressed sensing states that if an image  $x \in R^N$  has a sparse representation in some transform domain (such as DCT or wavelet), it can be acquired by taking the linearized random measurement,

$$y = \Phi x + \xi, \quad (1)$$

where  $\Phi \in R^{M \times N}$  is a random measurement matrix with  $M \ll N$ , the ratio  $M/N$  is called the measurement rate,  $y \in R^M$  is the obtained measurement vector, and  $\xi$  is the measurement noise. The core problem of CS is how to reconstruct the original signal  $x$  from the measurement vector  $y$  according to Eq. (1), which is an under-determined problem. By utilizing the sparsity of the image in some transform domain, the original image can then be approximately reconstructed by solving the following sparsity regularized optimization problem,

$$\min_x \|Ux\|_1 + \lambda \|y - \Phi x\|_2^2, \quad (2)$$

where  $U$  is the pre-defined transform operator such as DCT, wavelet, contourlet and total variation or dictionary adaptively learned from image set [4],  $\|\cdot\|_1$  is the  $l_1$  norm,  $\lambda$  is the regularization parameter. Some structured sparsity regularization can further exploit the cluster structure of non-zeros coefficients for reconstruction [5,6]. Provided the measurement matrix  $\Phi$  and the transform operator  $U$  satisfy the generalized restricted isometry property (RIP) [7,8], the signal  $x$  can be recovered reliably by solving Eq. (2).

**Related works:** The most common reconstruction algorithms leveraged the convex optimization for solving the  $l_1$ -minimization in Eq. (2), and many algorithms have been proposed [9]. Among them, iterative threshold shrinkage (ISTA) [10] is a well-

\* Corresponding author.

E-mail address: [qsliu@nust.edu.cn](mailto:qsliu@nust.edu.cn) (Q. Liu).

known algorithm for compressed reconstruction, which decomposes Eq. (2) into a gradient descent sub-problem and a threshold shrinkage sub-problem, iteratively solving these two sub-problems until meeting the convergence criteria. However, the convergence rate of IST becomes very slow when the measurement matrix  $\Phi$  is ill-conditioned or ill-posed. The two-step iterative threshold shrinkage (TwIST) [11] overcomes this shortcoming by implementing a nonlinear, second order iterative version of IST and exhibits a much faster convergence rate than IST. Similarly, a lot of endeavors have been made to propose more effective reconstruction algorithms, like approximate message passing [12], gradient projection [13], split bregman iteration [14], etc. However, all these reconstruction algorithms suffer from a time-consuming iterative optimization procedure, especially in the case of reconstructing the high-dimensional images. Meanwhile, when the measurement rate is very low, the condition number of measurement operator  $\Phi$  would be large and therefore lead to the deterioration of reconstruction quality.

Recently, deep learning has exhibited its superiority in many image processing problems, such as image enhancement [15], image super-resolution [16,17], image segmentation [18,19], etc. Deep learning has also been adopted to learn the compressed sensing reconstruction [20–22]. Existing deep learning based methods can be mainly classified into two categories. The first category is to design an elaborate network for learning the explicit mapping from the low dimensional compressed measurement vector  $y$  to the high-dimensional reconstruction  $x$ . A representative method is ReconNet that employs a tailor-made convolutional neural network for reconstruction in block-wise manner [23]. The reconstruction procedure of ReconNet is computationally light, as it only requires a single-pass forward computation through the trained network, which is about three orders of magnitude faster than traditional iterative method. At the same time, the reconstruction images quality produced by ReconNet is distinctly better than the classical methods [24,25] especially when the measurement rates are below 0.1. Some other networks, like DeepInverse [26] and DeepCodec [27], also follow this strategy for learning the non-iterative reconstruction mapping. The second category relies on learning the image prior model in terms of generative adversarial networks (GAN) which consists of a generator network  $G$  and a discriminator network  $D$  [28,29]. By means of the adversarial training, the generator  $G$  and the discriminator  $D$  can be mutually improved until the Nash equilibrium is achieved. Unlike the traditional methods that impose an sparsity assumption on the image  $x$ , the GAN model assumes the unknown signal  $x$  lies near the range of the pre-trained generative model  $G: R^K \rightarrow R^N$ , and the reconstruction is reduced to solve the following problem,

$$\begin{aligned} & \min_z \| \Phi \hat{x} - y \|_2^2 \\ & \text{s.t. } \hat{x} = G(z), \end{aligned} \quad (3)$$

where  $z \in R^K$  is low-dimensional random vector from the latent space,  $G(z) \in R^N$  is a generator network that produces the high-dimensional images. Eq. (3) could be solved by using gradient descent to optimize the latent code  $z$  such that the corresponding image  $G(z)$  matches the given measurement vector  $y$  with a minimal measurement error  $\| \Phi G(z) - y \|_2^2$ . Bora et al. [30] adopted Deep Convolutional GAN (DCGAN) [31] as the generative and discriminative model, and it showed that the reconstruction quality by DCGAN can significantly outperform the  $l_1$ -norm regularized convex optimization recovery. However, in [30], the generator  $G$  is not aware of the compressed sensing task, but only be trained for reproducing images that resemble the training dataset. To overcome this limitation, Kabkab et al. [32] proposed to train  $G$  specific to the task of CS recovery, which made the GAN task-aware and therefore enhanced the CS reconstruction performance. Although the adversarial learning is beneficial for capturing the image prior and

thus improving the reconstruction quality, the iterative solving of Eq. (3) is still computationally expensive compared with the non-iterative reconstruction in the first category. Meanwhile, the optimization error in finding the best  $z$  will degenerate the reconstruction quality.

**Our contributions:** In this paper, we propose the sub-pixel convolutional generative adversarial network termed SCGAN for the compressed sensing task. Our generative model learns the reconstruction  $\hat{x}$  from the measurement vector  $y$  rather than the latent code  $z$  in an adversarial manner. Without the need of optimizing the latent code  $z$  for reconstruction as done in [30,32], SCGAN can achieve non-iterative reconstruction of test images. The experimental results demonstrate that our model outperforms the conventional iterative CS reconstruction algorithms and some state-of-the-art deep learning based algorithms in terms of both reconstruction quality and computational cost upon the MNIST, F-MNIST and CelebA datasets. The main contributions of our work lie in three aspects.

- (1) The proposed SCGAN learns an explicit mapping from the compressed measurement  $y$  to the reconstruction  $\hat{x}$  in an adversarial manner, thereby the reconstruction to any testing sample can be obtained by simply feeding the measurement vector into the well-trained generator  $G$ , improving significantly the reconstruction quality and speeding up the CS reconstruction procedure greatly.
- (2) Our generator network uses multiple sub-pixel convolutions to progressively upscale the dimension of feature maps until reaching the dimension of the original image. The sub-pixel convolutions can extract more feature maps for resolution upscaling and thus promote the reconstruction quality.
- (3) We design a compound loss function, consisting of CS task-oriented reconstruction loss, measurement loss term and Wasserstein adversarial loss term, to encourage the output of generator to have a similar statistical distribution as the real images.

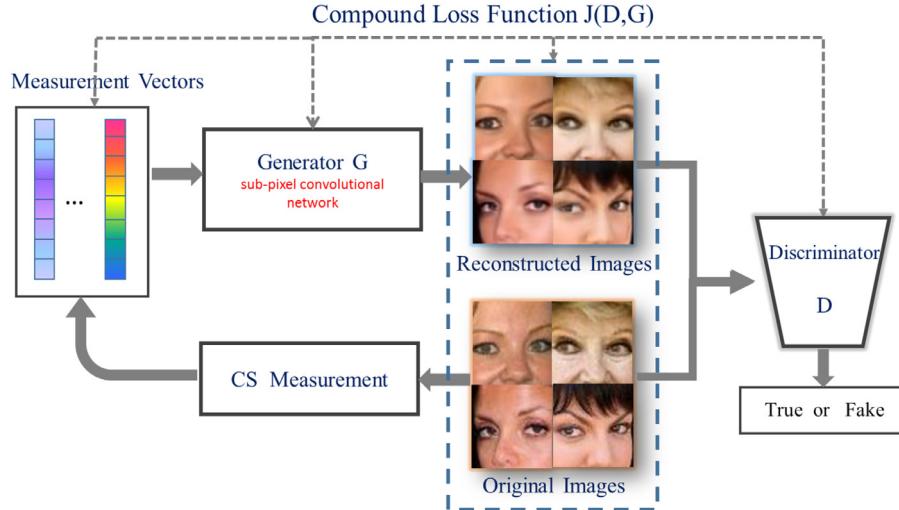
The reminder of the paper is organized as follows. Section 2 develops our method, including the design of Sub-pixel Convolutional GAN and the associated compound loss function for guiding the network parameters learning and model training algorithm. Experimental results and analysis are presented in Section 3, and Section 4 summarizes the paper.

