Super-Resolution ISAR Imaging for Maneuvering Target Based on Deep-Learning-Assisted Time–Frequency Analysis

Jiang Qian, *Member, IEEE*, Shaoyin Huang, Lu Wang, Guoan Bi, *Senior Member, IEEE*, and Xiaobo Yang

***Abstract*— Traditional range-instantaneous Doppler (RID) methods for maneuvering target imaging suffer from the problems of low resolution and poor noise suppression. We propose a new super-resolution inverse synthetic aperture radar (ISAR) imaging method based on deep-learning-assisted time–frequency analysis (TFA). Our deep neural network resembles the basic structure of a U-net with two additional convolutional-upsampling layers and l_1-norm loss function for super-resolution generation and noise suppression. The neural network is trained in advance to learn the mapping function between the low-resolution time–frequency spectrum inputs and their high-resolution references. Then, the linear TFA assisted by the pretrained network is integrated into the RID-based ISAR imaging system and is found to achieve sharply focused and denoised target image with super-resolution. Both the simulated and real radar data are used to evaluate the performance of the proposed method. Numerical experimental results demonstrate the superiority of the proposed ISAR imaging method over traditional ones.**

***Index Terms*— Deep learning, denoising, inverse synthetic aperture radar (ISAR), short-time Fourier transform (STFT), super-resolution.**

I. Introduction

INVERSE synthetic aperture radar (ISAR) is widely used in civil and military fields for long-distance, all-day, and all-weather observations [1]. ISAR transmits a linear frequency modulation (LFM) signal to obtain a high range resolution proportional to its bandwidth. High azimuth resolution is achieved by coherently integrating the observations of radar echoes after compensating for the target motion relative to the radar. Problem is easily solved when the motion of the target is uniform. After translational motion compensation, a conventional Fourier transform (FT), also known as the range-Doppler (RD) algorithm, can be performed to achieve a high azimuth resolution [2]. Super-resolution of radar imaging with effective denoising can be achieved by sparse representation with much higher computational complexity [3], [4]. Additional target structure can be further taken into consideration in sparse representation for an improved target feature preservation [5], [6]. However, target motion is generally complex due to the change of the target's direction and its time-varying rotation velocity. ISAR imaging by RD then suffers a blurred target image, as the Doppler frequency is time-varying and cannot simply be estimated by the FT [2]. In order to deal with the problem caused by the time-varying motion of maneuvering targets, time–frequency analysis (TFA) methods have been successfully applied to ISAR imaging, known collectively as the range-instantaneous Doppler (RID) methods [7].

TFA-based ISAR imaging for maneuvering target has been first studied in [8]. Different variants of TFA are proposed successively by a number of research groups. Linear TFA, such as short-time FT (STFT) and the Gabor transform, are the most well-known methods [9], [10] for their computational efficiency. However, there is a tradeoff on the resolution between time and frequency for the linear TFA. The maximum resolution of the instantaneous frequency is critically limited by the length of the window function. In order to improve the resolution of TFA, the Wigner–Ville distribution (WVD) and reassignment spectrum (RSP) have been proposed [11], [12]. However, WVD inevitably introduces an undesirable cross-term, and RSP may weaken the target energy, which mistakenly eliminates important target scatterers in radar imaging. Those disadvantages dramatically reduce the quality of the target images. The technique of combing TFA with compressed sensing (CS) to improve the resolution is studied in [13], which is computationally expensive and is difficult to meet the real-time requirements.

Recently, with the rapid development of deep learning, specific designed neural networks are applied to radar imaging for different applications. The deep convolutional

Manuscript received September 20, 2020; revised December 7, 2020; accepted January 2, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61401077 and Grant 61501375, in part by the Sichuan Science and Technology Program under Grant 2019YFG0099, in part by the China Postdoctoral Science Foundation under Grant 2015M580784, and in part by the the Natural Science Basis Research Plan in Shaanxi Province of China under Grant 2018JM6020. *(Corresponding author: Lu Wang.)*

Jiang Qian and Shaoyin Huang are with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China, also with the Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China, and also with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: jqian@uestc.edu.cn).

Lu Wang is with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: wang_lu@i2r.a-star.edu.sg).

Guoan Bi is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: egbi@ntu.edu.sg).

Xiaobo Yang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: cplus2@126.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TGRS.2021.3050189.

Digital Object Identifier 10.1109/TGRS.2021.3050189

neural network (CNN) is used [14] to suppress the sidelobe of ISAR image. Neural networks, especially deep ones, are usually data demanding. To solve the problem of insufficient training data, simulated data are used, where target images obtained by the traditional RD are fed into the model as inputs, and ISAR images by the ideal point model are applied as references. Complex CNN (CCNN) is proposed in [14] as radar data are inherently complex. The deep neural network can be also found in SAR imaging [15] with the presence of motion error. SAR imaging is formulated as a sparse recovery problem [15] and is solved by an unfolding neural network. The deep network is trained again by the simulated data. Experimental results demonstrate the success of transferring the learned model from simulated data to its application on the real radar data.

As mentioned earlier, traditional TFA methods are of great computational efficiency comparing to the sparse representation and compressive-sensing-based ISAR imaging. However, high resolution is still a big challenge for such algorithms due to the time–frequency (TF) tradeoff. Is it possible to achieve high-resolution images for maneuvering target by TFA-based ISAR imaging while preserving its low-computational efficiency? We propose to find a mapping function, which can be modeled nonlinearly as a deep neural network, to transform the low-resolution TF image into its high-resolution counterpart. For example, an improved U-net, named FBPConvNet [16], is proposed to improve the resolution of medical images. This inspires us to develop our super-resolution RID method using deep neural network-assisted linear TFA for maneuvering target ISAR imaging. We follow the basic structure of U-net with necessary modifications for super-resolution generation to model the mapping function. Inspired by Gao *et al.* [14] and Mason *et al.* [15], we propose to train the deep neural network by our designed low-resolution TF images of the simulated radar data and their theoretical high-resolution TF references. The pretrained network is then applied to generate high-resolution and denoised images of TF spectra from the corresponding low-resolution ones. High-resolution ISAR images can then be achieved by integrating those generated TF spectrum images into the traditional RID method. Since the neural network is pretrained, the proposed ISAR imaging method works as efficiently as the traditional RID with linear TFA.

The main contributions of our work are summarized as follows. First, training data have been generated to effectively train the model. Signals of radar echoes are simulated according to the mechanism of electromagnetic wave for point scatterers, and their noisy low-resolution TF spectrum images will be fed into the network as inputs. Since the ground truth is known for the simulated data, ideal noiseless high-resolution TF spectra can be easily computed and be served as network references. Second, necessary network structures are made onto FPBConvNet, an improved U-net, to achieve the purpose of high-resolution and noise suppression. Specifically, additional two convolutional upsampling layers are added to upgrade the resolution by four times. To further suppress the noise, the l_1-norm of the reconstruction error is chosen as the loss function as both predicted TF images and their references are extremely sparse. Finally, it is found that the mapping function, though learned from simulated signals, is powerful and can be well generalized to real radar signals as shown later in the experiment part using several real measured data. Comparing to the ISAR imaging by sparse representation, the proposed method maintains the advantage of computational efficiency as that of RID, at the same time, provides ISAR image with a high resolution for a maneuvering target.

It should be mentioned that this work is built on our prior study in [17]. Important improvements and extensions are made in this article. Comparing to the prior study in [17], two practical problems of super-resolution and noise suppression are further considered in this work, where the deep neural network with U-structure for super-resolution generation and denoising is accordingly designed. In addition, the algorithm robustness against noise and quadratic motion error is also investigated in this article.

The remainder of this article is organized as follows. In Section II, we concisely introduce the ISAR signal model and the TFA-based ISAR imaging method. In Section III, we present, in detail, our proposed deep neural network-assisted ISAR imaging for maneuvering targets, including the detailed structure of the neural network, training data generation, and network training process. In Section IV, numerical experiments are conducted to verify the effectiveness of the proposed method. Comparisons are made with other existing algorithms using both simulated and several sets of real radar data. Conclusions are drawn in Section V.

II. ISAR Model and Imaging via Time–Frequency Analysis

In this section, we briefly review two ISAR imaging systems based on RD and RID and discuss their limitations on the problem of maneuvering target imaging.

A. ISAR Signal Model

After the translational motion compensation [18], [19], target can be regarded as a turnable model rotating around its origin, as shown in Fig. 1, where the geometry of ISAR imaging is given. We define $P(x, y)$ as the location of a point scatterer on (x, y) plane, θ_0 as the original instantaneous angle of the target relative to radar, and $\theta(t)$ as the instantaneous rotation angle, where $0 \le t \le T$ is the slow-time and T is the whole observing duration. The instantaneous distance $R(t)$ from radar to the scatterer P can be written as

$$R(t) = R_0 + y\cos(\theta(t) - \theta_0) + x\sin(\theta(t) - \theta_0). \quad (1)$$

Define the change of the instantaneous rotation angle as

$$\Delta\theta(t) = \theta(t) - \theta_0. \quad (2)$$

$\Delta\theta(t)$ can be expanded into its first-order Taylor series

$$\Delta\theta(t) \approx \omega t + \frac{1}{2}\gamma t^2 \quad (3)$$

where ω represents the velocity of the rotation and γ is its acceleration. Assume that radar transmits the linear

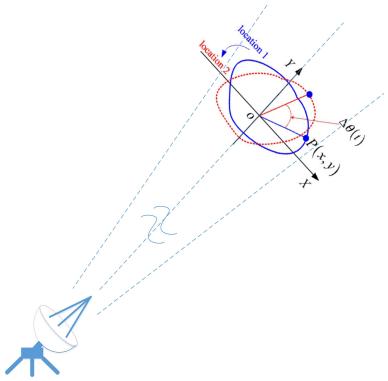

Fig. 1. Geometry of ISAR.

