Interference Mitigation for Automotive FMCW Radar Based on Contrastive Learning With Dilated Convolution

Wang, Jianping; Li, Runlong; Zhang, Xinqi; He, Yuan

Interference Mitigation for Automotive FMCW Radar Based on Contrastive Learning With Dilated Convolution

Jianping Wang, *Member, IEEE*, Runlong Li, Xinqi Zhang, and Yuan He, *Senior Member, IEEE*

Abstract— As one of the crucial sensors for environment sensing, frequency modulated continuous wave (FMCW) radars are widely used in modern vehicles for driving assistance/autonomous driving. However, the limited frequency bandwidth and the increasing number of equipped radar sensors would inevitably cause mutual interference, degrading target detection and producing safety hazards. In this paper, a deep learning-based interference mitigation (IM) approach is proposed for FMCW radars by using the dilated convolution for network construction and a designated contrast learning strategy for training. The dilated convolution enlarges the receptive field of the neural network, and the designated contrastive learning strategy enforces to distinguish better between interferences and desired signals. The results of numerical simulation and experimental data processing show that the dilated convolution-based IM network, compared to the traditional convolution-based ones, can achieve a higher Signal-to-Interference-plus-Noise-Ratio (SINR) and target detection rate. Moreover, the designated contrastive learning strategy enables a better and more stable IM performance without increasing the complexity of the network, which can facilitate faster signal processing.

Index Terms— Automotive radar, interference mitigation, deep learning, dilated convolution, contrastive learning.

I. INTRODUCTION

MODERN vehicles are equipped with a variety of sensors to perceive the surroundings for driver assistant system (ADAS)/full self-driving [1], [2]. Due to the advantages of low cost, long detection range, and adaptability to most weather and light conditions, frequency modulated continuous wave (FMCW) radar is an indispensable sensor for target's range or speed measurement, guard rail detection, etc [3], [4]. However, with the increasing number of radar sensors mounted on vehicles and other wireless devices, the probability of mutual interference has significantly increased [5].

FMCW radar works by transmitting a simple continuous wave modulated in frequency periodically. The interference from the aggressor radar with a different sweep slope from that of the victim radar may result in increased noise floor, overwhelming weaker targets and degrading the detection probability. On the other hand, the interference with the same sweep slope as the victim radar may cause a ghost target. Hence, mutual interference of FMCW radars has become a severe problem, and suppressing the interferences is crucial to guarantee accurate target detection of the FMCW radar systems. Due to the very low probability of interference with the same sweep slope, we will focus on, without loss of generality, the problem of suppressing the interference with a different sweep slope in this paper.

The interference mitigation (IM) approaches for FMCW radars can be classified into three categories: radar system coordination [6], radar waveform/system design [7], and signal processing methods [8], [9], [10], [11], [12], [13], [14], [15]. Considering the signal processing methods can be easily fit into the existing radar system, and the deep learning (DL)-based approaches have made great progress in signal processing problems, the DL-based interference mitigation methods for FMCW radars are investigated. In [9], the received beat signals after dechirping are used as the input of a recurrent neural network (RNN) with attention layers, which outputs the interference-suppressed signal samples. In the other DL-based approaches, the two-dimensional (2D) frequency spectrum [10], time-frequency (t-f) diagram [11], or the range-Doppler (RD) map [12], [13], [14], [15] of the beat signal contaminated by interferences is fed into a convolutional neural network (CNN). Then, the CNN can extract different features of the targets' beat signals and interferences, separating the desired beat signals from interferences for further processing. Due to the significantly different distributions of the targets' beat signals and the interferences in the t-f domain than that in the time domain or frequency domain, [11] has demonstrated the superior performance of interference mitigation in the t-f domain compared to approaches implemented in other domains. So, in this paper, the interference mitigation operation in the t-f domain would be adopted.

Although the aforementioned DL-based IM approaches for FMCW radars have shown a good performance, most of them chose to use a complex network architecture with many parameters (e.g., more than 1.2 million trainable parameters in [16]) to improve the representation capability of the network, which requires large memory to store the trained parameters and takes much computational time for inference. Thus, the real-time performance of the existing DL-based IM approaches

is still problematic, making it difficult to be implemented in resource-limited automotive radar systems. This motivates us to propose a new IM method to improve network performance without increasing the number of network parameters and computational complexity.

In the t-f domain, the received beat signal may exhibit a long duration time or a wide bandwidth range, which requires an IM network with a large receptive field to extract stable local features and process it. In [11], the fully convolutional network (FCN) constructed by the convolutional layers and activation layers was proposed and has shown a good IM performance. However, the receptive field of the FCN was limited by the depth of the network. To achieve a larger receptive field, the pooling layer is generally used in classical CNN-based network architectures (e.g., VGGNet [17]). In those networks, the operation of taking the maximum value or the average value of the selected area was used for pooling, which could degrade the accuracy of recovered signals (i.e., the signal distortion). The signal distortion would affect target detection or direction of arrival (DOA) estimation in the following processing. On the other hand, the convolution operation can be modified as the deformable convolution [18] to be more flexible or the dilated convolution [19] to achieve a larger receptive field. The dilated convolution was proposed by inserting the zeros between the consecutive coefficients of a convolution kernel, and thus enlarging its receptive field without increasing the number of parameters. If the dilated rates are chosen properly layer by layer, the dilated convolution neural network can extract features by covering continuous receptive field but avoid the signal distortion caused by the pooling operation. Thus, the dilated convolution is a better choice to build the interference mitigation network with a large receptive field than the traditional convolution followed by pooling layers.

Moreover, most existing DL-based IM approaches utilized clean radar signals as labels to realize supervised learning, and only the differences between clean signals and the corresponding recovered signals are evaluated with the mean square error (MSE) as a metric. However, interferences of FMCW radars have certain characteristics [11] and can be synthesized based on an analytical interference signal model. The prior feature (i.e., the generated interference signal) has not yet been exploited for training in the existing approaches.

Different from the supervised learning, contrastive learning (CL) provides an alternative to utilize the prior feature of interferences. In CL, the input samples consist of an input sample, "positive" sample, and "negative" sample. The "positive" sample means a sample with similar characteristics to the input sample, which usually belongs to the same category. The "negative" sample is a sample of a different category than the input sample. Then, the CL compared those in a feature space. Trained with the contrastive loss, the optimizer target to pull together the input sample and the "positive" sample, and pull apart the input sample from the "negative" sample in the feature space. Thus, the neural network can extract the feature of data without labeling.

The traditional CL-based applications focus on classification, where the "positive" sample is the input sample after
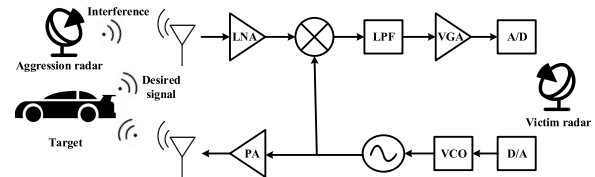

Fig. 1.   The block diagram of an interfered automotive FMCW radar system.

data enhancement (e.g., rotation), and the "negative" samples are samples other than the input sample in a batch. Another network specifically designed for the downstream task is used to output the results. In recent years, contrastive learning has been extended to address regression problems, where the loss function is composed of the supervised loss and the contrastive regularization [20], [21]. For image dehazing [20], the hazy image and the corresponding clear image are used as the "negative" sample and the "positive" image, respectively, for contrast learning. Through the contrastive regularization, it achieves a better performance than that trained with a single supervised loss. However, the hazy images used as the "negative" samples contain the information of the desired image, which would cause the network to learn and abandon certain features of desired images and deteriorate the quality of recovered image. Moreover, the supervised loss includes the penalty of the difference between the output and the label, which makes the contrastive regularization partly redundant. Meanwhile, it is hard to choose a suitable hyper-parameter to make a trade-off between the two terms.

In this paper, we presented a DL-based interference mitigation approach for FMCW radars in the t-f domain. Unlike existing works that use a more complex network to improve the representation ability, in this paper, our contributions to improve IM performance without increasing the complexity of the network are two-fold: (1) the dilated convolution with a proper dilated rate is selected to build the network to realize a larger receptive field, with which the IM network is then trained by the contrastive learning; (2) a new contrastive loss function is proposed for removing the additional feature extraction network and hyper-parameter, which can achieve a better and more stable performance than that in [20]. Thanks to the introduced prior feature of interferences through contrastive learning, the target's beat signal can be better separated from the interferences and noise.