## 2. Compressed sensing oriented sub-pixel convolutional GAN

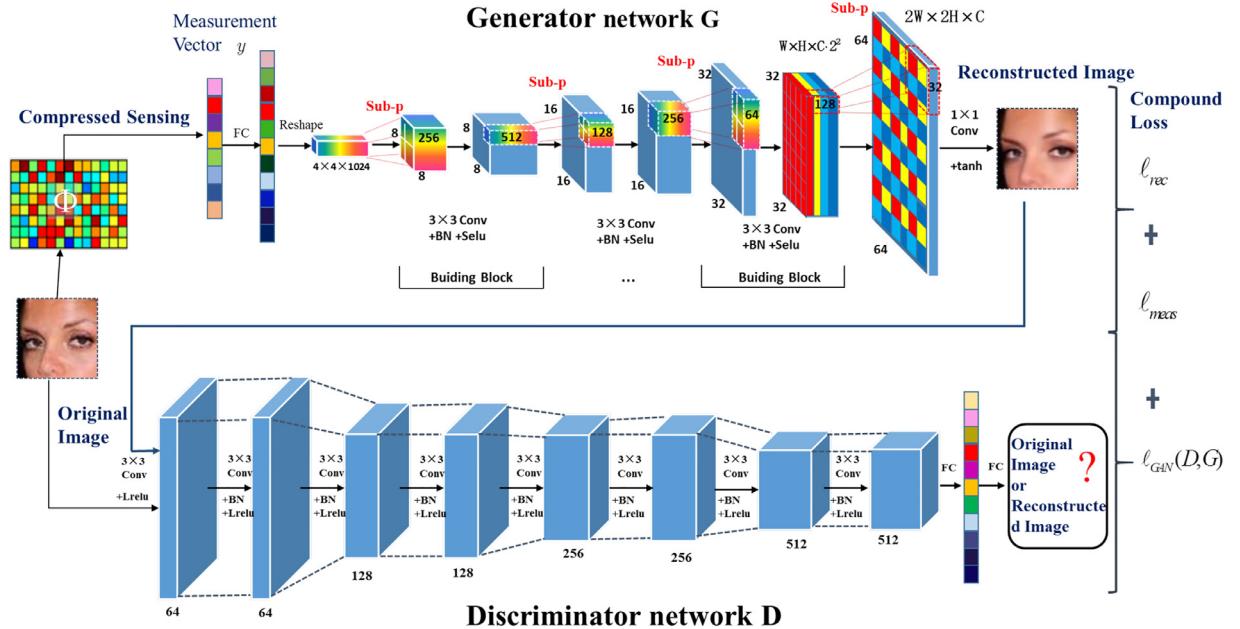
In this section, we propose a novel sub-pixel convolutional generative adversarial network for non-iterative compressed sensing reconstruction of images. Fig. 1 presents the framework of our proposed model. Different from the existing GAN approaches [30,32] taking the random noise  $z$  as the latent code, our generator admits the measurement vector  $y$  as the latent code and is engrossed in learning the explicit map from  $y$  to the high dimensional image space of  $x$ . The discriminator  $D$  is set as a binary neural network classifier that discriminates whether the input is a real image or fake image generated by  $G$ . The CS task-driven compound loss function is also designed to conduct the learning of  $D$  and  $G$  in an alternating fashion. Guided by this compound loss and well-designed network architecture, the generator  $G$  can be well-trained for generating the reconstruction image from the compressed measurement.

### 2.1. SCGAN network architecture

The proposed sub-pixel convolutional generative adversarial network learns CS reconstruction under the guidance of discriminator. The network architectures of the generator and discrimi-



**Fig. 1.** Illustration of our CS oriented Sub-pixel Convolutional GAN model.

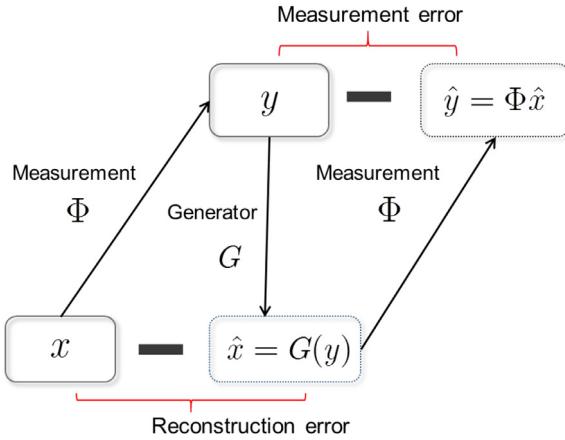


**Fig. 2.** The network architecture of our sub-pixel convolutional generative adversarial network for CS reconstruction of images, in which the arrowheads denote the operations in each layer and the three-dimensional cubes represent the generated feature maps with the spatial size and the number of feature maps labeled along the axes. The generator network  $G$  (the upper part) and discriminator network  $D$  (the bottom part) are trained under the compound loss function.

nator are shown in Fig. 2. The deep sub-pixel convolutional network is set as the generator. The input to the generator is the  $M$ -dimensional compressive measurements  $y$ . The first layer is a fully connected layer that takes compressive measurements  $y$  as the input and reshapes the outputs as 1024 feature maps with the dimension size  $4 \times 4$ . The following layers stack multiple building blocks consisting of sub-pixel convolutions (sub-p),  $3 \times 3$  convolutions, batch normalization (BN) and scaled exponential linear unit (SELU) activation, in which sub-pixel convolutions are responsible for increasing the resolution of the feature map layer by layer. The last layer of our generator is the  $1 \times 1$  convolution for reducing the feature maps to a single channel for gray image or three channels for color image. The final output is also undergone the processing of the tanh activation, such that the pixel values of final reconstruction image can be restricted to the range of  $-1$  and  $1$ . In the

entire generator network, the filter size for normal convolution is  $3 \times 3$  and the number of stride is 1. In addition, BN is used to accelerate deep network training by scaling and shifting the value of activation over a mini-batch and SELU [33] is also adopted to output the normalized activations to the next layer. In order to prevent the network from overfitting, the convolutional dropout regularization is also used in the feature maps of the final layer with a dropout probability 0.5.

Compared with DCGAN, the main strength of our architecture is the use of sub-pixel convolutions rather than the transposed convolution for resolution up-sampling. The commonly-used transposed convolution upscales the image by inserting zero values and later filling in with meaningful values. These zero values have no gradient information that can be backpropagated through. As the statement in [34], sub-pixel convolutions essentially use the regu-



**Fig. 3.** The CS task driven image fidelity constraint.

lar convolutional layer in low-resolution feature map followed by a phase shift (PS) operation to upscale their resolution. Specifically, PS is an periodic shuffling operator that rearranges the elements of low resolution feature maps  $f^{L-1}$  of shape  $W \times H \times Cr^2$  to high resolution features map  $f^L$  of shape  $rW \times rH \times C$ . This operation can be mathematically described as

$$f_{x,y,c}^L = PS(f^{L-1})_{x,y,c} = f^{L-1}_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, c \cdot r \bmod (y,r) + c \bmod (x,r)}, \quad (4)$$

where  $r$  is the upscaling factor,  $\lfloor \cdot \rfloor$  is the rounding function and  $\bmod$  is the modulo operation. Without filling in the zeros, the sub-pixel convolutions exclusive in low-resolution feature map have more representation power than transposed convolutions. As shown in Fig. 2, we set each sub-pixel convolution layer (sub-p) with upscaling factor  $r=2$  and successively employ multiple sub-p operation to upscale the feature maps until the spatial size of final output matches with the original image. Unlike our setting of multi-step resolution upscaling, ReconNet's first layer directly projects and outputs feature maps with the same dimension as the original image, which results in additional computational burden for the subsequent layers.

For the discriminator, its network structure is shown in the Fig. 2 (the bottom part). We adopt the network architecture as in [35]. In detail, the network contains eight convolution layers. After each two layers of convolution, the number of feature maps is doubled and the resolution of feature maps is halved. The convolutional filters size is always set as  $3 \times 3$ , and the activation function is selected as LeakyReLU ( $\alpha=0.2$ ). Following the eight convolution layers, two full-connection layers are employed for the final classification.