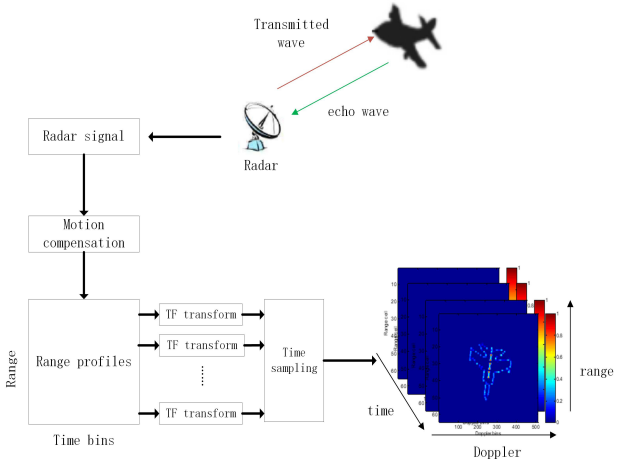

Fig. 2. Flowchart of ISAR imaging by RID.

frequency-modulated (LFM) signal

$$s(\tau) = \text{rect}\left(\frac{\tau}{T}\right)\exp\left\{j2\pi\left(f_c\tau + \frac{1}{2}k\tau^2\right)\right\} \tag{4}$$

where f_c is the carrier frequency, k represents the chirp rate, and τ denotes the fast time. Then, the echo of radar can be expressed as

$$S(\tau,t) \approx A \times \text{rect}\left(\frac{\tau - \frac{2R(t)}{c}}{T}\right) \times \exp\left\{j2\pi f_c\left(\tau - \frac{2R(t)}{c}\right) + j\pi k\left(\tau - \frac{2R(t)}{c}\right)^2\right\} \tag{5}$$

where A is the amplitude, c is the light velocity, and t represents the slow time. After range compression and motion compensation, (5) can be further expressed as

$$S(\tau,t) \approx A \times \text{sinc}\left\{Tk\left(\tau - \frac{2(R_0+y)}{c}\right)\right\} \times \exp\left\{-j\frac{4\pi}{\lambda}(R_0+y)\right\}\exp\left\{-j\frac{4\pi x}{\lambda}\left(\omega t + \frac{1}{2}\gamma t^2\right)\right\} \tag{6}$$

where λ is the wavelength. When target rotates in a constant manner, i.e., $\gamma \approx 0$, FT operated on the slow time τ in (6) can successfully achieve a focus target image. Procedures of range compression, motion compensation, and FT are collectively known as the range-Doppler (RD) imaging method. However, for maneuvering target with time-varying rotation, i.e., $\gamma \neq 0$, blurring in azimuth direction will be observed for RD method, which severely degrades target image quality.

B. ISAR Imaging via Time–Frequency Analysis

For maneuvering targets with time-varying Doppler frequency, TF transforms to estimate the instantaneous frequency are used to replace the traditional FT in RD. The resulted ISAR imaging system using different TF methods is known as the RID method in the literature.

Fig. 2 gives the flowchart of the RID-based radar imaging system. Prior to the TF image reconstruction, standard motion compensation has to be done. Then, TF transform instead of FT is applied along the azimuth for instantaneous Doppler estimation. Suppose that radar data consist of M slow-time instants, and each range profile at a certain time instant has a length of N. The Fourier-based RD will generate one 2-D RD image of size $M \times N$, while the TF-based RID will generate a 3-D time-RD image cube of size $M \times N \times M$ as one TF spectrum image will be given at each slow time slot. The 3-D time-RD image cube contains a focused RD image at each individual slow time slot, and the difference of images from one time slot to another provides the time-varying target attitude.

Different TFA can be used in the RID-based ISAR imaging framework. As discussed in Section I, linear TFA using STFT is of the greatest computational efficiency while suffering the problem of low resolution. WVD-based RID can achieve high image resolution at the cost of undesirable cross-terms due to its nonlinearity. In this article, we aim to build a new RID ISAR imaging system by using a deep neural network-assisted linear TFA to achieve a target image of high-resolution while maintaining its computational efficiency.

III. Neural Network-Assisted Super-Resolution ISAR Imaging

In this section, we first represent our signal model and review the typical STFT method. Then, we present our proposed deep-learning-assisted RID radar imaging system to solve the low-resolution problem of traditional linear TFA.

A. Time–Frequency Analysis Method

According to (6), for target with multiple scatterers, the radar signal in azimuth at certain range cell $R_0 + y$ can be considered as a linear combination of multiple LFM components

$$s(t) = \sum_{n=1}^{N} s_n(\tau,t) = \sum_{n=1}^{N} a_n \exp\left(j2\pi f_n t + j\pi k_n t^2\right) \tag{7}$$

where N represents the number of scatterers. a_n in (7) denotes the complex amplitude of the nth scatterer located at (x_n, y_n).

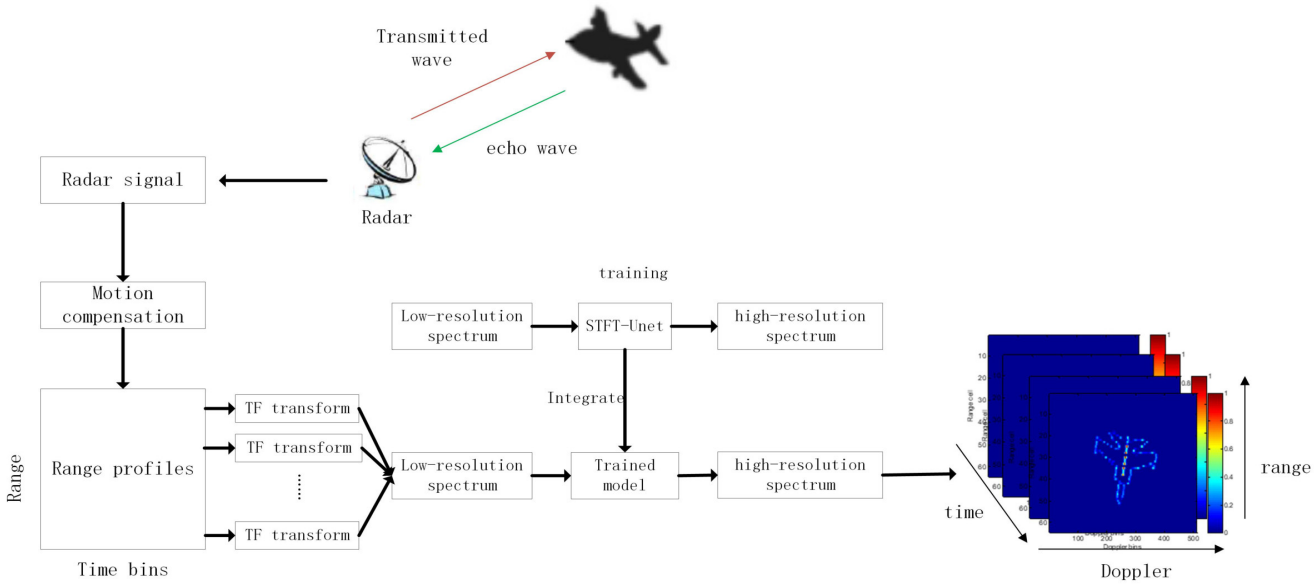

Fig. 3. Flowchart of ISAR imaging by RID based on the deep-learning-assisted TFA.

The complex amplitude a_n is composed of term A, sinc function, and the constant phase $\exp\{-j((4\pi)/\lambda)(R_0+y)\}$ in (6). The Doppler frequency is linearly modulated and expressed as

$$D_n = f_n + k_n t \tag{8}$$

where $f_n = 2x_n\omega_n/\lambda$ and $k_n = 2x_n\gamma_n/\lambda$. The above LFM signal model is commonly used for maneuvering target and is appropriate when a short observation time is considered. It is important to note that the proposed method is not restricted to the LFM signals. It can be easily extended to tackle more complicated signals with severe time-varying frequency, as the deep neural network is capable enough to deal with very complex images.

As seen from (8), the instantaneous Doppler frequency (IDF) is proportional to the azimuth location x_n at certain time t. Therefore, ISAR image of the scatterer can be obtained once its range and the IDF are known. To estimate the time-varying Doppler, the TF distribution STFT can be applied to (7) along the slow time t

$$\mathrm{STFT}(t,\omega) = \int_{-\infty}^{+\infty} s(\tau)h(\tau - t)e^{-j\omega\tau}\mathrm{d}\tau. \tag{9}$$

The spectrum is defined as follows [9]:

$$S(t,\omega) = |\mathrm{STFT}(t,\omega)|^2. \tag{10}$$

By submitting spectrum in (10) into the flowchart shown in Fig. 2, a series of target images in different slow times can be obtained, which is known as the STFT-based RID. However, the resolution is low due to the TF tradeoff.