The remainder of this paper is organized as follows. Section II briefly presents the signal model of interfered FMCW radars and the signal processing flow of our proposed IM approach. Section III elaborates the built IM network based on the dilated convolution and the proposed optimal setting of the dilated rate in each convolutional layer in the network, along with the contrastive loss function and contrastive strategy. In Sections IV, the numerical simulation setting for data synthesis is described, and the experimental results of the proposed method on the simulated data are presented. Section V shows the comparison results of the proposed method with state-of-the-art IM approaches, and the superiority of the proposed method was further demonstrated through a qualitative IM result in the complex traffic scenario. After that, the generalization performance of the proposed
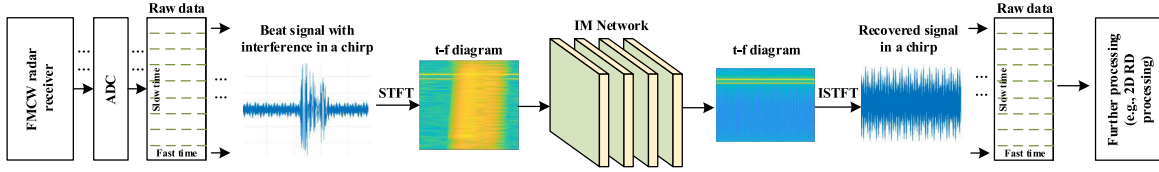
Fig. 2. Signal processing flow of our proposed approach for automotive radar interference mitigation (IM).

approach is demonstrated through the measured data in Sections VI. Finally, conclusions are drawn in Section VII.

II. RADAR SIGNAL MODEL

The block diagram of an interfered FMCW radar system is shown in Fig. 1. FMCW radar works by transmitting continuous waves, whose operating frequency increases or decreases linearly in a period during the measurement. The transmitted signal impinges upon targets and then scattered signals are reflected back to the receiver. At the same time, transmitted signals from surrounding aggression radars may also be received at the receiver of the victim radar.

After dechirping and low-pass filtering, the acquired beat signal affected by interferences can be expressed as [22]:

$$\mathbf{y}(t) = \sum_{k=1}^{N} \sigma_k \exp\left[j2\pi \left(-f_c \tau_k - K\tau_k t + \frac{1}{2} K \tau_k^2 \right) \right]$$
$$+ \mathcal{F}_{lp}\left[p^*(t) \sum_{m=1}^{M} f_m(t) \right] + \mathbf{n}(t) \qquad (1)$$

where $j = \sqrt{-1}$ is the unit of imaginary number. N (or M) is the number of targets (or interferences). δ_k means the amplitude of the received beat signal of kth target. t is used as a variable to represent fast time and $0 < t < T_{\text{sw}}$ with sweep duration T_{sw}. f_c is the center frequency, K is the chirp rate of the FMCW waveform $p(t)$ of the victim radar, and τ_k is the time delay of the scattered signal from the kth target relative to the transmitted one. $p^*(t)$ is the complex conjugate of $p(t)$ and used for dechirping, $f_m(t)$ denotes the mth interference and $\mathcal{F}_{lp}$ is the low-pass filtering (LPF) operator. $\mathbf{n}(t)$ represents thermal noise and measurement errors.

To suppress the interferences and noise in the received beat signal, the signal processing flow of our proposed IM method is shown as Fig. 2. After dechirping and low-pass filtering, the Short-time Fourier Transform (STFT) algorithm is used to transfer the received beat signal in each single chirp into a corresponding t-f diagram. The obtained t-f diagram of the interfered beat signal is fed into the IM network. After being processed, the Inverse Short-time Fourier Transform (ISTFT) can be applied to the recovered spectrogram, and the interference-suppressed beat signal in each chirp is obtained. Then, the 2D Range-Doppler processing can be done with the recovered signal in a frame, obtaining the interference-suppressed RD map for target parameter estimation.

III. METHODOLOGY

In this section, the dilated convolution is reviewed and used to build the interference mitigation network with a larger receptive field, and the selection of the dilated rate in each convolutional layer in the proposed network is discussed in
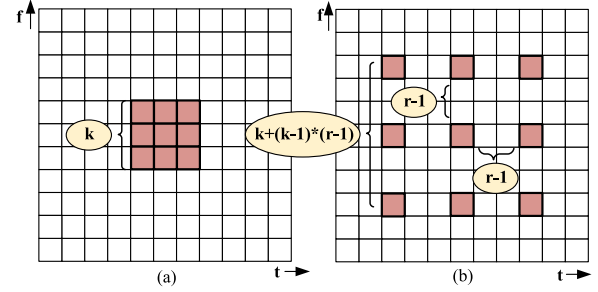

Fig. 3. Comparison of the traditional convolution operation with the dilated convolution: (a) traditional convolution (i.e., dilated convolution (r=1)), (b) dilated convolution whose dilated rate is r. The red area marks the receptive field of one pixel in the convolutional layer.

Section III-A. After that, the contrastive loss function and contrastive strategy are demonstrated in Section III-B, followed by a detailed description of the training procedure.

A. Dilated Convolution

In traditional convolutional neural networks, convolutional kernels can only extract the features of the spectrum elements adjacent to the center of the kernel, i.e., within the receptive field of the kernel size (see Fig. 3 (a)). Although it is possible to enlarge the receptive field by increasing the depth of the convolutional layers, each additional layer can only increase the receptive field by $k - 1$ (if the kernel size is $k \times k$). So, increasing the receptive field of the network based on the traditional convolutional kernel is inefficient.

In the classical architectures of CNNs (e.g., ResNet [23]), the pooling layer or a stride greater than 1 in the convolutional layer is used to reduce the size of feature maps and enlarge the receptive field. However, the signal may be distorted due to pooling by taking the average or the maximum value in the selected area. A possible better solution is to utilize the dilated convolution [19], where a dilation rate of r is set. Different from the traditional convolution, $r - 1$ zeros are inserted between the consecutive coefficients of a convolution kernel, as shown in Fig. 3 (b). Through this modification, the convolutional kernel can extract features of the spectrum elements over a greater field of view, and the receptive field of a convolutional layer with a kernel of size $k \times k$ can be increased from $k \times k$ to $[(k - 1)(r - 1) + k]^2$. Moreover, if the dilated rates are set to a group of proper values, all the spectrum elements can be processed (i.e., perceived) without signal loss. Hence, useful local and global features would be extracted by the kernel, resulting in better interference mitigation performances of the network. It should be noted that the inserted zeros are only used to explain the increasement of the receptive field of the dilated convolution kernel, but do not change the number of its effective kernel coefficients and do not need extra computation. Therefore, the dilated convolution
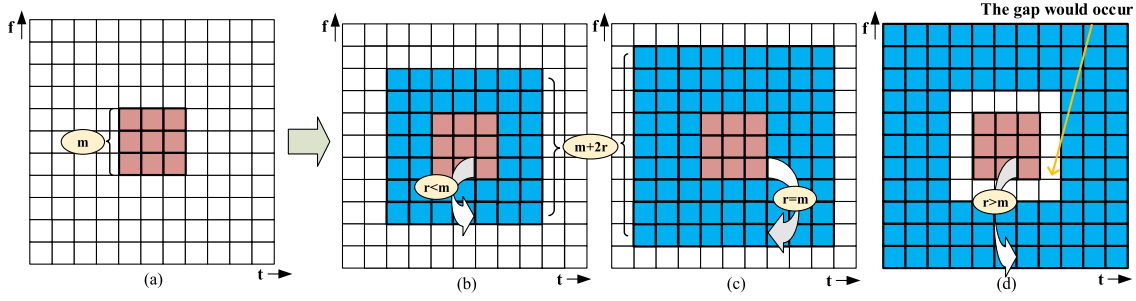
Fig. 4. Discussion about the setting of the dilated rate in each convolutional layer in the Dilated Convolution Block: (a) is the previous layer. (b)(c)(d) is the current layer using the dilated convolution with a different dilated rate: (b) $r < k$, (c) $r = k$, (d) $r > k$. The red area marks the receptive field of one pixel in the previous layer, and the total area of blue and red marks the receptive field of one pixel in the current layer.

does not introduce additional complexity in terms of both training parameters and computational load.