## 2.2. Loss function

After deploying the network architecture of the generator and discriminator, we specify the compound loss function for regulating the learning of network parameters, which is defined as the following min-max problem,

$$J(D, G) = \min_G \max_D [\ell_{rec}(G(y), x) + \ell_{meas}(y, \Phi(G(y))) + \ell_{GAN}(G(y), D)], \quad (5)$$

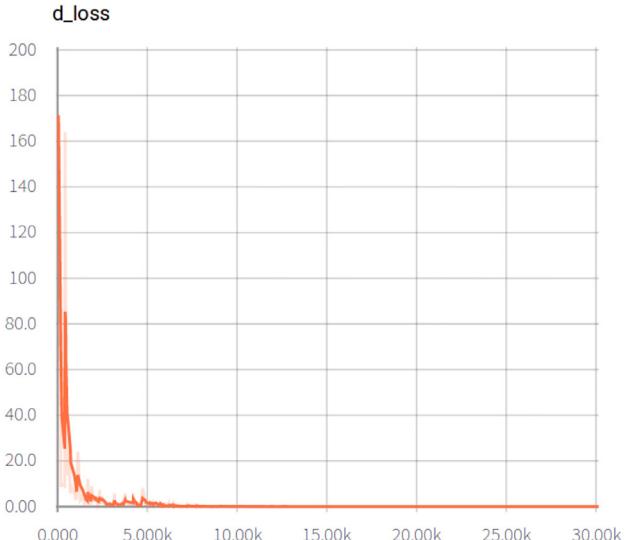
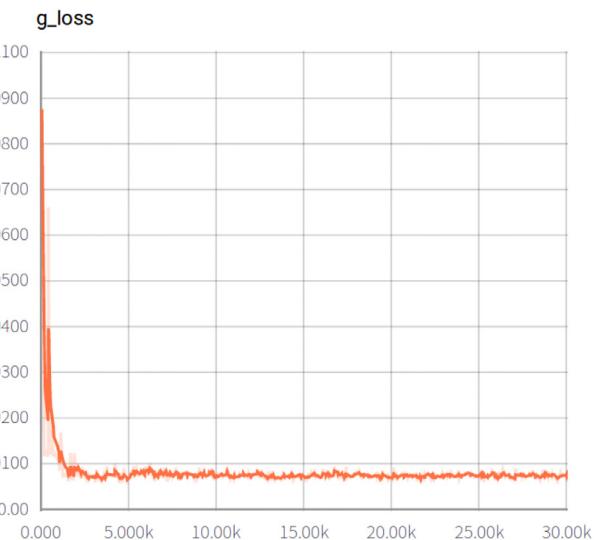
where  $\ell_{rec}$  is the reconstruction error term measuring whether the generated image  $G(y)$  matches with the original image  $x$ ,  $\ell_{meas}$  is the error between the original CS measurement  $y$  and the generated measurement  $\Phi(G(y))$ ,  $\ell_{GAN}$  is the adversarial loss term advocating the distribution of generated images to be consistent with the distribution of real images from the training sample. The core objective of Eq. (5) is to guide the generator to reconstruct the original image from the measurement vector. To achieve this objective, three terms  $\ell_{rec}$ ,  $\ell_{reg}$ ,  $\ell_{GAN}$  of Eq. (5) are defined as follows.

The reconstruction error term  $\ell_{rec}$  is to measure the reconstruction error, which is defined as

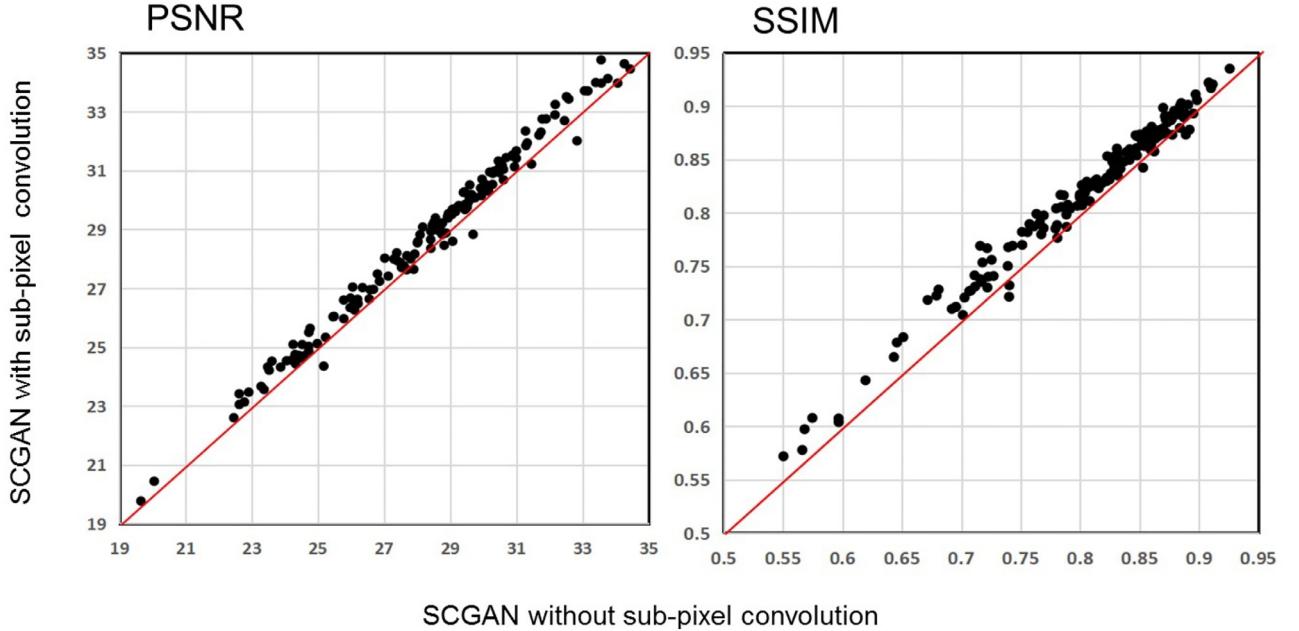
$$\ell_{rec} = \|\hat{x} - x\|_2^2, \quad (6)$$

where  $\hat{x} = G(y)$  is the reconstruction image generated by  $G$ . Apart from the commonly used pixel-wise mean squared error between the original image  $x$  and the generated image  $\hat{x}$ ,  $\ell_{meas}$  term requires the generated image  $G(y)$  can have the similar measurement with the original measurement vector  $y$ ,

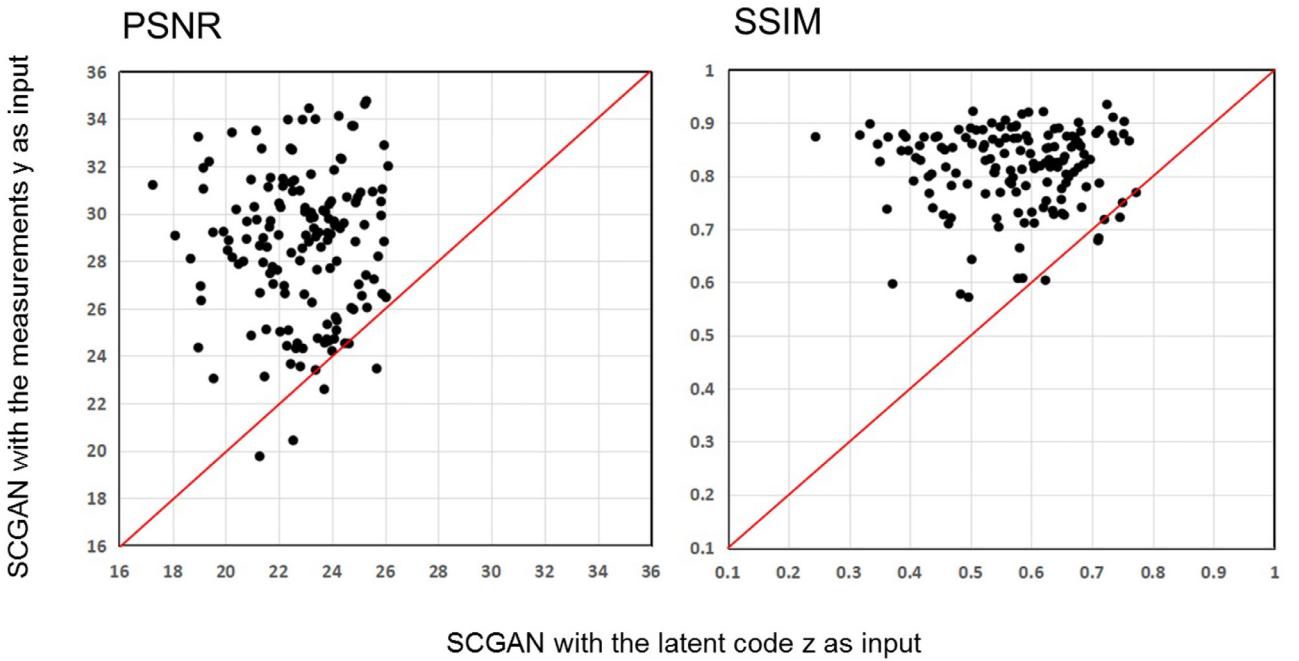
$$\ell_{meas} = \|\Phi\hat{x} - y\|_2^2. \quad (7)$$



**Fig. 4.** the loss function curves of generator and discriminator verse iteration numbers upon the CelebA dataset.



**Fig. 5.** A comparison of the proposed SCGAN model with and without sub-pixel convolution. Multiple sub-pixel convolutions give an improvement on 150 test images of CelebA-Cropped dataset.



**Fig. 6.** A comparison of the proposed SCGAN model with the measurement  $y$  and the latent code  $z$  as input. Learning the explicit mapping from the low-dimensional compressed measurement  $y$  gives an important improvement on 150 test images of CelebA-Cropped dataset.

As shown in Fig. 3, we first get the reconstructed image as  $\hat{x} = G(y)$  in terms of the generator  $G$ , and continue to perform another CS measurement  $\phi$  on the reconstructed image  $\hat{x}$  to get  $\hat{y}$ . The difference between  $y$  and  $\hat{y}$ ,  $x$  and  $\hat{x}$  should both be minimized. Those two loss terms form the image fidelity constraint in a feedback loop. In this way, the generator  $G$  is optimized specifically for the CS task.