B. Neural Network-Assisted Linear TFA for ISAR Imaging

To improve the spectrum resolution, we try to utilize a deep neural network as a mapping function to transform the low-resolution spectra to their high-resolution counterparts and incorporate them into the traditional RID to achieve high-resolution ISAR images for a maneuvering target. The whole process of our proposed algorithm can be summarized in Fig. 3.

The proposed method follows the basic flowchart of the RID-based ISAR imaging system. A well-pretrained neural network model is further incorporated to transform the low-resolution spectrum into the one with high-resolution. As shown in Fig. 3, low-resolution TF spectrum by STFT of each range cell is first fed into the trained model, and a high-resolution spectrum will be generated with concentrated IDF. At a particular slow time, a high-resolution ISAR image is obtained by extracting the frequency vectors from the TF spectrum and concatenating them according to their locations in range. The neural network in Fig. 3, named STFT-Unet, aims to assist the linear TFA to achieve a high-resolution spectrum estimation. The detailed structure and the training process of the STFT-Unet are given in the following.

C. Structure of STFT-Unet

In the article, we adopt the U-net, i.e., FBPConvNet, as the base network and necessary modifications are made for the purpose of super-resolution. U-net is well-known for its capability of image segmentation. The structure of the proposed network STFT-Unet is shown in Fig. 4. The network has a shape of "U," where the left-hand side is the contracting path and the right-hand side is the expanding path. The contracting path includes two convolution layers, each followed by an activation function of ReLU. The max-pooling operation with stride 2 is used to downsampling the feature map. The number of feature channels is doubled, and the size of the feature is half-reduced in each layer. The expanding path consists of an upsampling layer followed by a convolution layer that halves the number of feature channels and doubles the size of the feature. The operations of upsampling followed by convolution

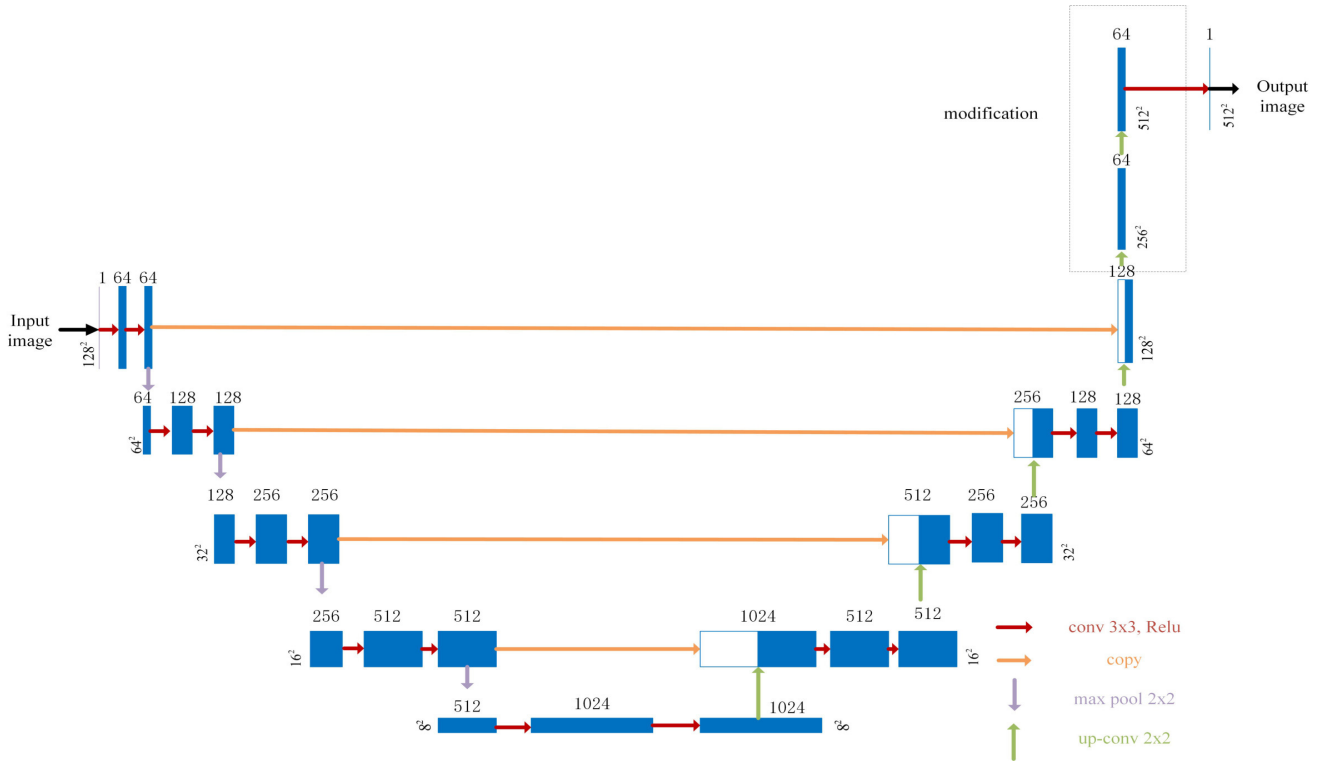

Fig. 4. Structure STFT-Unet, sizes of input image, and feature map are indicated at the bottom corner of each layer.

are called upconvolution. The resulted feature map is then concatenated with the previous one from the contracting path. Two additional convolution layers are added at the end of the network, each followed by na ReLU. It should be noted that four downsampling and six upsampling operations are used in total in our proposed network. The extra upconvolution operations are aimed to enlarge the size of the predicted image to match that of the preference image so that a resolution improvement of four times can be achieved. The detailed characteristics of the network can be summarized as follows, which are helpful for feature extraction, high-resolution image construction, and image denoising.

1) Multichannel Convolutional Layer: The convolutional layer is used to extract the feature information from the image. Different convolutional layers extract information on different levels. High-level features provide more semantic information, and low-level features retain more detailed and finer information [20]. The network STFT-Unet adopts multiconvolutional layers to extract features in different scales, which is similar to the telescopic operations of wavelet transform [21]. The extracted feature map is fundamental for high-resolution image reconstruction.

2) ReLU Activation Function: In order to model the nonlinear relationship between the input and output image, the activation function of ReLU can be used after each convolutional layer [22], [23]. In our application, since the prediction of TFA is nonnegative, ReLU is reasonably chosen as the activation function in the proposed network.

3) Pooling Layer: Redundant information can be retained by the convolutional layers, which is harmful to the efficient training of the network. In order to filter out the insignificance information, we adopt the max-pooling as the pooling layer in the STFT-Unet. The convolutional layer followed by the max-pooling is helpful to alleviate the problem of translation variance [24].

4) Information Sharing: In order to fully utilize both the low-level feature map and the high-level feature extracted by different layers, integration of multiscale information is used by feature concatenation to effectively improve the performance of CNN [25]. Concatenation of features from different convolutional layers is also found to prevent the gradient vanishing of the neural network [26].

5) Modifications: Two major modifications are made to FBPConvNet. The first is to add two upsampling layers at the end of the network to enlarge the predicted image by four times the input image. This step is important to achieve the super-resolution of the prediction. The second is the choice of the loss function, which determines how the predicted image approaches the reference. Here, we shortly discuss some well-known loss functions. One of the well-known loss functions is the cross-entropy [27], which is mainly used for classification and is not suitable for our task. Another loss function is the L_1-norm with the definition of $L_1(P) = (1/N)\sum_{p\in P}|\hat{y}(p) - y(p)|$, where p represents the pixel of the image, $y(p)$ is the pixel magnitude of the reference image, and $\hat{y}(p)$ is the pixel magnitude of the predicted image. L_2 loss is another commonly used loss function, which is defined as $L_2(P) = (1/N)\sum_{p\in P}|\hat{y}(p) - y(p)|^2$. In this application, both the predicted image and its ground-truth preference are sparse in nature as they are supposed to be the spectrum of

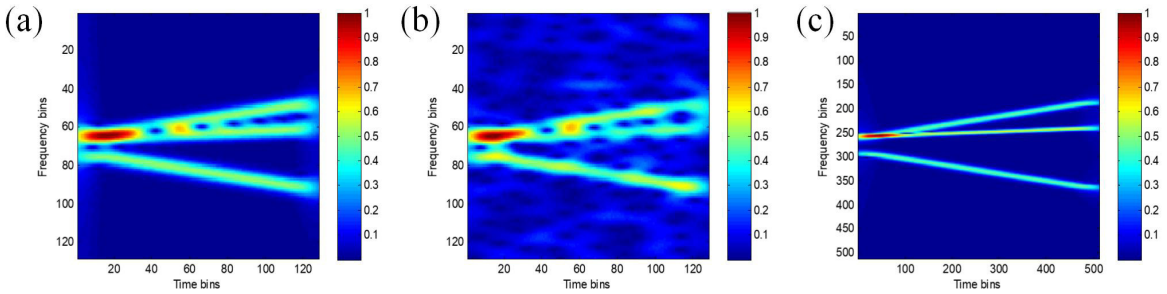

Fig. 5. TFD result. (a) Low-resolution STFT spectrum of size 128×128 (noiseless). (b) Low-resolution STFT of size 128×128 with a 5-db Gaussian noise (noisy). (c) Ideal high-resolution STFT of size 512×512 (reference image).

radar echo of several scatterers. The sparsity of the reference image and the prediction will be shown later in the training of the neural network. Therefore, L_1 loss function will be a reasonable distance measure between the prediction and the reference, which is utilized in our model.