The setting of the dilated rate (i.e., r) in each convolutional layer would determine the receptive field of the network. The purpose of choosing an appropriate value of dilated rate is to enlarge the receptive field as much as possible and keep all the elements in the matrix processed. If the dilation rates in all the convolutional layers are set to a fixed value greater than 2, there would be a gridding effect [24], and a large portion of information would be lost. In theory, a) only when the greatest common divisor of the dilated rates is no more than 1, all the elements in the matrices can be processed without information loss. Since the convolution operation satisfies the commutative law, b) the case where the dilated rate increases layer by layer was considered for discussion. Based on the condition a) and the assumption b), a Dilated Convolution Block can be constructed following the two guidelines below.

1) The dilated rate in the first convolutional layer of the Dilated Convolution Block has to be set to one;
2) From the second layer and on, if the receptive field of a previous layer is $m \times m$, then the current layer takes a dilated rate $r \leq m$, which leads to a receptive field $m + 2r$ for each sample (pixel) in the output of the current layer. However, if a dilated rate $r > m$ is taken in the current layer, there would be a gap (i.e., white cells in Fig. 4 (d)).

According to the previous experiments in [11], the FCN with 11-layer convolution showed the best performance in FMCW radar interference mitigation. Thus, based on the FCN, the DFCN(r_1, r_2, r_3) was constructed by including the dilated convolution, where r_i represents the dilated rate in i^{th} convolutional layer of the Dilated Convolution Block. The network architecture is shown in Fig. 5, where Conv(k, k, n, r) represents the convolution layer with a kernel size of $k \times k$ and a dilated rate of r, and the number of filters is n. The kernel size of 3×3 (i.e., $k = 3$) is used in this paper. As described before, we suggest that the dilated rate of each convolutional layer in the Dilated Convolution Block is set to 1, 3, and 9, which can guarantee no "hole" and achieve a larger receptive field compared to the setting of 1, 2, and 5 [25], [26], [27], [28], [29] (the common choices in dilated convolution-based networks). In the following experimental part, due to the used dilated rates (r_i) are all single digits, the DFCN(r_1, r_2, r_3) will be denoted as DFCN($r_1 r_2 r_3$) for simplicity.
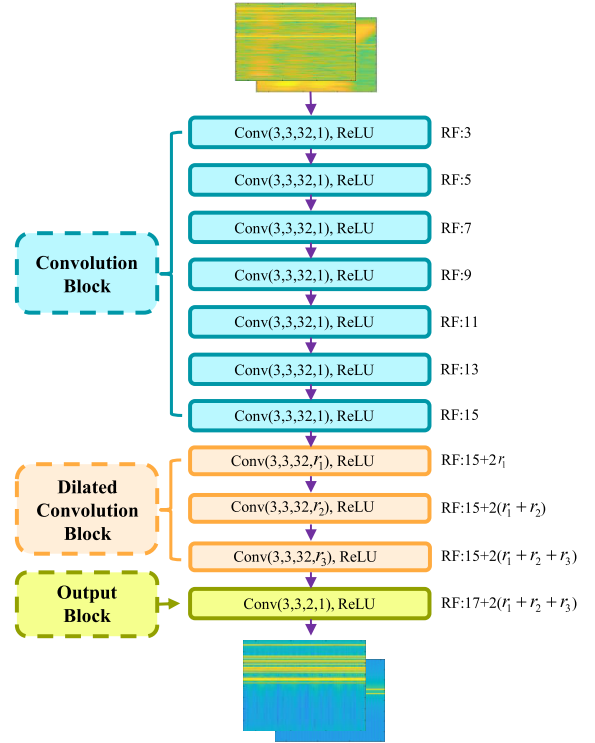
Fig. 5. He network architecture of the proposed dilated convolution based fully convolutional network (DFCN($r_1 r_2 r_3$)).

B. Contrastive Learning

1) Contrastive Loss Function: in [20], the dehazing network (similar to the IM network) was first used to output the recovered image. Then, an additional feature extraction network was implemented to extract the feature of the output image (i.e., recovered image), the "positive" image (i.e., clean image), and the "negative" image (i.e., hazy image), which would be compared in the representation space. The loss function is composed of a supervised loss function and a contrastive regularization. The supervised loss function is used to calculate the difference of the output and the label while the contrastive regularization can restrain the distance of the three feature vectors. The complete loss function is expressed as follows:

$$LA = \|\mathbf{Y} - \mathbf{Y_P}\|_F^2 + \lambda \cdot \frac{\mathcal{D}(\mathcal{G}(\mathbf{Y_P}), \mathcal{G}(\phi(\mathbf{Y})))}{\mathcal{D}(\mathcal{G}(\mathbf{Y_N}), \mathcal{G}(\phi(\mathbf{Y})))} \qquad (2)$$

where $\mathbf{Y_P}$, $\mathbf{Y_N}$, and $\mathbf{Y}$ denote the "positive", "negative", input samples, respectively. $\phi(\mathbf{Y})$ denotes the output of the dehazing

Fig. 6.   The architecture of the proposed interference mitigation network with contrastive learning.

network, $\mathcal{G}(\mathbf{x})$ is the feature extraction network, and $\mathcal{D}(\mathbf{a}, \mathbf{b})$ calculate the euclidean distance between the feature vector $\mathbf{a}$ and $\mathbf{b}$. Thus, the neural network $\mathcal{G}$ would determine the features of sa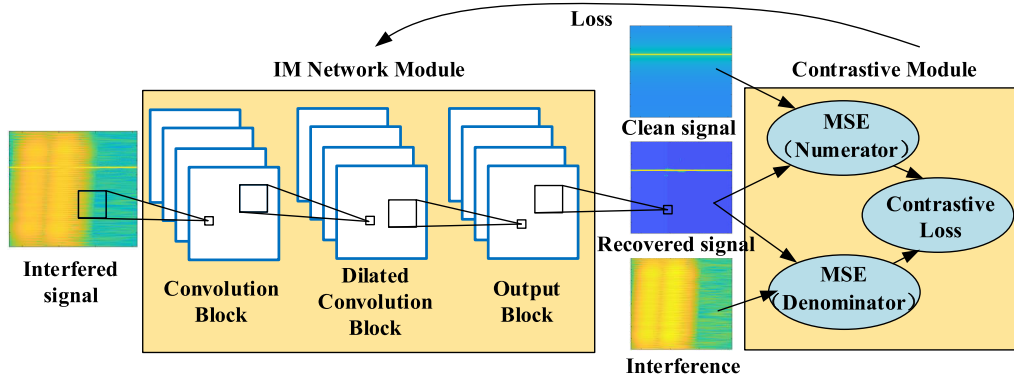mples for comparison and affect the stability of the performance during training. Moreover, the hyper-parameter λ of the regularization term is also hard to choose.

Due to these shortcomings, we reconsidered how the contrastive loss function should be constructed. Firstly, the feature extraction network was used for high-level feature extraction of natural images and following comparison. However, the t-f diagram of radar signals is relatively simple to understand than the natural images, which follows that the t-f diagrams of beat signals can be directly compared without the feature extraction network. Secondly, since the recovered signal and the label have been compared using the euclidean distance in the contrastive regularization, the supervised loss function and the hyper-parameter were not needed. Thus, we proposed a new and more simple contrastive loss function:

$$LB = \frac{\mathcal{D}(\mathbf{Y_P}, \phi(\mathbf{Y}))}{\mathcal{D}(\mathbf{Y_N}, \phi(\mathbf{Y}))} \tag{3}$$

Instead of using an additional network for feature extraction, the output of the IM network is directly compared with the "positive" sample and the "negative" sample. Meanwhile, the supervised loss function can be removed and the hyper-parameter is not needed, which enables a more stable training performance.

Compared with only using MSE as the loss function in the existing DL-based IM approaches, the proposed contrastive loss can not only calculate the distance between the "positive" sample (i.e., the label) $\mathbf{Y_P}$ and $\phi(\mathbf{Y})$, but also the distance between the "negative" sample $\mathbf{Y_N}$ and $\phi(\mathbf{Y})$ (see Fig. 6). The closer between the recovered spectrogram and the "positive" sample and the larger between the recovered spectrogram and the "negative" sample, the smaller the loss function would be. Due to the additional comparison with the "negative" sample, the extracted features of input samples can be further away from the interferences. The residual interference can be further removed, and the network can achieve a better IM performance than that trained with only MSE.