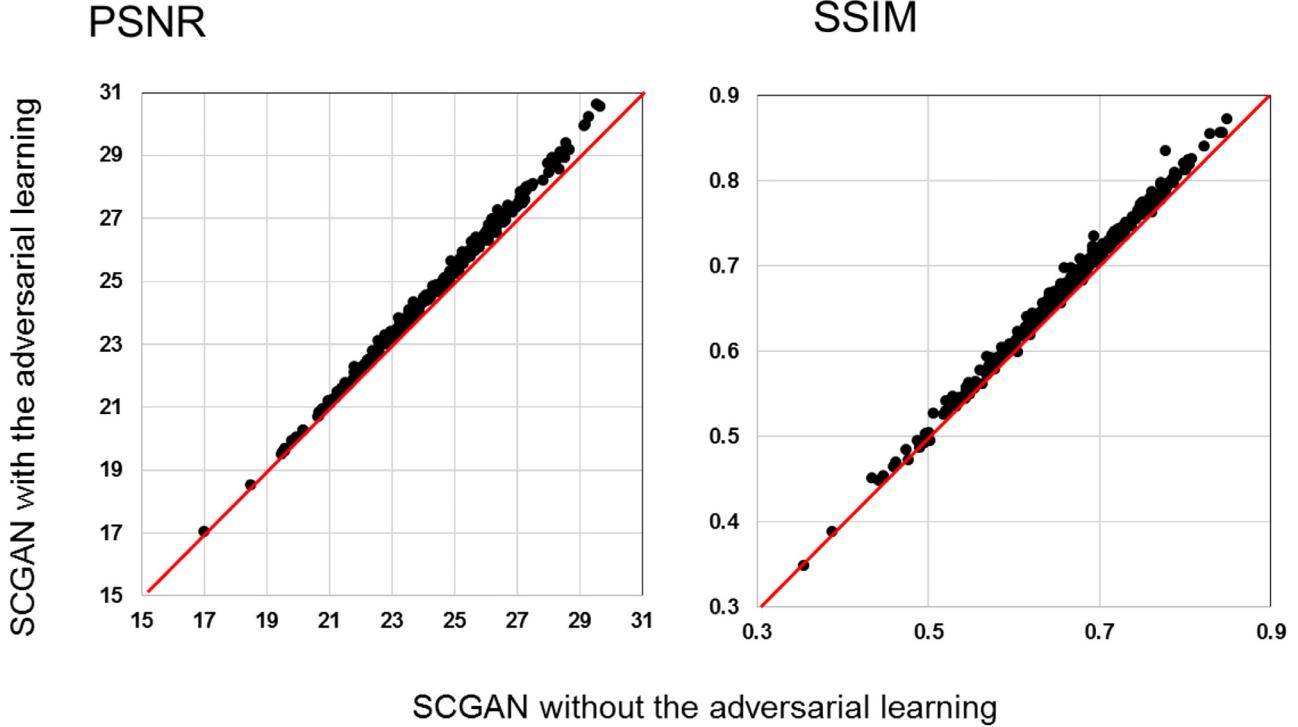
Finally, we define the adversarial loss  $l_{GAN}$  as

$$\ell_{GAN} = E_{x \sim P_{data}} D(x) + \gamma E_{y \sim P_y} (1 - D(G(y))), \quad (8)$$

where  $P_{data}$  and  $P_y$  are the statistical distributions of the original image and the random measurement vector respectively,  $\gamma$  is the

regularization parameter coordinating the importance of the two terms. This loss term mainly advances the output of generator has a similar distribution to the original image, and such that the generated image can deceive the discriminator  $D$ .

$G$  and  $D$  are jointly learned in an adversarial fashion. However, the generative adversarial network is prone to instability in the training process, e.g. mode collapse. Wasserstein GAN is an effective algorithm to stabilize the training. According to the setting of Wasserstein GAN [36], a log operation is removed in Eq. (8). The output of  $D$  is no longer a probability, hence we do not apply the sigmoid activation at the output of the discriminator network. Fi-



**Fig. 7.** A comparison of the proposed SCGAN model with and without adversarial learning. The adversarial learning improves the reconstruction quality.

nally, RMSProp is used to avoid the discontinuity of the discriminator during training [36].

### 2.3. Model training

**Eq. (5)** can be solved by alternately optimizing the generator  $G$  and the discriminator  $D$ . After the SCGAN model is trained well, the generator can be used to reconstruct the test image from the measurement vector. The overall training algorithm is shown in **Algorithm 1**.

With the deployment of Wasserstein GAN, our network learning is more stable in the training phase. Taking the CelebA dataset as example, the function value curves for  $D$  and  $G$  are shown in **Fig. 4**. We can see that the loss function values with respect to generator and discriminator steadily decrease and converge.

## 3. Experimental results and analysis

We test the performance of the proposed SCGAN model on multiple public datasets, i.e., the CelebA face dataset (CelebA), the Fashion-MNIST clothing articles dataset (F-MNIST) and the MNIST handwritten digits dataset (MNIST). In order to analyze the proposed SCGAN model in detail, we first conduct some ablation experiments to verify the validity of multiple components of SCGAN and then compare SCGAN with some state-of-the-art deep learning based algorithms, ReconNet [23], DCGAN [30], CSGAN [32] and conventional iterative reconstruction algorithms, LASSO and TwIST [11], in terms of both reconstruction quality and reconstruction time. These datasets are detailed below.

**The CelebA face dataset (CelebA):** This is a large scale face attributes dataset containing 202,599 color face images and 10,177 number of identities. Each face image has the resolution of  $178 \times 218$ . As in [30] and [32], a cropped version with size  $64 \times 64$  is constructed by cropping the face area of the original high resolution  $178 \times 218$  image. The cropped and the full resolution version are denoted as CelebA-Cropped and CelebA-Full respectively.

---

**Algorithm 1** Sub-pixel convolutional GAN for CS reconstruction.

**Input:**  $\alpha$  the learning rate parameter,  $s$  the batch size,  $L$  the number of iterations of  $D$  per the iteration of  $G$ .

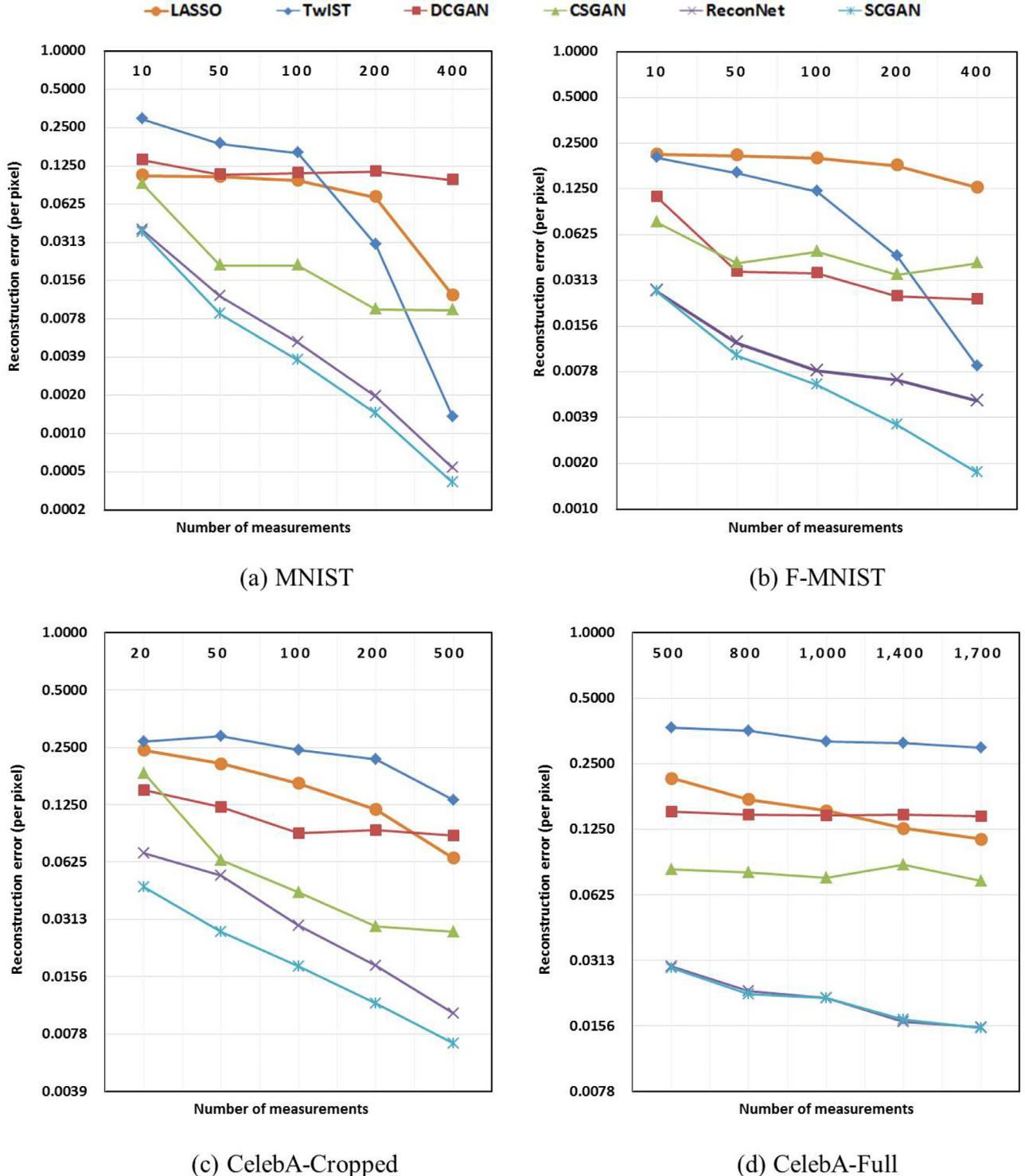
- 1: Initialize  $D$  and  $G$
- 2: **for** specified number of training iterations **do**
- 3:     **TRAINING DISCRIMINATOR**
- 4:     **for** specified  $L$  steps **do**
- 5:         Draw a batch of  $s$  training images  $\{x^{(1)}, \dots, x^{(s)}\}$
- 6:         Sample a batch of  $s$  training images, and compute the measurement vectors  $y^{(i)} = \Phi x^{(i)}$
- 7:         Update  $D$ 's parameters  $\omega$  by RMSProp method:
- 8:         
$$g_\omega \leftarrow \nabla_\omega \frac{1}{s} \sum_{i=1}^s D_\omega(x^{(i)}) + \gamma (1 - D_\omega(G_\theta(y^{(i)})))$$
- 9:         
$$\omega \leftarrow \omega + \alpha \cdot \text{RMSProp}(\omega, g_\omega)$$
- 10:     **end for**
- 11:     **TRAINING GENERATOR**
- 12:         Draw a batch of  $s$  training images  $\{x^{(1)}, \dots, x^{(s)}\}$  and denote the corresponding reconstructed images as  $\{\hat{x}^{(1)}, \dots, \hat{x}^{(s)}\}$
- 13:         For all sampled examples, compute the measurement vectors  $y^{(i)} = \Phi \hat{x}^{(i)}$
- 14:         Update  $G$ 's parameters  $\theta$  by RMSProp method:
- 15:         
$$g_\theta \leftarrow -\nabla_\theta \frac{1}{s} \sum_{i=1}^s \{\|\hat{x}^{(i)} - x^{(i)}\|_2^2 + \|\Phi \hat{x}^{(i)} - y^{(i)}\|_2^2 + \gamma (1 - D_\omega(G_\theta(y^{(i)})))\}$$
- 16:         
$$\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$$
- 17: **end for**