D. Training of the Neural Network

With the whole structure of the neural network, we discuss in this section its training processing, including the generation of its input and reference. The inputs of the network are a series of STFT spectra of the simulated radar signals generated according to (7), where f_n is drawn from a uniform distribution within the range of 6.25, i.e., $U(-6.25, 6.25)$, γ_n is also drawn from $U(-6.25, 6.25)$, the number of scatterers N is randomly selected from 1 to 5, the pulse repetition frequency is set to be 64 Hz, and 128 pulses are generated within 2 s. To consider the noise effect, random Gaussian noise is added to the simulated signals with a signal-to-noise ratio (SNR) changing from 5 to 10 dB. STFT spectra of the simulated signals are computed according to (10) as the input of the network. In our setting, the input image is of size 128×128, and the reference image is expected to be of 512×512 since four times upsampling operations are implemented in the proposed network. In order to generate the corresponding reference image, both the resolutions of f_n and γ_n will be theoretically increased by a factor of 4. The reference image is generated according to (10). In summary, the network is trained to approximate the function, which can transform the noisy low-resolution spectral image to its noiseless ideal one with high resolution.

An illustrative example of designed input and references is given in Fig. 5. In this example, three LFM signals are used, and the resolution of STFT is low. Fig. 5(b) shows the noise effect, where shifts and discontinuity can be observed for the instantaneous frequency given by STFT due to the noise. The noiseless high-resolution TFD is shown in Fig. 5(c) as the reference image of the neural network.

It is worth noting that the injection of random noise to the input image not only mimics the real radar signal but also enables the network the potential of noise suppression. The proposed network is implemented by the Tensorflow toolkit. The parameter settings of STFT-Unet are summarized in Table I.

TABLE I
TRAINING PARAMETERS

Number of training data	500
Loss function	L1
Gradient optimization	Adam
Learning rate	0.0001
Batch_size	1
Epoch	2

IV. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

In this section, extensive numerical experiments are conducted to verify the effectiveness of the proposed method. First, the quality of the predicted image by our pre-trained network is evaluated in terms of reconstruction error and noise suppression. Then, comparisons using the Mig-25 radar data set are made for ISAR imaging between the proposed method and several RID baselines. The robustness of the proposed algorithm is tested under the scenarios of noisy data and data with quadratic motion error. Finally, several sets of real measured data from different targets are used to test the performance of the proposed method.

A. Prediction Quality Evaluation

The network of STFT-Unet can be trained using the training process discussed in Section III-D. In order to evaluate the prediction performance of the trained STFT-Unet network in terms of resolution and the effect of denoising, the noiseless STFT spectrum of low-resolution and the one with 5-dB noise are fed into the trained STFT-Unet. The input and the predicted images are shown in Fig. 6, where images with higher-resolution are given by the STFT-Unet both in noisy and noiseless cases. In the noiseless case, the predicted image in Fig. 6(d) is very close to the reference image with an obvious improvement in its resolution. Fig. 6(e) shows the predicted image from the trained STFT-Unet with noisy input of an SNR of 5 dB. Comparing to the input image in Fig. 6(b), it can be found that the proposed STFT-Unet has a strong capability of noise suppression.

To quantitatively analyze the prediction performance of STFT-Unet, we establish three indicators to evaluate the predicated image quality and the final ISAR image quality. One is

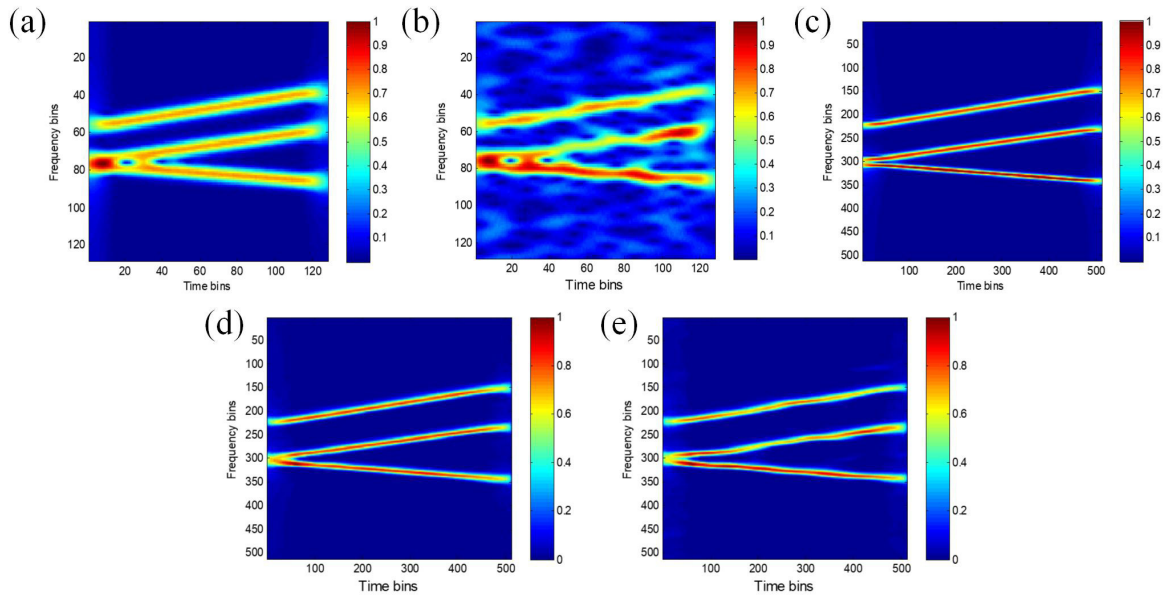

Fig. 6. TFD result. (a) Noiselss STFT. (b) STFT with a 5-dB noise. (c) Reference image. (d) Prediction of trained STFT-Unet with input in (a) and (e) prediction of trained STFT-Unet with input in (b).

the mean square error (mse) expressed as follows:

$$\text{mse} = \frac{\sum_{i=1,j=1}^{m,n} \left(y_{i,j} - y_{i,j}^{r}\right)^2}{mn} \tag{11}$$

where $y_{i,j}^{r}$ represents the reference image, and $y_{i,j}$ is the image to be evaluated. The mse evaluates the level of similarity between $y_{i,j}^{r}$ and $y_{i,j}$. A lower mse indicates a higher similarity.

Another indicator is the SNR improvement, denoted as "Imp_SNR," which is calculated by

$$\text{Imp\_SNR} = 100 \times \frac{(\text{SNR}_{\text{STFT-Unet}} - \text{SNR}_{\text{STFT}})}{\text{SNR}_{\text{STFT}}}\% \tag{12}$$

where $\text{SNR}_{\text{STFT-Unet}}$ represents the estimated SNR of image predicted by STFT-Unet and SNR_{STFT} represents the estimated SNR of the original noisy input. The SNR of the image is estimated according to

$$\text{SNR} = 10 \log 10 \left(\frac{E\_\text{target}}{E\_\text{noise}}\right) \tag{13}$$

where E_{target} and E_{noise} represent the estimated energies of target and noise, respectively. Since the ground truth of the signal parameters is known, the target region can be masked out to estimate the target energy. Noise energy is then estimated by ruling out the target region. Higher Imp_SNR shows a better noise suppression performance. Both mse and Imp_SNR evaluate the quality of the predicted high-resolution TF spectral images. In order to qualitatively evaluate and compare the quality of target ISAR images obtained by different methods, the metric of image entropy is calculated, which will be used in the testing experiments later in Sections IV-D and IV-F.

The mses between evaluation and reference images are shown in Table II. For images of size 128×128, we interpolate them to the size of 512×512 when calculating the mse. It can be seen from Table II that lower mses are achieved by the predicted images than the original input images. The metric of Imp_SNR is calculated for images in Fig. 6(b) and (e). We can qualitatively achieve an SNR improvement of $\text{Imp\_SNR} = 41.12\%$.

TABLE II
MSE BETWEEN THE EVALUATED IMAGE AND REFERENCE IMAGE

$y_{i,j}$	$y_{i,j}^{r}$	MSE
noiseless STFT as Fig.6(a)	reference image as Fig.6(c)	0.0207
noisy STFT of a SNR of 5dB as Fig.6(b)	reference image as Fig.6(c)	0.0372
prediction of STFT-Unet with noiseless input as Fig.6(d)	reference image as Fig.6(c)	0.0014
prediction of STFT-Unet with noisy input as Fig.6(e)	reference image as Fig.6(c)	0.0026

TABLE III
CONSUMING TIME OF DIFFERENT MODELS

Methods	consuming time (s)
RSP	0.0422
STFT	0.0013
STFT-Unet	0.3413

As aforementioned, the proposed model is trained off-line, and making a higher resolution prediction merely needs a forward computation, which is computational efficient. In order to show how efficient it is, the running time comparison of different methods is made and given in Table III. It takes 0.341 s for the proposed model to make a high-resolution prediction of the image, which is fast within 1 s though slightly slower than simply performing the TFA.