2) Contrastive Strategy: in [30], the authors have researched the key components of the success of the constrastive learning. One of them is to construct the "positive"

sample and the "negative" sample to be compared. For the IM task, the clean reference signal is the desired signal and utilized as the "positive" sample. Considering the interfered signal includes the desired beat signal, the recovered signal may be away from the target's beat signal if we push away the recovered signal from the interfered signal. In particular, in the high Signal-to-Interference-plus-Noise-Ratio (SINR) cases, there are weak interference and noise in the interfered signal, and then pushing the recovered signal away from the interfered signal is not crucial. To avoid this, we only use the generated interferences and noise as the "negative" sample. By comparing the recovered signal with various interferences, the network can pay attention to the residual interference components in the recovered signal and remove them. In the following experiments, the DFCN($r_1\,r_2\,r_3$)-CL(LA/interfered) means the DFCN trained with CL using the LA loss and the interfered signal as the "negative" sample. DFCN($r_1\,r_2\,r_3$)-CL(LB/interfered) means the DFCN trained with CL using the LB loss and the interfered signal as the "negative" sample. DFCN($r_1\,r_2\,r_3$)-CL(LB/interference) means the DFCN trained with CL using the LB loss and the interference and noise as the "negative" sample.

C. Training Procedure

For the network training, the t-f diagrams of received radar signals are preprocessed (i.e., data split and normalization) as described in [11]. The Adam optimizer was adopted for training with a fixed learning rate of 0.001 and 12 input samples per batch. The training procedure was ended after 100 epochs when good convergence was observed. Moreover, the code was implemented using Keras and TensorFlow tools, and all the models were trained on a single NVIDIA GeForce RTX 3090 graphics processing unit (GPU). Each network was trained five times and the results were averaged to reduce the variance.

IV. SIMULATION RESULTS

In this section, the setups of numerical simulations for data synthesis are first introduced, and then the evaluation metrics are given. After that, the FCN was used for IM in the t-f domain to show its effectiveness. Then, the DFCNs were implemented to demonstrate the better performance caused by

TABLE I
PARAMETERS OF THE VICTIM RADAR

Parameter	Value	Parameter	Value
Center frequency	76.5 GHz	Moving speed	0 m/s
Duration of a sweep T_{sw}	25 μs	Window type	Hamming
Bandwidth	0.4 GHz	Window length	256
Chirp rate K	16 MHz/μs	Overlap length	255
Sampling frequency	40 MHz	FFT points	256
Maximum detection distance	168.75 m		

TABLE II
PARAMETERS OF THE TARGETS AND INTERFERENCES

Parameter of Targets	Value	Parameter of Interferences	Value
Number	$\mathcal{U}\{0, 20\}$	Number	$\mathcal{U}\{1, 20\}$
Distance	$\mathcal{U}(8, 168.75)$ m	Amplitude[2]	$\mathcal{U}(0, 3)$
Amplitude[2]	$\mathcal{U}(0, 3)$	Center frequency	76.5 GHz
Phase	$\mathcal{U}(0, 2\pi)$	Chirp rate	$\mathcal{U}(-2K, 2K)$
Velocity	$\mathcal{U}(0, 30)$ m/s	Duration	$\mathcal{U}(0, T_{sw})$
		Delay time	$\mathcal{U}\left(\frac{-T_{sw}}{2}, \frac{T_{sw}}{2}\right)$

[1] $\mathcal{U}$ means the uniform distribution.
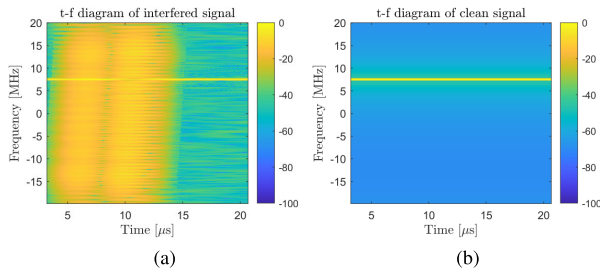[2] Note that the amplitude is a relative value instead of a true value.

Fig. 7. t-f diagrams of (a) a simulated beat signal contaminated by interferences and (b) its label.

dilated convolution and the superiority of the proposed optimal setting of dilated rates. Finally, the contrastive learning was used for training, verifying the advantages of the proposed contrastive loss function and contrastive strategy.

A. Simulated Datasets

Due to the difficulties of collecting labeled experimental data with radars, we use the simulated radar signals for training, and both the simulated and measured radar signals for testing.

The similar approach as in [11] is used for data synthesis for automotive radars. To simulate the automotive FMCW radar (i.e., millimeter wave radar), the parameter setting of a victim radar is illustrated in Table I. Meanwhile, the detailed intervals of the values of the parameters of targets and interfering signals are illustrated in Table II, where K and T_{sw} refer to the chirp rate and sweep duration of the victim radar in Table I. After synthesizing the beat signal in the time domain, the STFT was used to transform it into a t-f diagram. The parameters of the STFT algorithm are listed in Table I. In practice, a large dataset is generally hard to be acquired for the deep-learning-based radar signal processing methods. Considering this reality and, meanwhile, evaluating the generalization capability of the related DL-based methods,

a small set is used for training while a relatively large dataset is used for test. Specifically, 540 and 180-chirp radar signals (with a ratio of 3:1) are generated as training and validation data sets, respectively. For adequate quantitative evaluation, a relatively large number of signal samples (1440 chirps) are generated for quantitative testing. As an example, the t-f diagrams of a simulated interfered radar signal and its clean counterpart are shown in Fig. 7.

B. Evalution Metrics

The interference mitigation operation is one of the steps in the radar signal processing flow. The quality of the recovered signal after IM is critical for performance evaluation of the IM approach and further processing including target detection, DOA estimation, etc. Hence, the *SINR* and the *correlation coefficient* ρ of the recovered signal relative to its clean reference are used to evaluate its accuracy, and their definition can be seen in [11]. The SINR can show how much residual interference exists in the recovered signal and the signal distortion. Meanwhile, the correlation coefficient can measure the similarity of the recovered signal with its clean reference.

Besides, the target detection performance and the parameter estimation accuracy of the recovered signals were evaluated after further signal processing operations. Without loss of generality, the recovered signals after IM were further processed with the Fast Fourier Transform (FFT) to obtain the range profiles of targets. Note that the signal samples were padded with seven times more zeros before the aforementioned FFT operation, which would result in smooth range profiles and facilitate the estimation of targets' distances later. Then, the cell-averaging constant false alarm rate (CA-CFAR) detector was used to detect the targets on the range profiles. Specifically, the numbers of training cells and guarding cells were set to 150 and 14, respectively. The false alarm rate was fixed to 1×10^{-4}. After that, the distances of detected targets were estimated by extracting the positions of the related peaks in the range profiles.

For quantitative evaluation, the detection rate and the range estimation error are used as metrics to indicate the target detection performance and the accuracy of the estimated ranges of detected targets with the recovered signals after IM, respectively. The range estimation error (REE) is defined as:

$$\text{REE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}|\tilde{R}_k - R_k|^2} \qquad (4)$$

where N is the number of targets. R_k denotes the range of the kth target, and $\tilde{R}_k$ is its estimated counterpart.

C. Effects of Dilated Convolution

Here, the effectiveness of the dilated convolution and the performance improvement introduced by the proposed optimal setting of dilated rates are demonstrated.

Firstly, we used the normal FCN without dilated convolution to process the interference-contaminated beat signals. The performance results are shown in Fig. 8. The proposed FCN can suppress the interference components and noise in
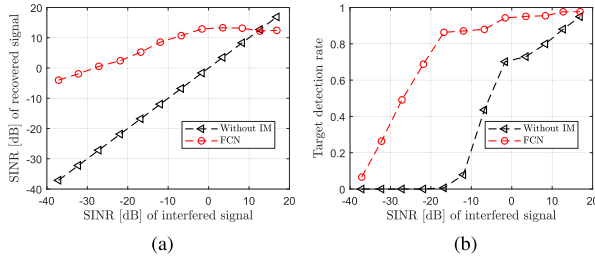
Fig. 8. Interference mitigation performance of the proposed FCN using the normal convolution: (a) the SINR of recovered signals, (b) the target detection rate after IM.

the t-f diagram of received beat signals. It is obvious that the recovered signals after IM show a higher SINR than that without IM, and the target detection rate improves. The average SINR of the interfered signal is -9.5056 dB. After IM, it improves to 7.1266 dB. As the SINR of the interfered beat signals is between -40 dB and -20 dB, the target detection rate of the beat signals without IM is zero. After using the FCN for IM, more targets can be detected, and the average target detection rate increases to 37.69% in the low SINR cases.