**Output:** Well-trained  $G$  and  $D$

---

The reconstruction experiments are conducted on both the CelebA-Cropped and CelebA-Full dataset. We select the 1 to 162,770 images as the training set, the 162,771 to 182,637 pictures as the validation set, and the 182,638 to 202,599 pictures as the test set.

**The Fashion-MNIST clothing articles dataset (F-MNIST):** This dataset is provided by the research department of Zalando. F-Mnist dataset consists of 10 fashion categories, including 60,000 gray-



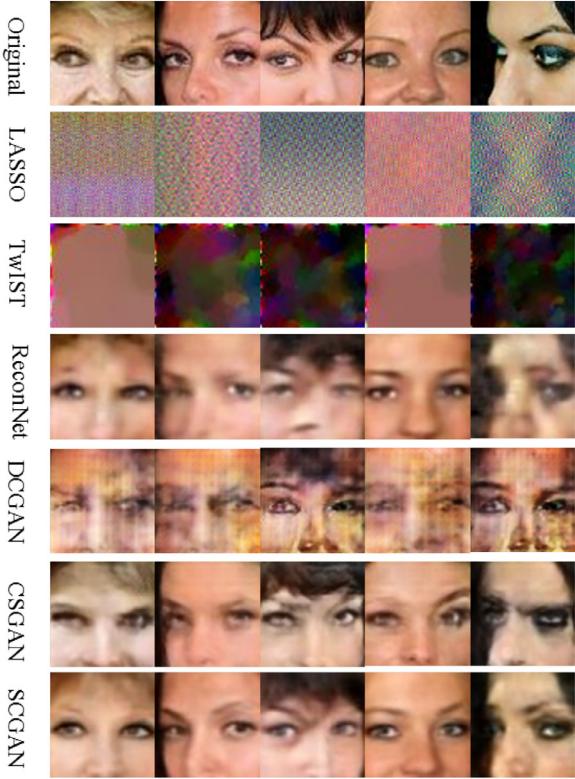
**Fig. 8.** The MSE plots of various algorithms upon MNIST, F-MNIST, CelebA-Cropped and CelebA-Full datasets with multiple measurements numbers. The vertical axis are the per-pixel mean-squared reconstruction error (in log2 scale) and the horizontal axis is the number of measurements.

scale images of size  $28 \times 28$  and along with a test set of 10,000 images.

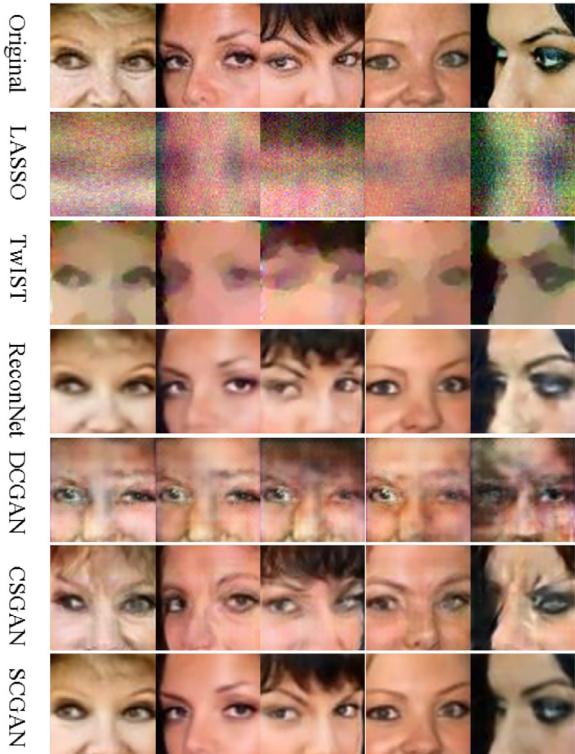
**The MNIST handwritten digits dataset (MNIST):** This dataset contains a total of 70,000 handwritten digits images. 60,000 examples are selected as training set and the remaining as test set. Each example is normalized as gray-scale  $28 \times 28$  pixels and placed in the center of the image.

### 3.1. Experimental setting

With regard to the proposed SCGAN algorithm, the training is divided into two stages. In the first stage, only the generator are pre-trained with the setting of iteration numbers as 10, the learning rate as 0.01, such that the training results can quickly converge; In the second stage, we train the generator and the discriminator



(a) Reconstructions with 100 measurements



(b) Reconstructions with 400 measurements

**Fig. 9.** Reconstructed images of various algorithms upon CelebA-Cropped dataset using different numbers of measurements. Top to bottom rows are the original images, and reconstructions by LASSO, TWIST, ReconNet, DCGAN, CSGAN and our SCGAN respectively.

according to algorithm 1 and fine-tune the parameters of the generator obtained from the first stage with the setting of iterations number as 30, the learning rate as 0.002. The learning rate is reduced to 0.001 when the iteration number exceeds 15. The training of all networks is carried out on the NVIDIA GeForce GTX TITAN X GPU using TensorFlow as backend.

The generator and discriminator of CSGAN and DCGAN both follow the deep convolutional network architecture of [31]. The characteristics of CSGAN is to joint optimize  $z$  and network parameters for the CS task. The source code of CSGAN, DCGAN, ReconNet and TwIST are downloaded from the links given in the original papers. Adam optimizer is adopted for training the CSGAN and DCGAN in the tensorflow framework. ReconNet bases upon the Caffe framework. LASSO performed directly  $l_1$  sparse constraint on the pixel values for MNIST and F-MNIST, and in discrete cosine transform domains for CelebA-Cropped and CelebA-Full datasets. ReconNet, DCGAN, CSGAN and our SCGAN directly reconstruct the entire image for the CelebA-Full dataset. The hyper-parameters of these comparing algorithms are set to match the ones in the original papers. For all the experiments, measurement matrix  $\Phi$  is set as a zero-mean random Gaussian matrix with independent and identically distributed entries and we assume the measurement is noiseless.

### 3.2. Our model analysis

The main strengths of the SCGAN model for compressed reconstruction lie in three aspects, namely the employment of sub-pixel convolution for upscaling image resolution, learning the direct mapping from compressed measurement to the reconstruction and the adversarial learning manner. In order to evaluate the benefits of these three strengths, we implement another three versions of SCGAN model and conduct three groups of experiments.

The first version named SCGAN-v1 is to transform the SCGAN model by replacing the sub-pixel convolution as the conventional transposed convolutions, and we compare their reconstruction quality based on peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index [37]. This experiment is conducted on 150 face images from the CelebA-Cropped dataset with 500 measurements. The reconstruction results of SCGAN and SCGAN-v1 for each sample are arranged as tuples and depicted in Fig. 5. It can be seen that most of the points are above the straight line, which demonstrates the advantage of using sub-pixel convolution for the CS reconstruction.