B. ISAR Imaging Using Mig-25

In this section, different TFA methods are compared with the proposed method on ISAR imaging by using the data set Mig-25 provided by the U.S. Naval Research

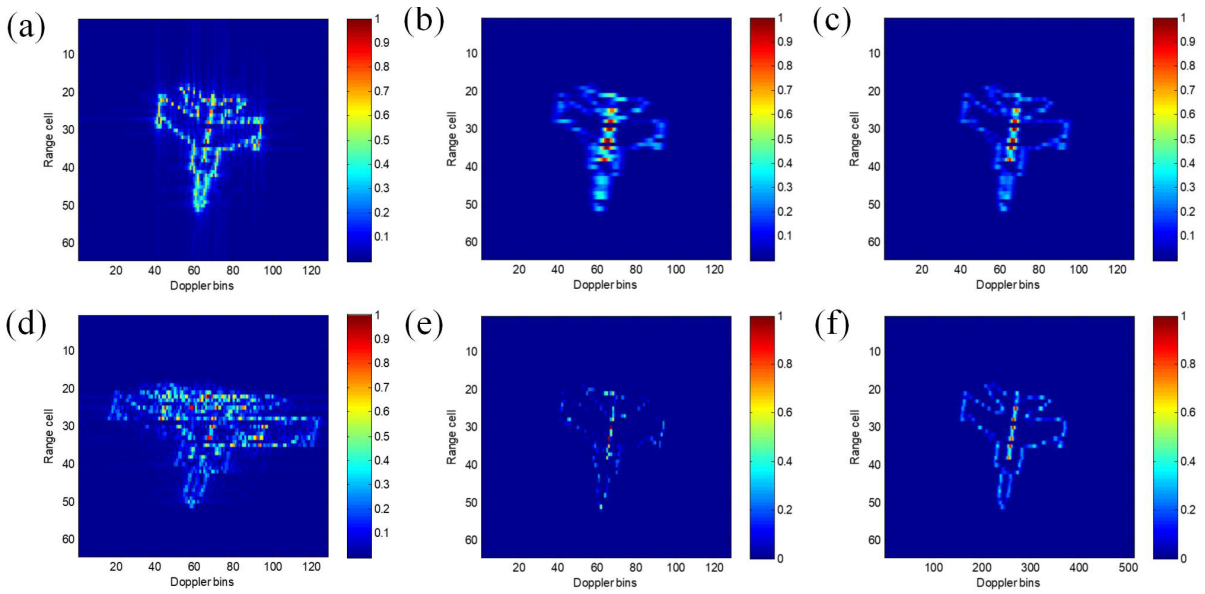

Fig. 7. ISAR imaging result by RID using different TFA methods. (a) RD. (b) STFT. (c) Gabor. (d) WVD. (e) RSP. (f) STFT-Unet.

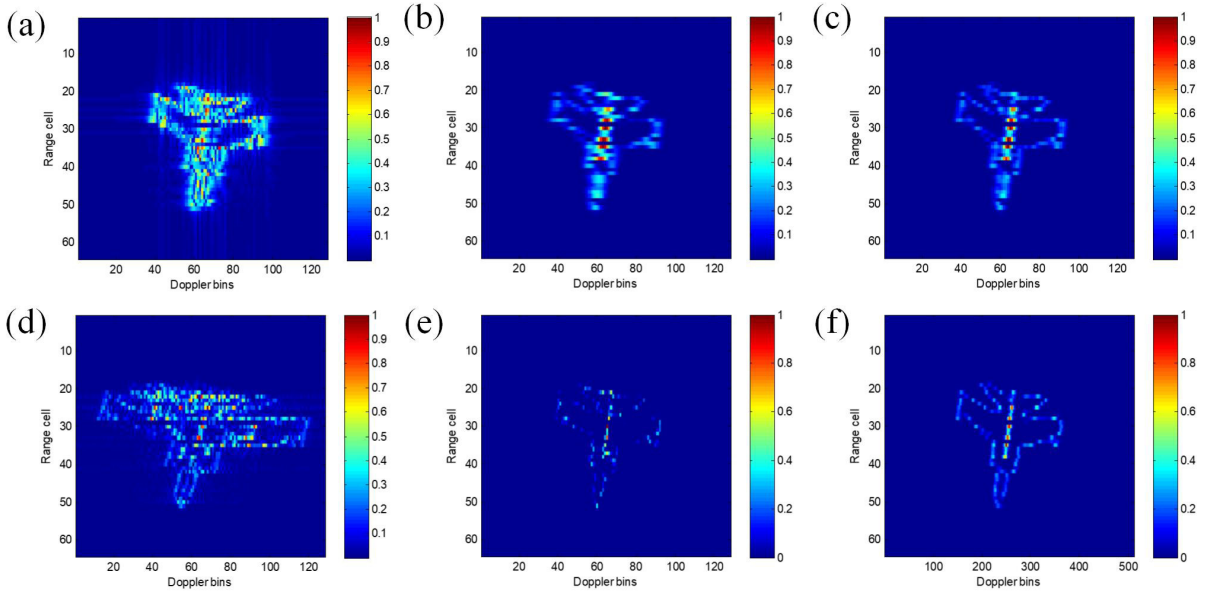

Fig. 8. ISAR imaging result under the presence of motion error. (a) RD. (b) STFT. (c) Gabor. (d) WVD. (e) RSP. (f) STFT-Unet.

Laboratory available online: http://airborne.nrl.navy.mil/~vchen. The radar system parameters of data set Mig-25 are summarized in Table IV. The 128 pulses with a dwell time of range from -0.0043 to 0.0043 are used in this experiment. A variety of TFAs are used in RID for ISAR imaging, including WVD, Gabor transform, and RSP. The obtained target images are given in Fig. 7.

In this experiment, since a small number of pulses are used, the target is approximate of a uniform motion. Therefore, traditional RD can achieve a focused target image, as shown in Fig. 7(a). The ISAR imaging results generated by STFT and Gabor transform are presented in Fig. 7(b) and (c), where the obtained target images are of low resolution. In particular, target scatterers on the head of the aircraft cannot be separated.

TABLE IV
RADAR SYSTEM PARAMETER SETTING FOR THE MIG-25 DATA SET

Carrier frequency	9GHz
Bandwidth	512MHz
Range bin	64
Azimuth bin	128
Pulse repetition frequency	15 KHz

Linear TFAs of STFT and Gabor transform cannot yield accurate and detailed characteristics of the target due to their limitation on the resolution. Fig. 7(d) and (e) shows the ISAR imaging results generated RID using WVD and RSP. It can be found that the target images generated by WVD and RSP

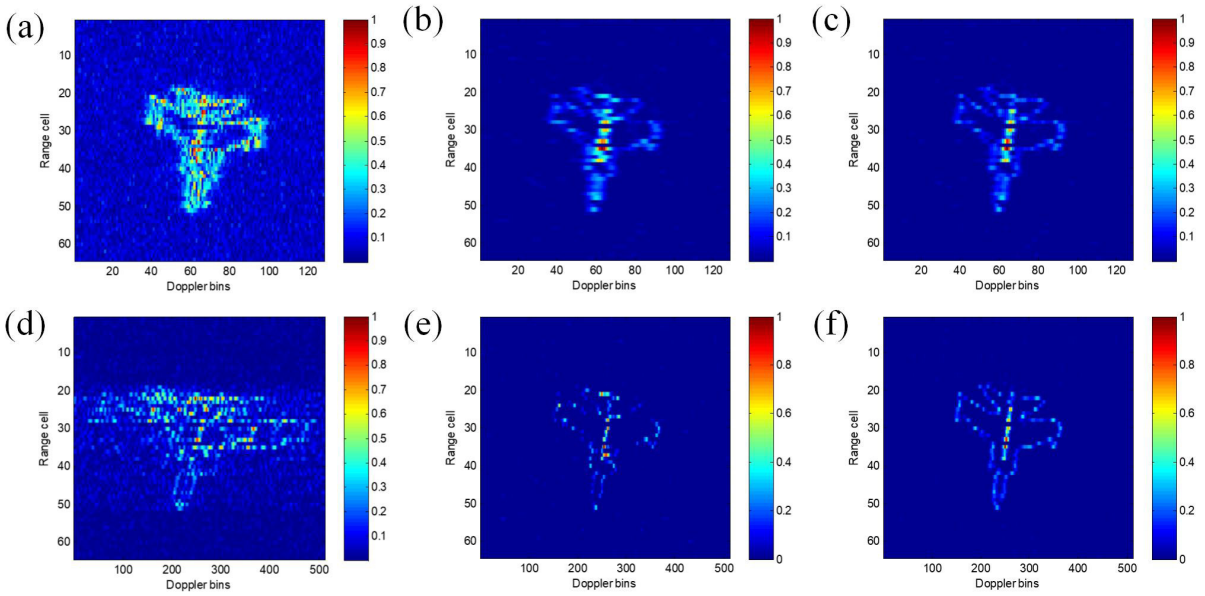

Fig. 9. ISAR images achieved by different algorithms in presence of the motion error and noise. (a) RD. (b) STFT. (c) Gabor. (d) WVD. (e) RSP. (f) STFT-Unet.

have higher resolution than the linear TFAs. However, since the target scatterers at the same range cell are of the same frequency change, severe undesirable cross-term will dominate the TF spectrum by WVD. As shown in Fig. 7(d), the quality of the ISAR image generated by the WVD-based RID is low due to the undesirable cross-term. Although the TFD generated by RSP has no cross-term, RSP suffers the problem of missing target scatterers. As shown in Fig. 7(e), many scatterers on the right-wing of the aircraft have been eliminated incorrectly. The ISAR image generated by RSP cannot provide a complete characteristic of the target. Fig. 7(f) shows the ISAR image achieved by RID with the STFT-Unet-assisted linear TFA. It can be seen that the STFT-Unet-assisted linear TFA not only achieves super-resolution ISAR images but also avoids the problem of cross-term. We also find that the shape of aircraft generated by the trained STFT-Unet is closer to the shape of aircraft generated by RD than other methods.