Then, the dilated convolution was implemented to build the IM network (i.e., DFCN). We set the dilated rate of each convolution layer in the Dilated Convolution Block according to three different strategies: (a) 1, 2, and 5, (b) 1, 2, and 7, and (c) 1, 3, and 9. Specifically, (a) is the commonly used in dilated convolution-based network architectures, and (b) is the improved version based on (a) through our analysis. (c) is proposed in Section III-A and the optimal setting of dilated rates. The performance comparison of the networks built based on the three strategies are shown in Fig. 9 and Table V. Both the evaluation results and the receptive field of each network are listed. It is obvious that using the dilated convolution can enlarge the receptive field and improve the IM performance without increasing the number of parameters of the network. Using the dilated convolution with the setting of (a), compared to the FCN using the normal convolution, the average SINR increases from 7.1266 dB to 7.6382 dB, and the target detection rate can be improved from 74.36% to 78.86%.

Moreover, among the IM networks with the three settings of dilated convolutions, the settings (b) and (c) can achieve a larger receptive field than the setting (a). Both the achieved SINR and the target detection rate improve with the increase of the receptive field. Specifically, the IM network with the dilated convolution of the setting (c) has the largest receptive field (43) compared to the settings (a) and (b) (with receptive fields of 33 and 37, respectively); thus, it achieves the best SINR of the recovered signals (i.e., 7.8791 dB) and the detection rate of targets (i.e., 79.03%). By contrast, the IM networks with the dilated convolutions of settings (a) and (b) result in similar SINRs of recovered signals (7.6392 dB and 7.6429 dB) and the detection rate of targets (78.86% and 78.29%) due to their comparable receptive fields (33 and 37). The DFCNs with a larger receptive field can remove more interferences than the normal FCN, and the noise floor of the recovered signal processed by the DFCNs decreases even more. A larger difference of amplitude between targets' beat signals and the noise floor (i.e., a larger SINR of the recovered signal after IM)

achieves a significantly higher detection rate of the DFCNs. Besides, the range estimation error of the IM networks is compared. The signals recovered with the DFCNs and the normal FCN lead to almost equally small range estimation error (i.e., about 0.06m).

Thus, the dilated convolution is an effective method to improve the IM performance of the neural networks by increasing the receptive field. Although the dilated convolution has been utilized in previous research, most of the networks chose to use the setting of dilated rates (a) (i.e., 1, 2, and 5), which is not optimal as revealed by our experiments. To the best of our knowledge, it is the first time to implement the dilated convolution in the FMCW radar interference mitigation task, and verify the strategy of how to choose the dilated rate in each convolutional layer, which can provide a good guideline for further research of dilated convolution-based applications.

D. Effects of Contrastive Learning

In this part, the interference signal was added as the input of the network, and we used the contrastive learning for training. The performance of the network trained with contrastive learning and the one trained with only MSE are compared. Meanwhile, the contrastive loss function and strategy in [20] were also implemented for comparison to verify the superiority of the proposed contrastive loss and the contrastive strategy.

1) Contrastive Loss Function: firstly, the contrastive loss proposed in [20] (i.e., LA) was used as the loss function. In [20], the feature extraction network was utilized to extract the high-level feature of the output sample of the IM network, the "positive" sample and the "negative" sample. Then, they were compared in a specific feature space using the Euclidean distance. As a result, the feature extraction network would determine the features for comparison. If an improper network was used, the feature comparison operation would be invalid, resulting in no improvement or worse performance. Besides, the additional feature extraction network was trained together with the interference mitigation network, which would affect the convergence and performance of the IM network. Moreover, the LA was composed of the MSE and a contrastive regularization. A hyper-parameter was used to adjust the ratio between them. The hyper-parameter was hard to choose for training, and this causes more difficulties during the network's training.

To verify the performance of the LA loss, in our experiment, several classical CNN-based networks including the VGGNet [17] and ResNet [23], and the lightweight networks including the EffectiveNet [31] and MobileNet [32] were used as the feature extraction network because of their good performance in past tasks. The hyper-parameter was selected from {0.1, 1, 2, 5, 10}. The interfered signal was used as the "negative" sample, and the clean reference signal as the "positive" sample. The input sample was compared with only a "positive" sample and a "negative" sample.

The performance of the networks trained using contrastive learning with LA is shown in Tables III and IV. When the hyper-parameter $\lambda = 1$, the results in Table III show that the recovered signal with a higher SINR is obtained by using the ResNet or MobileNet as the feature extraction network.
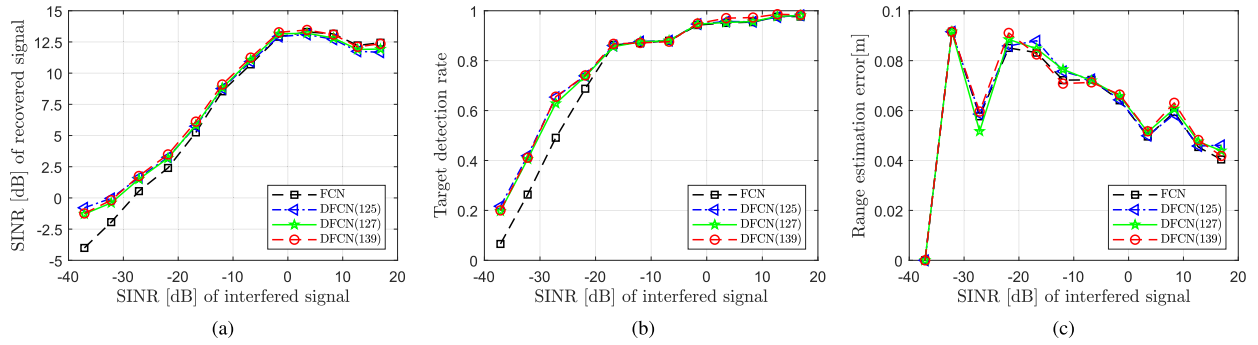
Fig. 9.   The interference mitigation performance comparison of the DFCNs with different settings of dilated rates: (a) the SINR of the recovered signals, (b) target detection rate after IM, (c) range estimation error after IM.

TABLE III

RESULTS OF THE FCNs TRAINED WITH LA USING DIFFERENT FEATURE EXTRACTION NETWORKS

Method	Hyper-parameter	Feature Extraction Network	SINR (dB)	$\mid \rho \mid$	$\angle\rho$ [rad]	detection rate	REE[m]
FCN-CL(LA/interfered)	**1**	-	7.1266	0.7244	0.2319	0.7436	0.0601
		VGGNet	6.6080	0.7128	0.2278	0.7263	0.0607
		ResNet	7.1831	0.7251	0.1907	0.7509	0.0599
		EfficientNet	5.6249	0.6954	0.2598	0.7066	0.0592
		MobileNet	**7.2554**	0.7266	0.2217	0.7523	0.0611

TABLE IV

RESULTS OF THE FCNs TRAINED WITH LA USING DIFFERENT HYPER-PARAMETERS

Method	Feature Extraction Network	Hyper-parameter	SINR (dB)	$\mid \rho \mid$	$\angle\rho$ [rad]	detection rate	REE[m]
FCN-CL(LA/interfered)	**MobileNet**	0.1	7.1185	0.7230	0.2386	0.7447	0.0585
		1	**7.2554**	0.7266	0.2217	0.7523	0.0611
		2	6.9487	0.7146	0.2269	0.7288	0.0589
		5	6.9785	0.7243	0.2343	0.7439	0.0614
		10	7.0064	0.7207	0.2473	0.7230	0.0602