The second version named SCGAN-v2 is to transform the SCGAN model by taking the random latent code  $z$  as input and generate the reconstruction image by solving Eq. 3 following the strategy of DCGAN [30] and CSGAN [32]. The experimental results are shown in Fig. 6. The SCGAN model outperforms SCGAN-v2 with significant improvement in terms of PSNR and SSIM measure upon the CelebA-Cropped dataset with 500 measurements, which verifies the effectiveness of learning the explicit mapping from the low-dimensional compressed measurement to the high-dimensional reconstruction.

The third version named SCGAN-v3 is to train our sub-pixel convolutional generator without the adversarial learning manner. Comparative experiments between these two algorithms are conducted upon 300 face images from the CelebA-Full dataset with 1000 measurements. The PSNR and SSIM values of each reconstructed face image using the two method are shown in Fig. 7. It can be shown that the adversarial learning does help to improve the reconstruction quality even with very few measurements. This improving is mainly attributed to the fact that the generator can be promoted to learn the statistical distribution of real images.

As regard to the selection of convolution kernel size for our SCGAN model, we test four sizes including  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , and

**Table 1**

The reconstruction performance and time complexity of SCGAN with different convolution kernel size upon the CelebA-Full dataset.

Kernel size	$3 \times 3$	$5 \times 5$	$7 \times 7$	$9 \times 9$
MSE	0.0153	0.01578	0.0159	0.0163
PSNR	24.7692	24.6465	24.6072	24.5079
SSIM	0.6742	0.6720	0.6707	0.6669
Time complexity (ms)	3.4340	4.2313	5.6927	7.5997

**Table 2**

PSNR and SSIM values of different algorithms upon the CelebA-Cropped dataset at multiple measurement numbers.

Number of measurements	LASSO		TwIST		DCGAN		CSGAN		ReconNet		SCGAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
20	12.5482	0.0351	11.9598	0.0249	14.7166	0.1607	13.9502	0.1699	18.3923	0.3184	<b>20.2133</b>	<b>0.4418</b>
50	13.3525	0.1187	11.6943	0.0560	15.6927	0.1974	18.4961	0.3575	19.2434	0.3725	<b>22.5931</b>	<b>0.5614</b>
100	14.4518	0.2195	12.4257	0.1247	16.9252	0.2201	20.2503	0.4557	22.0527	0.5249	<b>24.4723</b>	<b>0.6497</b>
200	15.8287	0.3308	13.0145	0.2132	16.6623	0.2483	22.0001	0.5439	24.2389	0.6352	<b>26.4187</b>	<b>0.7289</b>
500	18.3790	0.4991	15.7259	0.6862	17.0053	0.2535	22.2607	0.5441	26.7663	0.7487	<b>28.3954</b>	<b>0.8024</b>

Note: The bold values indicate the best PSNR or SSIM of these algorithms.

**Table 3**

PSNR and SSIM values of different algorithms upon the CelebA-Full dataset at multiple measurement numbers.

Number of measurements	LASSO		TwIST		DCGAN		CSGAN		ReconNet		SCGAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
500	13.0526	0.1437	10.6549	0.0479	14.6169	0.3360	17.2020	0.3341	21.8701	<b>0.5606</b>	<b>21.9757</b>	0.5593
800	14.0905	0.1900	10.8725	0.1265	14.7535	0.3442	17.3586	0.3302	23.0199	0.6054	<b>23.1654</b>	<b>0.6082</b>
1000	14.6090	0.2176	11.0776	0.1907	14.7389	0.3419	17.6170	0.3634	23.2940	0.6136	<b>23.3771</b>	<b>0.6499</b>
1400	15.4107	0.2560	11.5471	0.2706	14.7252	0.3416	16.8997	0.3137	<b>24.4683</b>	0.6579	24.3853	<b>0.6600</b>
1700	15.8987	0.2835	11.9279	0.3233	14.8254	0.3449	17.8043	0.3569	24.7121	0.6639	<b>24.7692</b>	<b>0.6742</b>

Note: The bold values indicate the best PSNR or SSIM of these algorithms.

**Table 4**

PSNR and SSIM values of different algorithms upon the F-MINIST dataset at multiple measurement numbers.

Number of measurements	LASSO		TwIST		DCGAN		CSGAN		ReconNet		SCGAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	7.5850	0.2904	7.1914	0.0058	10.3294	0.3571	11.9798	0.3695	<b>16.6220</b>	<b>0.6032</b>	16.5811	0.6004
50	7.6343	0.2665	8.3121	0.1592	15.4760	0.6117	14.7582	0.5537	20.2356	0.7689	<b>21.3385</b>	<b>0.8088</b>
100	7.8320	0.2527	9.7027	0.3487	15.7943	0.6171	14.3564	0.4892	22.2870	0.8395	<b>23.4056</b>	<b>0.8589</b>
200	8.4231	0.2533	14.3431	0.5437	17.1320	0.6920	15.8138	0.5883	22.6020	0.8324	<b>26.2512</b>	<b>0.9193</b>
400	10.2824	0.3484	22.3436	0.7173	17.1243	0.6902	15.1816	0.5177	23.8831	0.8650	<b>29.2229</b>	<b>0.9511</b>

Note: The bold values indicate the best PSNR or SSIM of these algorithms.

$9 \times 9$ . The Table 1 presents the influence of convolution kernel size on the reconstruction performance and computational efficiency of SCGAN upon the CelebA-Full dataset with 1700 measurements. It can be seen that as convolution kernel size increases, the reconstruction quality decreases while the time complexity of reconstructing a test image increases. Thus, we set the convolution kernel size as  $3 \times 3$  in our SCGAN model.

### 3.3. Experimental comparison

We compare performance of different CS reconstruction algorithms qualitatively and quantitatively. For quantitative comparison, the mean squared error (MSE), PSNR and SSIM measures are used to evaluate the reconstruction quality. Following [32], the number of measurements range from 10 to 500 for the MNIST, F-MNIST and CelebA-Cropped datasets. As regard to the CelebA-Full dataset, we set the number of measurements from 500 to 1700. In all experiments, the average results are reported on the test set unseen during the training phase. Fig. 8 shows the per-pixel mean-squared reconstruction error of different algorithms. In addition, Tables 2–5 also report the average PSNR and SSIM values of these algorithms upon the four datasets using different numbers of

measurements. The best PSNR or SSIM values of these algorithms are marked in bold.

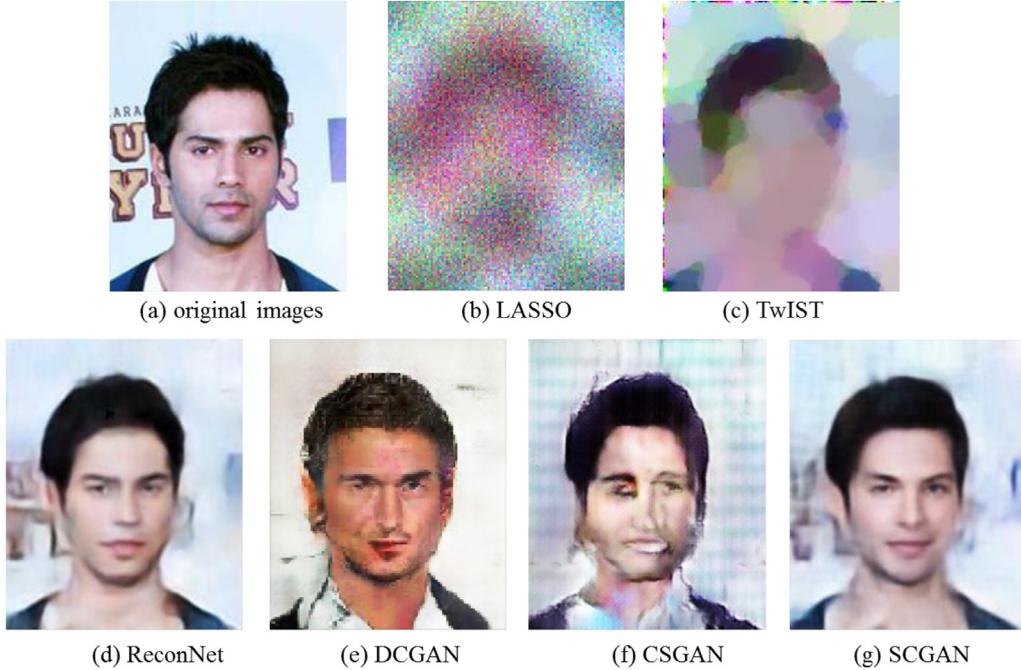
We first analyze the performance of LASSO and TwIST methods. Upon the MINIST and F-MINISST dataset, TwIST has lower MSE than LASSO and the MSE of TwIST rapidly reduces with the measurement number increasing. However, LASSO performs better than TwIST on the CelebA-Cropped and CelebA-Full datasets. As regard to the deep learning based algorithms, they are obviously better than TwIST and LASSO upon the complex face dataset. CSGAN achieves better or comparable reconstruction results than DCGAN upon all the datasets, which demonstrates the compressed sensing task-aware learning is beneficial for reconstructing unseen test samples more reliably. After the number of measurements exceeds 100, both the performance of CSGAN and DCGAN tend to saturate, and additional measurements give no significant performance improvement. This limitation may be attributed to the mechanism that DCGAN and CSGAN models learn the reconstruction mapping from the latent vector  $z$  rather than measurement vector  $y$  to the high dimensional image space, thus the growth of measurement number can not necessarily bring the improvement of reconstruction quality. Our model and ReconNet both learn explicit reconstruction mapping from the random measurement to the recon-

**Table 5**

PSNR and SSIM values of the MNIST dataset using different algorithms at multiple measurement numbers.