C. ISAR Imaging Using Mig-25 With Quadratic Motion Error

To test the robustness of the proposed method on ISAR imaging for maneuvering targets, quadratic motion error is manually added to the Mig-25 data set according to

$$\text{error}(\tau, t) \approx A \times \exp\left\{-j\frac{4\pi}{\lambda}at^2\right\} \tag{14}$$

where $A = 1$, $a = 1000$, and $\lambda = (c/f_c)$ with $c = 3 \times 10^8$ and $f_c = 9$ GHz. The motion error is added into the radar echo of each range cell. The resulted Mig-25 data with motion error are used to test the aforementioned methods, and the obtained target images are shown in Fig. 8. Due to the motion error, target of Mig-25 is no longer of the uniform motion. ISAR image generated by RD is, therefore, blurring in azimuth, as shown in Fig. 8(a).

Comparing to the image given by RD in Fig. 8(a), linear TFA-based RIDs generally produce promising ISAR images. However, the resolution of linear TFA of STFT and Gabor transform is too low to distinguish scatterers on the target. RSP and the proposed method yield focused ISAR images with high resolution. However, important scatterers on the wings and head of the aircraft are missing on the image achieved by RSP, as shown in Fig. 8(e). The image given by the proposed method is focused, complete, and of a much higher resolution.

D. ISAR Imaging Using Mig-25 With Presence of Motion Error and Noise

The robustness of the proposed method against noise is demonstrated using the Mig-25 data with motion error in this part. In addition to the motion error, 5-dB random Gaussian noise is added to the Mig-25 data set. The image given by RD using the noisy data set is given in Fig. 9(a), where an unfocused target image completely buried in noise is observed. ISAR images obtained by different TFA-based RID methods are given in Fig. 9(b)–(e), where images with the relatively clearer background are shown as RID methods remove parts of the noise by TF spectrum.

Images with low resolution are noted in Fig. 9(b) and (c), while RSP cannot provide a continuous and complete target shape, and the problem is worsened in a noisy situation. WVD completely fails in this complex scenario. The ISAR image generated by the proposed method is given in Fig. 9(f), which shows a focused aircraft with higher resolution and clearer background. Comparing to the RD method, the proposed method demonstrates a strong noise suppression capability, which is built on two successive steps of TF representation and U-net transformation.

As shown in Table V, both the RSP and the proposed method achieve smaller entropy as the images are of high-resolution and well-focused. A smaller entropy value also indicates a better denoising effect as the background noise is

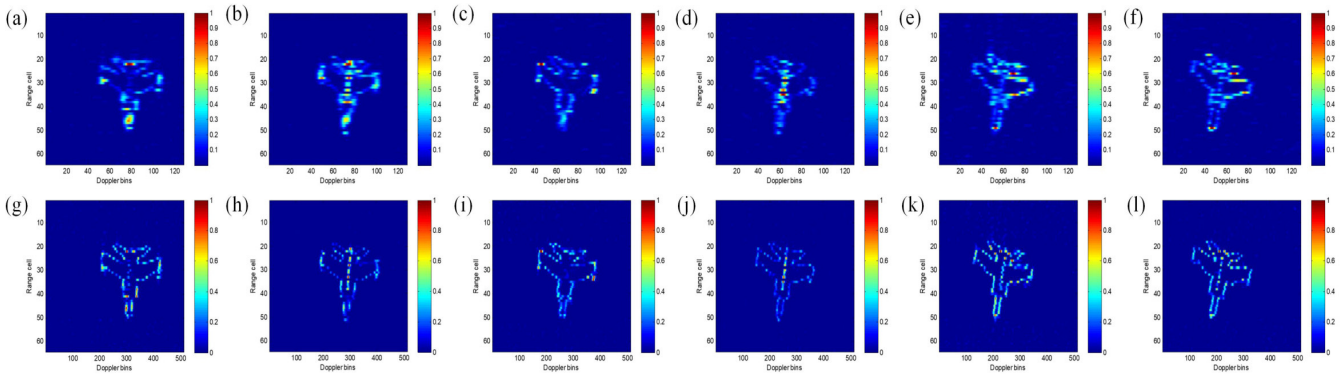

Fig. 10. Target attitude change against slow time achieved by (a) STFT with pulses from 1 to 128, (b) STFT with pulses from 65 to 192, (c) STFT with pulses from 129 to 256, (d) STFT with pulse from 193 to 320, (e) STFT with pulses from 257 to 384, (f) STFT with pulses from 321 to 448, (g) STFT-Unet with pulses from 1 to 128, (h) STFT-Unet with pulses from 65 to 192, (i) STFT-Unet with pulses from 129 to 256, (j) STFT-Unet with pulses from 193 to 320, (k) STFT-Unet with pulses from 257 to 384, and (l) STFT-Unet with pulses from 321 to 448.

TABLE V
ENTROPY OF ISAR IMAGES (WITH AN SNR OF 5 dB)

Methods	Entropy
RD	7.5072
RSP	5.6282
WVD	8.2380
Gabor Transform	7.0935
STFT	6.9907
STFT-Unet	6.3439

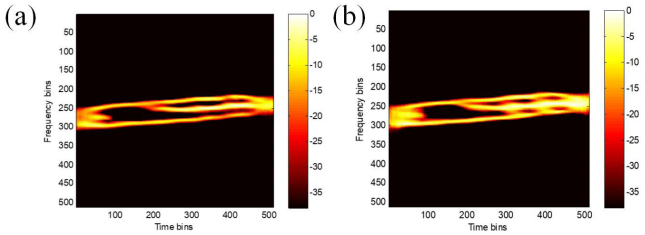

Fig. 11. Predicted TF spectral images by models with different losses using the Mig-25 data set under the SNR of 5 dB. (a) L1-norm loss. (b) L2-norm loss.

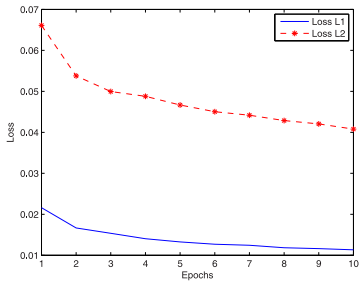

Fig. 12. Convergence of models with different losses in training process.

TABLE VI
RADAR SYSTEM PARAMETER SETTING FOR THE B-727 DATA SET

Carrier frequency	9GHz
Bandwidth	150MHz
Range bin	128
Azimuth bin	128
Pulse repetition frequency	15 KHz

suppressed and a clearer target image is achieved. Since RSP tends to oversparsify the target image, which favors a relative smaller image entropy, important target information is missing, and the resulted image is poor.

E. Aircraft Attitude

As mentioned in Section II-B, one advantage of TFA-based RID is to provide a 3-D time-RD image cube, which shows changes in the target attitude over slow time. In this section, the whole Mig-25 data set of 512 pulses with motion error and artificially added noise is used to show the change of the target attitude; six sets of overlapped pulses in slow time with a size of 128 are selected, i.e., pulses from 1 to 128, 65 to 192, . . . , 321 to 448. A series of target images are presented in Fig. 10, where images achieved by the STFT-based RID are given as comparisons. Both algorithms show that the target's head rotates from right to left along slow time. The proposed method provides more target details as the resolution is greatly improved. Higher image contrast can also be observed for the images obtained by the proposed method as the background noise is further suppressed.

To illustrate the advantages of the L1-norm over the loss function of L2-norm for our proposed network, comparisons are made from two perspectives by using the noisy Mig-25 data set. One is to compare the resolution of the predicted results, and the other is to compare the convergence of the models with different losses. The predicted results in Fig. 11 show that the model with L1 loss achieves a prediction with a relatively higher resolution. Convergence curves in Fig. 12 also suggest a faster convergence rate of the model with l1-norm loss.