TABLE V

RESULTS OF THE ABLATION EXPERIMENTS

Method	Dimensions of receptive field	SINR (dB)	$\mid \rho \mid$	$\angle\rho$ [rad]	detection rate	REE[m]
FCN	23	7.1266	0.7244	0.2319	0.7436	0.0601
FCN-CL(LA/interfered)	23	7.2554	0.7266	0.2217	0.7523	0.0611
FCN-CL(LB/interfered)	23	7.2476	0.7264	0.2395	0.7519	0.0612
FCN-CL(LB/interference)	23	7.3184	0.7279	0.2329	0.7528	0.0627
DFCN(125)	33	7.6382	0.7559	0.2178	0.7886	0.0614
DFCN(125)-CL(LA/interfered)	33	7.8810	0.7590	0.2248	0.7939	0.0620
DFCN(125)-CL(LB/interfered)	33	7.8839	0.7521	0.2358	0.7863	0.0619
DFCN(125)-CL(LB/interference)	33	7.8856	0.7615	0.2246	0.7977	0.0619
DFCN(127)	37	7.6429	0.7526	0.2240	0.7829	0.0612
DFCN(127)-CL(LA/interfered)	37	7.8223	0.7533	0.2246	0.7847	0.0617
DFCN(127)-CL(LB/interfered)	37	8.1077	0.7630	0.2266	0.7988	0.0594
DFCN(127)-CL(LB/interference)	37	8.0888	0.7619	0.2289	0.7992	0.0602
DFCN(139)	43	7.8791	0.7575	0.2276	0.7903	0.0615
DFCN(139)-CL(LA/interfered)	43	7.9063	0.7573	0.2275	0.7896	0.0619
DFCN(139)-CL(LB/interfered)	43	8.0009	0.7563	0.2320	0.7915	0.0610
DFCN(139)-CL(LB/interference)	**43**	**8.1564**	0.7619	0.2300	0.8008	0.0614

Among them, the MobileNet achieves the highest SINR of 7.2554 dB and target detection rate of 75.23%. Similarly, in Table IV, the IM results of the networks trained using LA with λ of different values are presented. Different values of λ lead to a big difference in network performance. When the hyper-parameter $\lambda = 1$, the trained network gets the best performance. So, in the following experiments, the MobileNet would be used as the feature extraction network and the
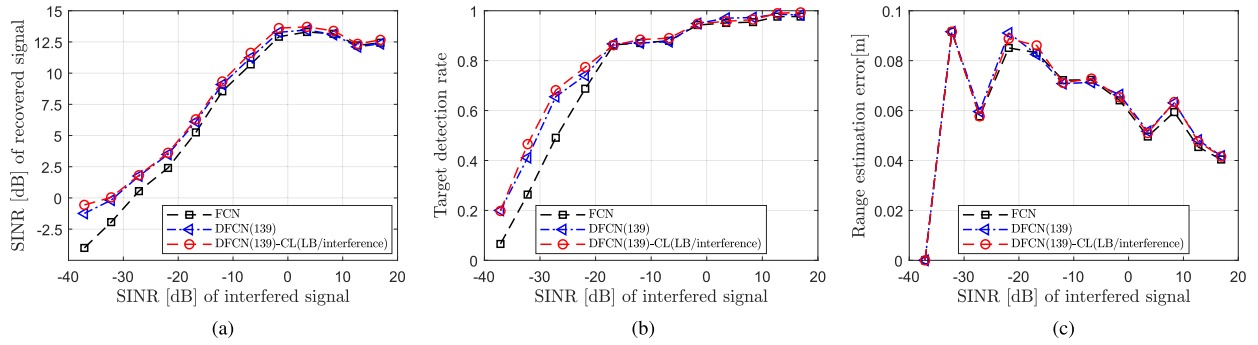
Fig. 10. The interference mitigation performance comparison of the normal FCN and the improved ones using our proposed methods: (a) the SINR of the recovered signals, (b) target detection rate after IM, (c) range estimation error after IM.

hyper-parameter $\lambda = 1$ is used for performance comparison with our proposed approach (LB loss function).

Further, the results of the networks trained with LA under different settings of dilated rates are illustrated in Table V. In the cases of different settings of dilated rates, all the networks trained using contrastive learning with the MobileNet as the feature extraction network can get the SINR improvement, which shows the effectiveness of the contrastive learning in the interference mitigation problem.

Finally, we used our proposed contrastive loss (i.e., LB) for training. As described in Section III-B, the feature extraction network and the hyper-parameter are not needed. For comparison, the interfered signal was firstly used to be the "negative" sample. The performance comparison between the proposed contrastive loss function (i.e, LB) and LA can be seen in Table V. In general, the network trained with LB can achieve a better IM performance (both the achieved SINR and target detection rate) than the one trained with LA, which is the best result obtained through parameter search. This is because that the more accurate the recovered signal, the higher its SINR, and thus the higher the detection rate of targets. Meanwhile, the networks trained using the contrastive learning (LA/LB loss) show a similar range estimation error as the networks trained using the MSE. Since there is no hyper-parameter to adjust and no feature extraction network to select, the proposed loss function can achieve a more stable experimental performance. Besides, the feature extraction network in LA would inevitably increase the computational complexity (i.e., the training time). Using LA as the loss function, it takes 1.75 hours to train the network for 100 epochs. However, only 1.46 hours are needed without the feature extraction network. Thus, it is a better solution to use our proposed contrastive loss function for training.

2) Negative Sample Selection: in [20], the authors chose to use the hazy image as the "negative" sample, which includes the desired pixels. In our experiments, the SINR of the interfered signal ranges from -40 dB to 20 dB in the dataset. In the high SINR scenarios (e.g., the SINR $>= 10$ dB), there is much weaker interference and noise in the interfered signal. If we push the recovered signal away from the "negative" sample, it may lead to a wrong convergence direction.

Instead, we use only the interference and noise as the "negative" sample to avoid the problem. Then, the performance comparison is demonstrated in Table V. The network trained using the interference signal as the "negative" sample can

TABLE VI
COMPARISON OF THE SELECTED DL-BASED IM APPROACHES

Method	Parameters	Memory (MB)	GFLOps	Average SINR (dB)
CNN	10194	0.22	1.283	2.9118
ResNet	1229442	14.6	154.920	3.7352
U-net	540154	6.41	2.076	6.4648
RV-FCN	84418	0.69	11.022	7.1266
CV-FCN	42370	0.37	11.022	6.8511
Ours	84418	0.69	11.022	8.1564

achieve a higher SINR and target detection rate than the one using the interfered signal as the "negative" sample. As the dilated rates were set to 1, 3, and 9, the SINR of the recovered signals can be improved from 8.0009 dB to 8.1564 dB, and the target detection rate increases to 80.08%, achieving the best performance in all the trained networks.

In conclusion, we used the contrastive learning for training in this part, which shows a better performance than the one trained with only MSE. Moreover, a new contrastive loss function and the choice of the "negative" sample are proposed and studied. The results show it can realize the superior performance and more stable training convergence.

To facilitate the comparison of the effect of the dilated convolution and contrastive learning, the evaluation results of the FCN, DFCN(139), and DFCN(139)-CL(LB/interference) are presented together in Fig. 10. It clearly shows the improvement of the IM networks with the dilated convolution and contrastive learning in terms of both the SINR of recovered signals and the target detection performance. Moreover, the dilated convolution can be easily integrated into the existing CNN-based networks, which means only the change of the dilated rate in the convolutional layer is enough. Meanwhile, the proposed contrastive learning does not need to modify the network architecture and introduces no additional parameters. The only expense is that the interference signal was needed during the training of contrastive learning. Thus, the proposed approach would become an easy-to-implement performance improvement method for DL-based interference mitigation.