Number of measurements	LASSO		TwIST		DCGAN		CSGAN		ReconNet		SCGAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	10.1026	0.5414	5.5225	0.0012	8.7231	0.2055	11.1678	0.4179	14.7993	0.5233	<b>14.8612</b>	<b>0.5421</b>
50	10.1896	0.5080	7.5428	0.0916	9.9335	0.2871	18.4221	0.8206	20.0300	0.8191	<b>21.7215</b>	<b>0.8879</b>
100	10.5685	0.4783	8.7379	0.2438	9.7617	0.3188	18.1993	0.8165	23.7355	0.9263	<b>25.3156</b>	<b>0.9531</b>
200	12.3195	0.5002	17.4029	0.4875	9.6081	0.2401	21.2316	0.9113	28.2581	0.9793	<b>29.5088</b>	<b>0.9824</b>
400	19.5859	0.7561	30.2395	0.6473	10.3606	0.3264	21.3980	0.9126	33.7970	0.9922	<b>34.7224</b>	0.9913

The bold values indicate the best PSNR or SSIM of these algorithms.



**Fig. 10.** Some reconstructed images of various algorithms using 800 measurements, when applied to a sample image from CelebA-Full dataset.

structure. The performance of our model and ReconNet stay ahead of CSGAN and DCGAN upon all these datasets.

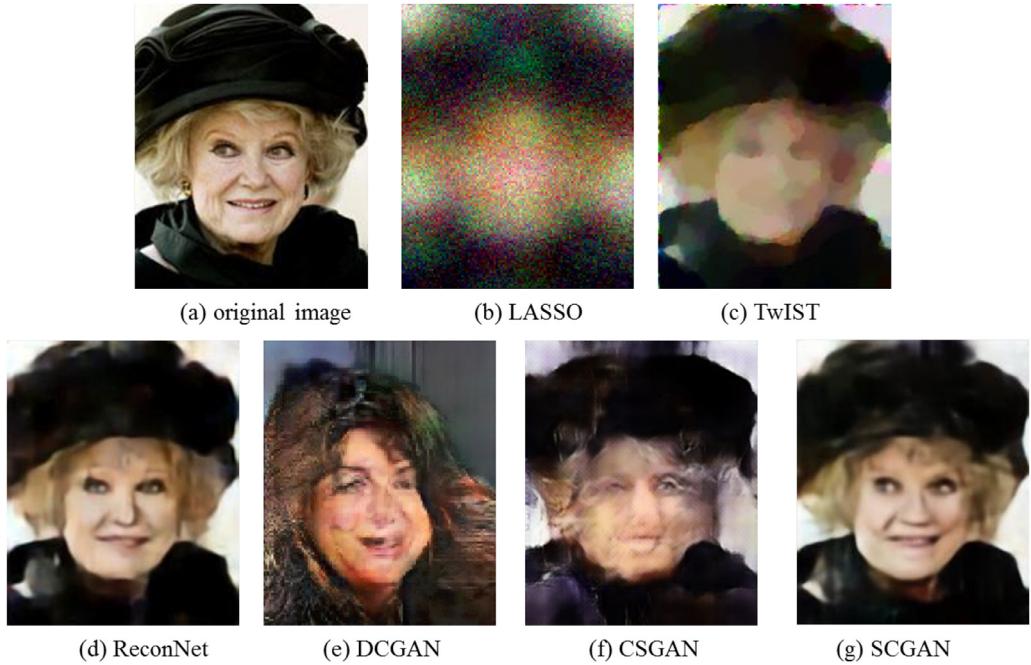
Overall, our algorithm has the optimal or comparable performance in all these datasets. Although ReconNet outperforms our model in a few case of measurement numbers, ReconNet involves the dense layer for projecting the measurement vector as the feature map with its dimension same as the original image and thus lead to the network to have much more parameters. The superiority of our model over ReconNet mainly owes to the exploiting of min-max adversarial training manner to capture inherent image distribution. Compared with CSGAN and DCGAN, the improvement of our algorithm mainly boils down to the design of sub-pixel convolutional generator network for learning explicit CS reconstruction from the measurement vector  $y$ .

Some reconstructed images of these algorithms upon the CelebA-Cropped, CelebA-Full and F-MNIST datasets are also shown in Figs. 9–12. In the whole, LASSO and TwIST have poor reconstruction quality and are even impossible to reconstruct the basic image structures in the case of low measurement numbers. TwIST can well recover the F-MINISST image with 400 measurements, but the reconstructed images of TwIST upon the face image is poor. The structures and textures in the reconstruction images of CSGAN seem visually reasonable, however the reconstruction images couldn't match well with the original image and even worse suffer from the issue of error reconstruction. As shown in Figs. 9–11, the reconstructed facial features of CSGAN partially deviate from the true patterns in the original images. In Fig. 12, CSGAN recovers the

cloth pattern whilst the ground-truth is the shoe. This error may result from the case that the unknown image is far from the range of the generator or the optimization procedure does not find the best latent vectors  $z$ . Different from CSGAN and DCGAN, our SC-GAN learns the direct reconstruction mapping and thus can avoid this issue, and the reconstructed images of SCGAN resemble the original images more accurately than CSGAN and DCGAN. ReconNet has good visual quality on these data sets. Compared with ReconNet algorithm, our model can recover more meaningful structures especially in the low measurement numbers.

### 3.4. Running time

In addition to assess reconstruction quality, we also evaluate the running time of our algorithm and the competing algorithms. We use an Intel Core i70500U CPU to run the LASSO and TwIST code. All the deep learning based algorithms are run in NVIDIA GeForce GTX TITAN X GPU. Table 6 reports the time complexity (in millisecond) of multiple reconstruction algorithms upon the three datasets. The minimum time complexity values of these algorithms are marked in bold. When the image size scales up, especially in the case of  $64 \times 64$  and  $178 \times 218$  color face image, the running times of TwIST and LASSO increase dramatically. Due to gradient descent optimization to find the optimal matching latent vector  $z$ , DCGAN and CSGAN are still computationally expensive. The computation of our algorithm and ReconNet is about three orders of magnitude faster than TwIST and LASSO reconstruction al-



**Fig. 11.** Some reconstructed images of various algorithms using 1700 measurements, when applied to a sample image from CelebA-Full dataset.



**Fig. 12.** Reconstruction results on F-MNIST with 400 measurements. Top to bottom rows are the original images (a), and reconstructions by LASSO (b), TwIST (c), ReconNet (d), DCGAN (e), CSGAN (f) and our SCGAN (g).

**Table 6**  
Average time complexity (in milliseconds) of various reconstruction algorithms upon the three datasets.

Dataset	LASSO	TwIST	ReconNet	DCGAN	CSGAN	SCGAN
MNIST	1038.21	1258.42	1.7887	616.18	623.43	<b>1.0837</b>
F-MINIST	1348.43	1599.91	1.6610	616.54	613.36	<b>1.1472</b>
CelebA-Cropped	81067.35	91129.50	2.1516	2560.40	2503.52	<b>1.7015</b>
CelebA-Full	131381.37	177652.62	10.4456	9069.45	8963.33	<b>3.4340</b>

The bold values used in this table indicate the minimum time complexity of these algorithms.

gorithms. By the utilization of the multi-steps resolution upscaling, our model has fewer network parameters than ReconNet, further reducing computational complexity.

#### 4. Conclusion

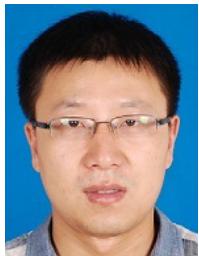
In this paper, we proposed a sub-pixel convolutional generative adversarial network for learning compressed sensing reconstruction of images. A compound loss function is designed to guide the compressed sensing oriented min-max adversarial learning. In the generator network, we use multiple building blocks consisting of normal convolutions, batch normalization, SELU activation and sub-pixel convolutions, to upscale the image resolution layer by layer until coinciding with the dimension of the original image. By feeding the low-dimensional measurement to the well-trained generator, the test image can be fast reconstructed by only single-pass forward computation in non-iterative manner. The experimental results on multiple data sets demonstrate the superiority of our proposed network model over some conventional and deep learning based reconstruction algorithms in terms of both reconstruction quality and time complexity. At present, our generator network mainly exploits common convolution operations for learning the reconstruction mapping. In the succeeding work, we expect to further improve the network structure such as introducing residual connections and attention mechanisms, and thus more detailed structures can be preserved in the reconstructed image. Furthermore, the proposed generative adversarial networks can be extended into the compressed sensing of image sequences by taking account of the inter-frame correlation.

#### Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant numbers ([61672292](#), [61825601](#), [61532009](#) and [61622305](#)), in part by the Key University Science Research Project of Jiangsu Province under Grant Number [18KJA520007](#).

#### References

- [1] E.J. Candes, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (12) (2005) 4203–4215, doi:[10.1109/tit.2005.858979](#).
- [2] F.D. Marco, A.D. Mark, T. Dharmpal, et al., Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.* 25 (2) (2008) 83–91, doi:[10.1109/MSP.2007.914730](#).
- [3] F. Shahrooz, Z. Dornoosh, A. Ashraf, J. Ljaremko, Multi-dimensional low rank plus sparse decomposition for reconstruction of under-sampled dynamic mri, *Pattern Recognit.* 63 (2017) 667–679, doi:[10.1016/j.patcog.2016.09.040](#).
- [4] Z. Dornoosh, F. Shahrooz, A. Ashraf, J. Ljaremko, Dependent nonparametric bayesian group dictionary learning for online reconstruction of dynamic mr images, *Pattern Recognit.* 63 (2017) 518–530, doi:[10.1016/j.patcog.2016.09.038](#).
- [5] D.-S. Pham, On group-wise  $l_p$  regularization: theory and efficient algorithms, *Pattern Recognit.* 48 (11) (2015) 3728–3738, doi:[10.1016/j.patcog.2015.05.009](#).
- [6] W. Yulong, T. YuanYan, L. Luoqing, Z. Xianwei, Block sparse representation for pattern classification: theory, extensions and applications, *Pattern Recognit.* 88 (2019) 198–209, doi:[10.1016/j.patcog.2018.11.026](#).
- [7] E.J. Candes, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathematique* 346 (9–10) (2008) 589–592, doi:[10.1016/j.crma.2008.03.014](#).
- [8] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Probl.* 24 (3) (2008) 035020, doi:[10.1088/0266-5611/24/3/035020](#).
- [9] Q. Saad, M.B. Rana, I. Wafa, et al., Compressive sensing: from theory to applications, a survey, *J. Commun. Netw.* 15 (5) (2013) 443–456, doi:[10.1109/JCN.2013.000083](#).
- [10] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* 4 (4) (2005) 1168–1200, doi:[10.1137/050626090](#).
- [11] J.M. Bioucas-Dias, M.A. Figueiredo, A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Trans. Image Process.* 16 (12) (2007) 2992–3004, doi:[10.1109/tip.2007.909319](#).
- [12] D.L. Donoho, A. Maleki, A. Montanari, Message-passing algorithms for compressed sensing, *Proc. Natl. Acad. Sci.* 106 (45) (2009) 18914–18919, doi:[10.1073/pnas.0909892106](#).
- [13] M.A. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE J. Select. Top. Signal Process.* 1 (4) (2007) 586–597, doi:[10.1109/jstsp.2007.910281](#).
- [14] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for  $l_1$ -minimization with applications to compressed sensing, *SIAM J. Imaging Sci.* 1 (1) (2008) 143–168, doi:[10.1137/070703983](#).
- [15] G. Kin, A. Adedotun, S. Soumik, Llnet: a deep autoencoder approach to natural low-light image enhancement, *Pattern Recognit.* 61 (2017) 650–662, doi:[10.1016/j.patcog.2016.06.008](#).
- [16] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (2) (2016) 295–307, doi:[10.1109/tpami.2015.2439281](#).
- [17] L. Wang, Z. Huang, Y. Gong, C. Pan, Ensemble based deep networks for image super-resolution, *Pattern Recognit.* 68 (2017) 191–198, doi:[10.1016/j.patcog.2017.02.027](#).
- [18] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, T. Tan, Feedback convolutional neural network for visual localization and segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, Online, (2018), doi:[10.1109/TPAMI.2018.2843329](#).
- [19] Y. Wei, X. Liang, Y. Chen, X. Shen, et al, Stc: a simple to complex framework for weakly-supervised semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (11) (2017) 2314–2320, doi:[10.1109/TPAMI.2016.2636150](#).
- [20] A. Mousavi, A.B. Patel, R.G. Baraniuk, A deep learning approach to structured signal recovery, in: 53rd Annual Allerton Conference on Communication, Control, and Computing, 2015, pp. 1336–1343, doi:[10.1109/allerton.2015.7447163](#).
- [21] A. Adler, D. Boubil, M. Zibulevsky, Block-based compressed sensing of images via deep learning, International Workshop on Multimedia Signal Processing, IEEE, 2017, doi:[10.1109/mmsp.2017.8122281](#).
- [22] M. Janki, M. Angshul, Rodeo: robust de-aliasing autoencoder for real-time medical image reconstruction, *Pattern Recognit.* 63 (2017) 499–510, doi:[10.1016/j.patcog.2016.09.022](#).
- [23] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, A. Ashok, Reconnet: non-iterative reconstruction of images from compressively sensed measurements, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 449–458, doi:[10.1109/cvpr.2016.55](#).
- [24] C. Li, W. Yin, H. Jiang, Y. Zhang, An efficient augmented lagrangian method with applications to total variation minimization, *Comput. Optim. Appl.* 56 (3) (2013) 507–530, doi:[10.1007/s10589-013-9576-1](#).
- [25] W. Dong, G. Shi, X. Li, F. Ma, Y. Huang, Compressive sensing via nonlocal low-rank regularization, *IEEE Trans. Image Process.* 23 (8) (2014) 3618–3632, doi:[10.1109/TIP.2014.2329449](#).
- [26] A. Mousavi, R.G. Baraniuk, Learning to invert: signal recovery via deep convolutional networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 2272–2276, doi:[10.1109/icassp.2017.7952561](#).
- [27] A. Mousavi, G. Dasarathy, R.G. Baraniuk, DeepCodec: adaptive sensing and recovery via deep convolutional neural networks, in: 55th Annual Allerton Conference on Communication, Control, and Computing, 2017, doi:[10.1109/allerton.2017.8262812](#).
- [28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, et al., Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
- [29] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.
- [30] A. Bora, A. Jalal, E. Price, A.G. Dimakis, Compressed sensing using generative models, in: International Conference on Machine Learning, 70, 2017, pp. 537–546.
- [31] R. Alec, M. Luke, C. Soumith, Unsupervised representation learning with deep convolutional generative adversarial networks, in: International Conference on Learning Representations, 2016.
- [32] M. Kabkab, P. Samangouei, R. Chellappa, Task-aware compressed sensing with generative adversarial networks, in: AAAI Conference on Artificial Intelligence, 2018, pp. 2297–2304.
- [33] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-normalizing neural networks, in: Advances in Neural Information Processing Systems, 2017, pp. 971–980.
- [34] W. Shi, J. Caballero, F. Huszár, e.a. Totz, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883, doi:[10.1109/cvpr.2016.207](#).
- [35] C. Ledig, L. Theis, F. Huszár, J. Caballero, Photo-realistic single image super-resolution using a generative adversarial network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 105–114, doi:[10.1109/cvpr.2017.19](#).
- [36] A. Martin, C. Soumith, B. Lon, Wasserstein generative adversarial networks, in: Proceedings of Machine Learning Research, 70, 2017, pp. 214–223.
- [37] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612, doi:[10.1109/tip.2003.819861](#).



**Yubao Sun** received the Ph.D. degree from Nanjing University of Science and Technology, Nanjing, China, in 2010. He is an associate professor in the School of Automation, Nanjing University of Information Science and Technology, China. His research interests are deep learned based compressed sensing, low rank and sparse representation, video analysis, hypergraph learning, and so on.



**Jiwei Chen** received the bachelor's degree in automation from the Nanjing University of Information Science and Technology, Nanjing, China, in 2017. He is current a master candidate and his research interests include deep learning and pattern recognition.



**Qingshan Liu** received the M.S. degree from Southeast University, Nanjing, China, in 2000, and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2003. He was an Associate Professor with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences. From 2010 to 2011, he was an Assistant Research Professor with the Department of Computer Science, Computational Biomedicine Imaging and Modeling Center, Rutgers University, The State University of New Jersey, Piscataway, NJ, USA. From 2004 to 2005, he was an Associate Researcher with the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing. His research interests include image and vision analysis. Dr. Liu received the National Science Fund for Distinguished Young Scholars in 2018 and the President Scholarship of the Chinese Academy of Sciences in 2003.



**Guangcan Liu** received the bachelor's degree in mathematics and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2004 and 2010, respectively. He was a postdoctoral researcher with the National University of Singapore, Singapore, from 2011 to 2012, the University of Illinois at Urbana-Champaign, Champaign, IL, USA, from 2012 to 2013, Cornell University, Ithaca, NY, USA, from 2013 to 2014, and Rutgers University, Piscataway, NJ, USA, in 2014. Since 2014, he has been a professor with the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China. His research interests touch on the areas of machine learning, computer vision, and image processing. He is a member of IEEE.