F. Experimental Results on Measured Data

In this section, the proposed method is tested using several real measured data sets, including data B-727 and data of three different ships. The B-727 data set can be found online: http://airborne.nrl.navy.mil/~vchen. The radar system parameters of the data set B-727 are shown in Table VI.

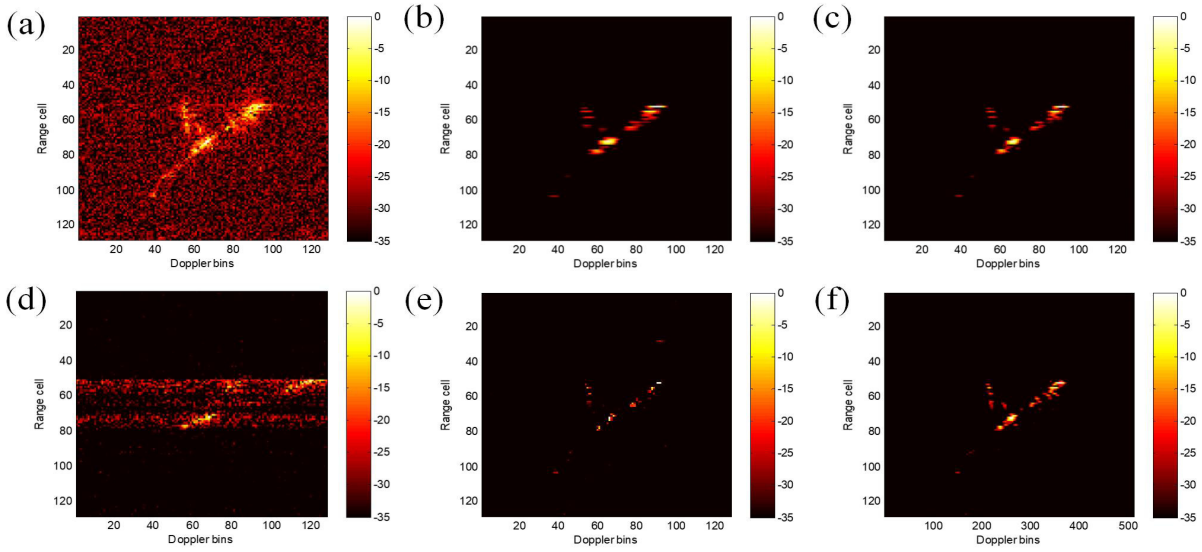

Fig. 13. ISAR images achieved by different methods using the B-727 data set. (a) RD. (b) STFT. (c) Gabor transform. (d) WVD. (e) RSP. (f) STFT-Unet.

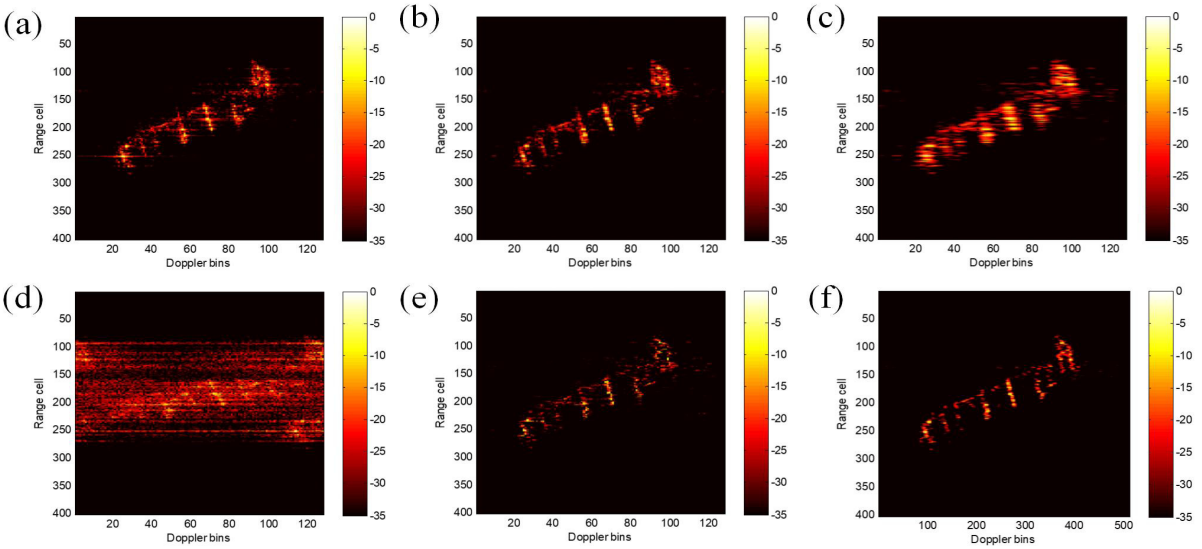

Fig. 14. ISAR images achieved by different methods using the ship data set-I. (a) RD. (b) STFT. (c) Gabor transform. (d) WVD. (e) RSP. (f) STFT-Unet.

TABLE VII
RADAR SYSTEM PARAMETERS FOR SHIP DATA

Carrier frequency	9GHz
Bandwidth	600MHz
Range bin	128
Azimuth bin	128
Pulse repetition frequency	300Hz

TABLE VIII
ENTROPY OF THE TARGET ISAR IMAGES

Methods	B-727	Ship 1	Ship 2	Ship 3
RD	7.7082	6.3496	7.7852	5.3949
RSP	2.9122	5.4898	6.3141	4.2009
WVD	5.9804	8.7614	9.6524	7.8917
Gabor	3.4384	6.8852	8.1532	5.7924
STFT	3.6518	6.1138	7.6924	5.5079
STFT-Unet	3.4325	5.7505	6.6557	4.7607

Note that the results of real measured data are given in the heat map and scaled in dB for better visualization. It is noticed that data B-727 is very noisy since its target image given by RD is completely immersed in noise, as shown in Fig. 13(a). Images with the clear background are obtained by STFT, Gabor transform, RSP, and the proposed STFT-Unet-based RID methods. In Fig. 13(e) and (f), images given by RSP and the proposed method show an obvious resolution improvement. However, it is observed in Fig. 13(e) that fake scatterers appear in the lower side of the target image given by RSP in such a strong noisy case, which is misleading and can be mistakenly treated as a lower wing. As shown in Fig. 13(f), STFT-Unet can improve the resolution of TFD without excessively weakening the scatter energy of the target. A focused target image with high-resolution can be obtained by the proposed method.

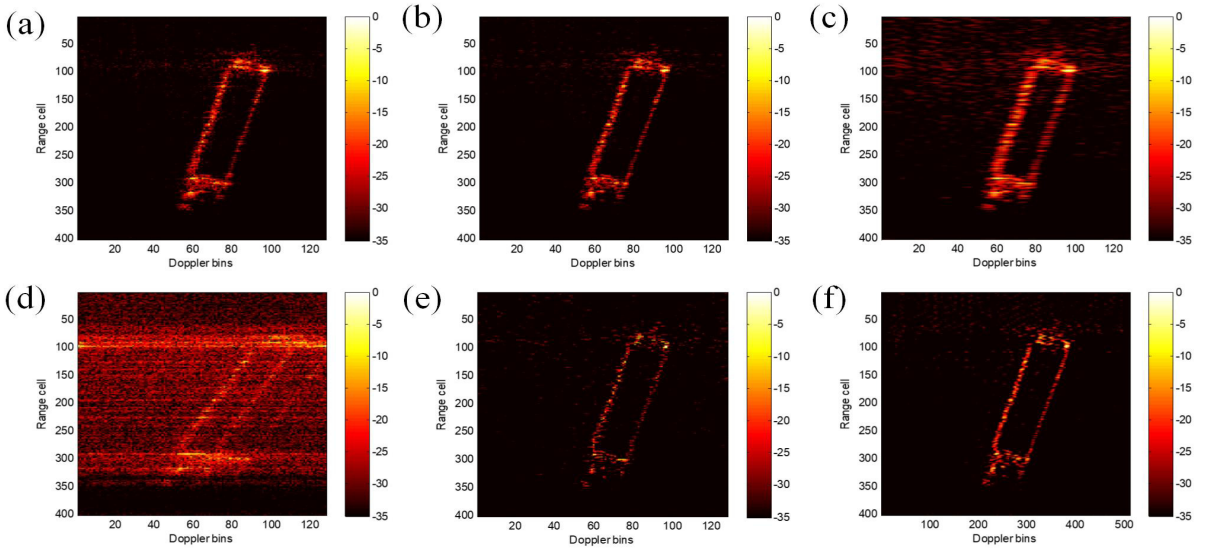

Fig. 15. ISAR images achieved by different methods using the ship data set-II. (a) RD. (b) STFT. (c) Gabor transform. (d) WVD. (e) RSP. (f) STFT-Unet.