V. COMPARISON EXPERIMENTS

The performance comparison of DL-based IM approaches and traditional signal processing methods in [11] shows that the DL-based approaches can generally achieve much better
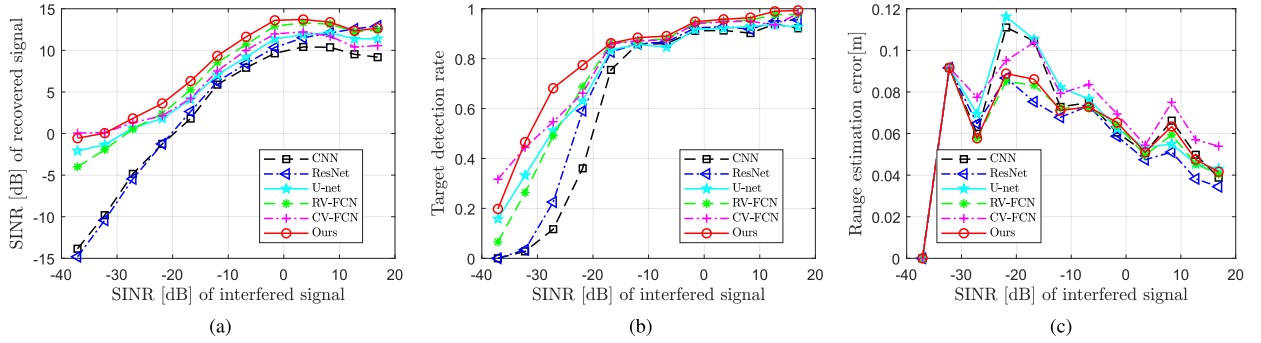
Fig. 11. The interference mitigation performance comparison of the proposed IM approach with the selected DL-based interference mitigation approaches: (a) the SINR of the recovered signals, (b) target detection rate after IM, (c) range estimation error after IM.
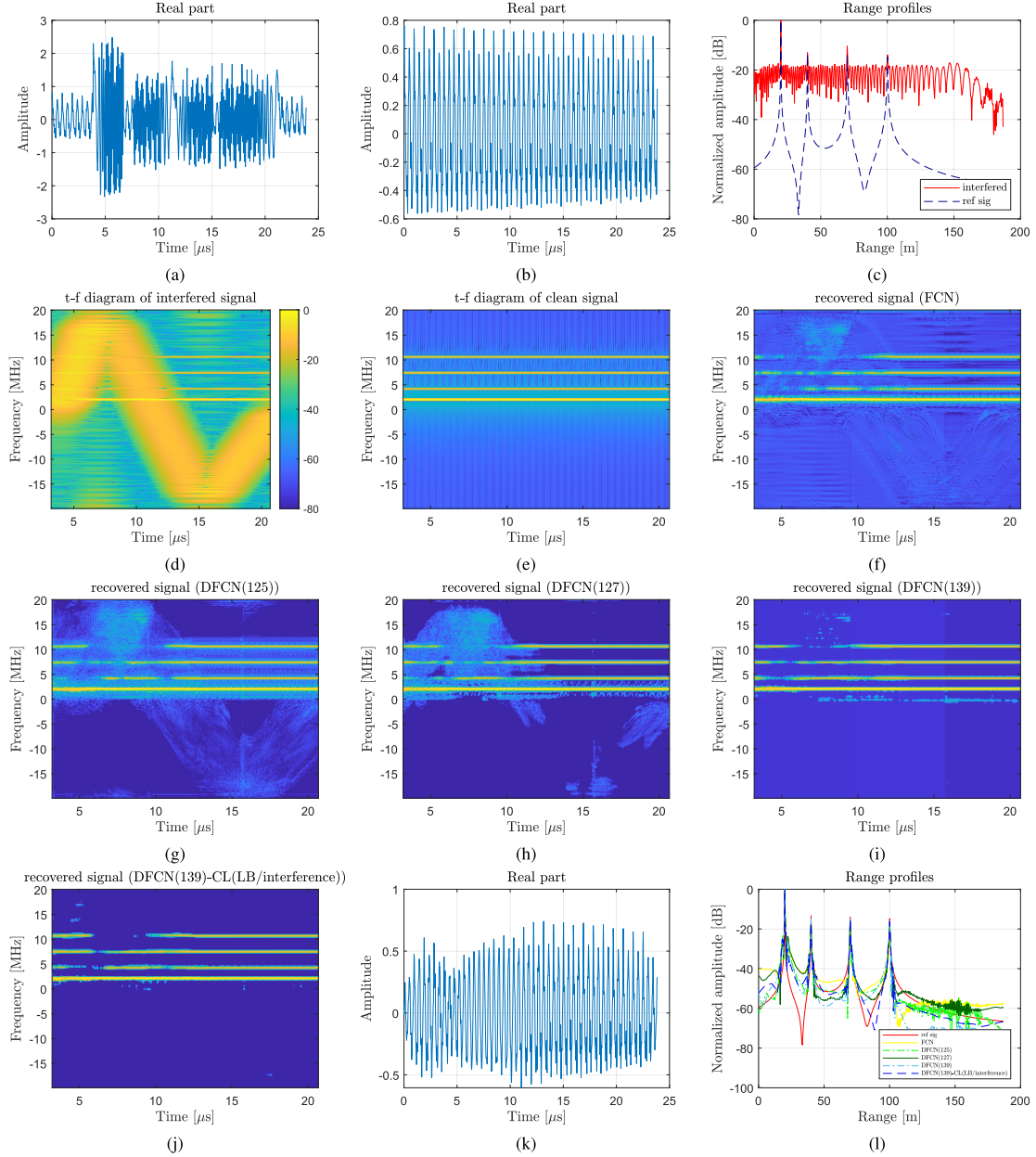

Fig. 12. Interference mitigation for the simulated radar signal whose SINR is -9.11 dB and SNR is 20 dB. (a) shows the acquired beat signal contaminated by mutual interference, (b) its range profiles, (c) its time-frequency diagram, (d) and (e) show an interference-free reference. (f) the result of IM using the FCN. (g)-(h) the results of IM using the DFCNs with different settings of dilated rates. (i) and (j) the result of IM (t-f diagram and signal waveform) using DFCN(139)-CL(LB/interference). (k) displays the corresponding range profiles obtained after IM.

results. So, here we mainly focus on the comparison of the proposed approach with the state-of-the-art DL-based IM approaches including the CNN [12], RV-FCN [11], CV-FCN [11], ResNet [16], and U-Net [14]. All the networks were
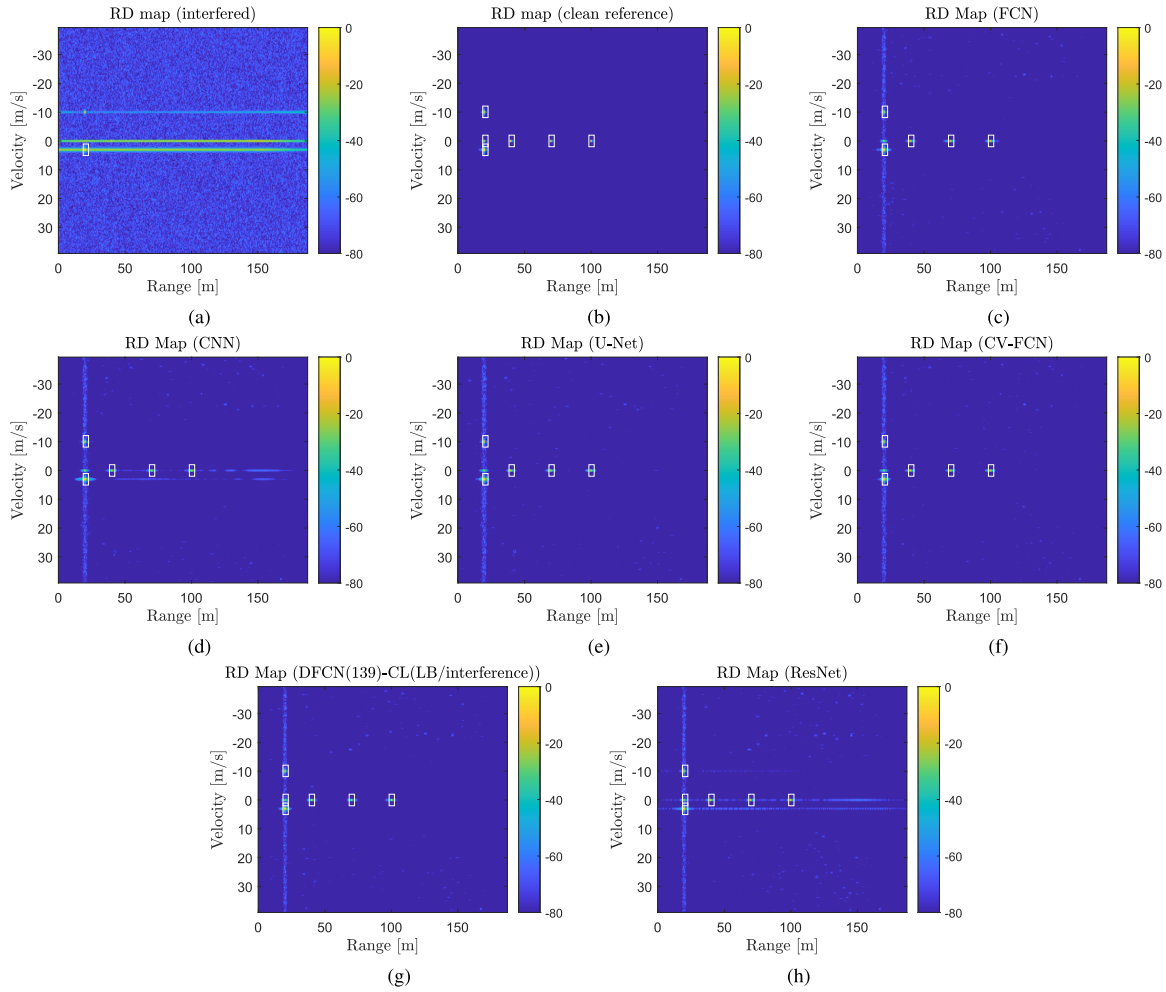
Fig. 13. The target detection results on the RD maps of (a) the interfered signal, (b) the clean reference signal, (c) the recovered signals processed by the FCN using the normal convolution, (d) the CNN-based approach, (e) the U-Net-based approach, (f) the CV-FCN-based approach, (g) the DFCN(139)-CL(LB/interference), and (h) the ResNet-based approach.

trained using only 540-chirp signal samples. The CV-FCN was trained using the prior-guided loss function [2] where the hyper-parameter was set to 256. The CNN, RV-FCN, ResNet, and U-Net were trained using MSE as the loss function. The DFCN proposed in this paper was trained using the contrastive loss function. Their performances are compared and shown in Fig. 11. The number of total parameters (i.e., scalability), memory usage, Giga Floating-point Operations (GFLOps) (i.e., computational complexity), and the achieved average SINR of these methods are listed in Table VI.