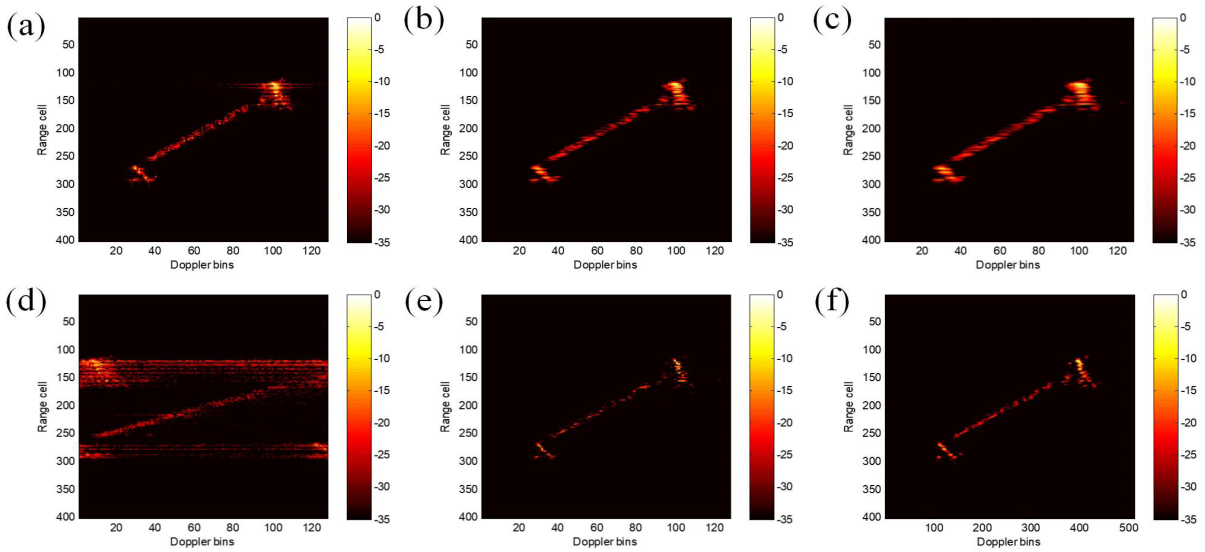

Fig. 16. ISAR images achieved by different methods using ship data set-III. (a) RD. (b) STFT. (c) Gabor transform. (d) WVD. (e) RSP. (f) STFT-Unet.

Data of three ship targets are tested in order to show the robustness of the proposed algorithm over the number of target scatterers. The radar system parameters are listed in Table VII. Ships in the data differ from each other in two aspects. One is the number of target scatterers, and the other is their motion velocities, which can be observed in Figs. 14(a), 15(a), and 16(a), where mainlobes of scatterers on three ships are broadened to different extents by the RD method. The focusing degrees of images indicate the motion difference among targets. Similar results as those of data B-727 are observed for the ship data despite their differences in numbers of scatterers and motion conditions. The image entropy obtained by different methods using the four real data sets is summarized in Table VIII, where our proposed STFT-Unet and RSP achieve the best image entropy. The proposed algorithm consistently provides the best target images with high resolution and improved image contrast. A careful comparison among images of different ships also shows a greater image quality improvement of the proposed method over the traditional RD if the image of RD is less focused.

V. Conclusion

In order to achieve super-resolution ISAR images for maneuvering target, a new RID method using deep neural network-assisted linear TFA is proposed. The neural network ensembles the structure of Unet and is trained by samples of pairs of low-resolution spectrum image and its high-resolution reference to learn the inherent mapping function. The trained STFT-Unet-assisted TF can not only predict a high-resolution spectrum without any undesirable cross-term but also suppress the background noise. Numerical experimental results demonstrate the effectiveness of the proposed method. As the STFT-Unet is pretrained, the prediction of the

high-resolution spectrum costs little. The proposed deep neural network-assisted ISAR imaging is as efficient as the traditional linear TFA-based RID method.

It is noticed that other more complex structures of the deep neural network could be possibly applied to replace U-net for the task of super-resolution image reconstruction and denoising. One of the further works could be the exploration of those networks and their adaptation to radar target images either by transfer learning or necessary modifications.

REFERENCES

[1] M. Soumekh, *Synthetic Aperture Radar Signal Processing*. New York, NY, USA: Wiley, 1999.

[2] J. Walker, "Range-Doppler imaging of rotating objects," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 23–52, Jan. 1980.

[3] L. Zhang *et al.*, "Resolution enhancement for inversed synthetic aperture radar imaging under low SNR via improved compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3824–3838, Oct. 2010.

[4] L. Zhao, L. Wang, L. Yang, A. M. Zoubir, and G. Bi, "The race to improve radar imagery: An overview of recent progress in statistical sparsity-based techniques," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 85–102, Nov. 2016.

[5] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5736–5750, Sep. 2014.

[6] L. Wang, L. Zhao, G. Bi, and C. Wan, "Sparse representation-based ISAR imaging using Markov random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3941–3953, Aug. 2015.

[7] C. Sun, B. Wang, Y. Fang, K. Yang, and Z. Song, "High-resolution ISAR imaging of maneuvering targets based on sparse reconstruction," *Signal Process.*, vol. 108, pp. 535–548, Mar. 2015.

[8] V. C. Chen and S. Qian, "Joint time-frequency transform for radar range-Doppler imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 486–499, Apr. 1998.

[9] X.-G. Xia, G. Wang, and V. C. Chen, "Quantitative SNR analysis for ISAR imaging using joint time-frequency analysis-short time Fourier transform," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 2, pp. 649–659, Apr. 2002.

[10] J.-H. Park and N.-H. Myung, "Enhanced and efficient ISAR image focusing using the discrete Gabor representation in an oversampling scheme," *Prog. Electromagn. Res.*, vol. 138, pp. 227–244, 2013.

[11] P. Jun, L. Dane, W. Guang-ming, and Y. Haotian, "Reconstruction of ISAR imaging using time-frequency distribution series method," in *Proc. 1st Asian–Pacific Conf. Synth. Aperture Radar*, Nov. 2007, pp. 351–354.

[12] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency reassignment: From principles to algorithms," in *Applications in Time-Frequency Signal Processing*, A. Papandreou-Suppappola, Ed. Boca Raton, FL, USA: CRC Press, 2003, ch. 5, pp. 179–203.

[13] J. Qian, G. Bi, and X. Lv, "ISAR maneuvering target imaging based on compressive time-frequency distribution," in *Proc. 13th Int. Radar Symp.*, May 2012, pp. 430–434.

[14] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "Enhanced radar imaging using a complex-valued convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 35–39, Jan. 2019.

[15] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1703–1708.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[17] S. Huang, J. Qian, Y. Wang, X. Yang, and L. Yang, "ISAR maneuvering target imaging based on convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 2551–2554.

[18] Q. Zhang, T. S. Yeo, G. Du, and S. H. Zhang, "Estimation of three-dimensional motion parameters in interferometric ISAR imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 2, pp. 292–300, Feb. 2004.

[19] J. Wang and D. Kasilingam, "Global range alignment for ISAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 351–357, Jan. 2003.

[20] Z. Cui, S. Dang, Z. Cao, S. Wang, and N. Liu, "SAR target recognition in large scene images via region-based convolutional neural networks," *Remote Sens.*, vol. 10, no. 5, p. 776, 2018.

[21] S. Mallat, "Understanding deep convolutional networks," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, Art. no. 20150203.

[22] G.-B. Huang and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Trans. Neural Netw.*, vol. 9, no. 1, pp. 224–229, Jan. 1998.

[23] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[24] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associate, 2015, pp. 2017–2025. [Online]. Available: http://papers.nips.cc/paper/5854-spatial-transformer-networks.pdf

[25] C. Du, Y. Wang, C. Wang, C. Shi, and B. Xiao, "Selective feature connection mechanism: Concatenating multi-layer CNN features with a feature selector," *Pattern Recognit. Lett.*, vol. 129, pp. 108–114, Jan. 2020.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[27] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.

Jiang Qian (Member, IEEE) received the B.Sc. degree in statistics and the M.Sc. degree in applied mathematics from Northwestern Polytechnical University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in signal and information processing from the National Lab of Radar Signal Processing, Xidian University, Xi'an, in 2011.

From 2011 to 2012, he was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From 2012 to 2013, he was a Post-Doctoral Research Fellow with the Center for Advanced Communications, Villanova University, Villanova, PA, USA. He is an Associate Professor with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, China. His major research interests include through-the-wall radar imaging, synthetic aperture radar (SAR)/inverse SAR imaging, interferometric SAR, and ground moving-target indication.

Shaoyin Huang received the B.Eng. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2017, where he is pursuing the M.Eng. degree.

His research interests include deep learning and radar imaging.

Lu Wang received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from Xidian University, Xi'an, China, in 2007 and 2010, respectively, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2014.

From 2015 to 2018, she was with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an. Since 2019, she has been a Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. Her major research interests include sparse Bayesian learning, radar signal processing, and the applications of machine learning.

Guoan Bi (Senior Member, IEEE) received the B.Sc. degree in radio communications from the Dalian University of Technology, Dalian, China, 1982, and the M.Sc. degree in telecommunication systems and the Ph.D. degree in electronics systems from the University of Essex, Colchester, U.K., in 1985 and 1988, respectively.

Since 1991, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include digital signal processing (DSP) algorithms and hardware structures, and signal processing for various applications, including sonar, radar, and communications.

Xiaobo Yang received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1986 and 1988, respectively.

He is also a Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include multiple-input multiple-output radar, through-the-wall radar, and statistical signal processing.