Compared to the traditional CNN and ResNet-based IM networks, the U-Net, RV-FCN, CV-FCN-based method, and our proposed method have shown better IM performance. The U-Net-based IM method is built based on the encoder-decoder architecture, and has more than 540 kilo parameters. As pooling layers are used to reduce the size of feature maps in the U-Net, its computational complexity is relatively small. However, the pooling operation may lead to signal loss, affecting the further DOA estimation, etc. Compared to the U-Net, the RV-FCN uses fewer filters in each convolutional layer; thus, it has fewer parameters. However, the RV-FCN shows a higher textitaverage SINR than the U-Net, but the SINR of the recovered signal processed by the RV-FCN is

lower than that of the U-Net in the low SINR scenarios. Meanwhile, the CV-FCN is constructed based on the complex-valued fully convolutional network. By using the prior-guided loss function for training, the CV-FCN with only 42370 parameters shows the highest SINR and target detection rate in the low SINR scenarios among all the networks. However, the IM performance of the CV-FCN depends on a hyperparameter whose optimal value is not easy to choose. By contrast, the proposed IM network with the dilated convolution can realize a larger receptive field and get better elimination of interference components with the help of contrastive learning. As a result, the proposed approach achieves the highest average SINR and target detection rate. Regarding the range estimation error, the ResNet-based IM approach achieves the best performance, followed by our proposed method. The ResNet eliminates interferences and noise by learning their features rather than directly learning the features of the desired beat signals. Although it does not remove interferences and noise very well (i.e., lower SINR and target detection rate), the ResNet causes less distortion to the desired beat signal, thus resulting in a lower range estimation error of targets, especially in the high SINR scenarios. In summary, the comparisons have shown that our proposed IM approach achieves the best performance.
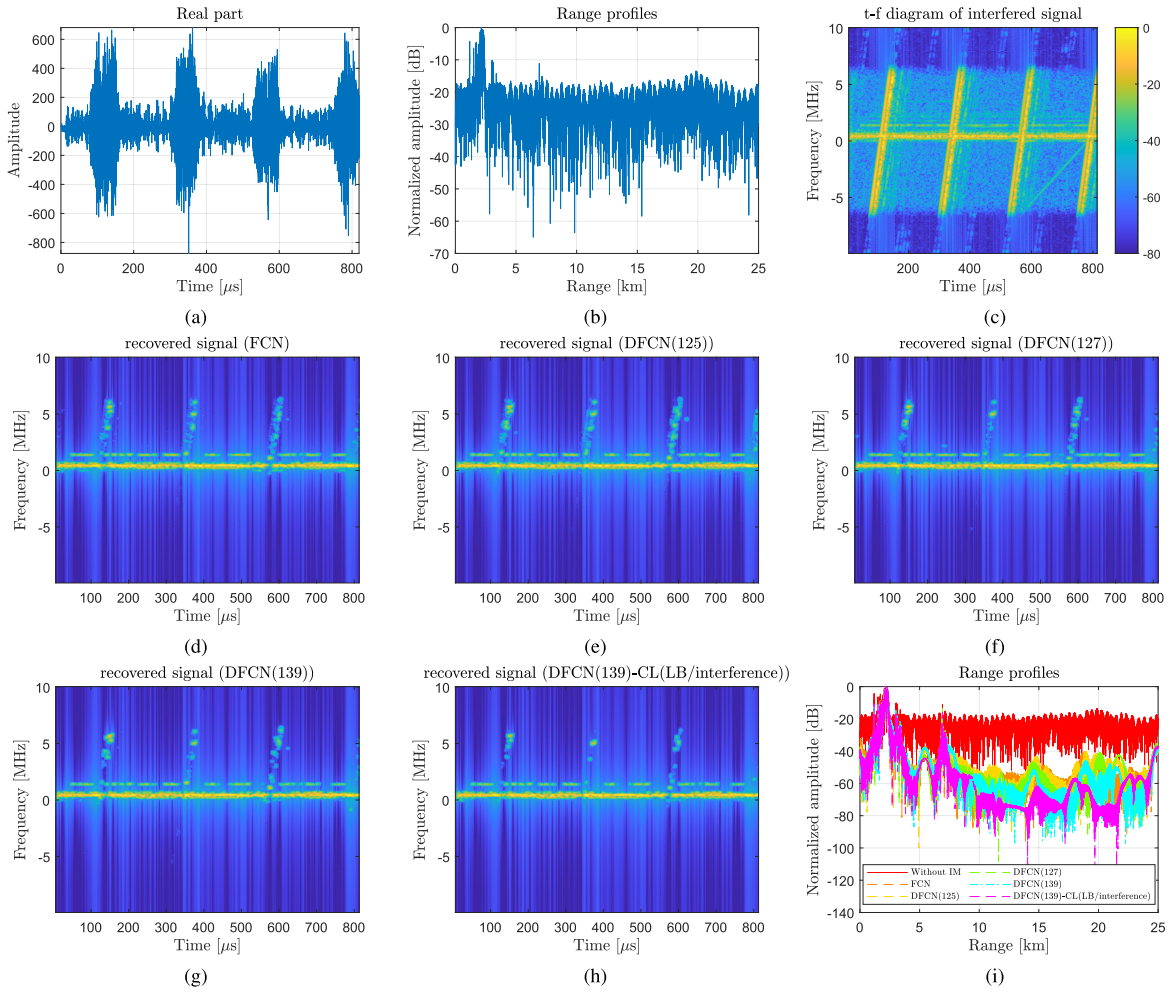
Fig. 14.   Interference mitigation for measured radar signals in the scene of A13 highway with traffic flow in Delft. (a) shows the acquired beat signal contaminated by mutual interference (b) its range profiles and (c) its time-frequency diagram. (d) the results of IM of the FCN. (e)-(f) the results of IM of the DFCNs with differernt settings of dilated rates. (g) the result of IM of the DFCN(139)-CL(LB/interference). (h) displays the corresponding range profiles obtained after IM. All the models were trained using only simulated data.

With much fewer parameters compared to the ResNet and U-Net, the proposed network uses less memory and takes only 2-3 ms to process the signal in a sweep with 1000 points.

As shown in Fig. 12, a complex highway scenario is considered to demonstrate the interference mitigation performance of the proposed approach. The IM results of an interfered beat signal of six point targets at the distance of [100, 70, 40, 20, 20, 20] m are illustrated. Due to strong interferences, the weak targets are almost immersed in the raised noise floor (see Fig. 12(a)-(c)). After being processed with the FCN using the normal convolution, the interferences and noise are significantly suppressed. However, some residual interference components and noise are still observed (Fig. 12(d)-(f)). By contrast, the proposed FCN with the dilated convolution and contrastive learning strategy further mitigates the residual interferences and noise (Fig. 12(g)-(j)); consequently the noise floor of the range profiles decreases as well (Fig. 12(l)).

To further evaluate the effects of the proposed IM approach on range Doppler processing, the beat signals in the 256 consecutive sweeps were generated for the scenario above with six point targets at [100, 70, 40, 20, 20, 20] m. The targets' Doppler velocities were set to [0, 0, 0, 3, 0, -10] m/s relative to the primary radar, and three aggressor FMCW radars moved

away from the primary radar with Doppler velocities of 0 m/s, 3 m/s and -10 m/s, respectively. To evaluate the detection performance and the accuracy of speed estimation on the RD maps of the beat signals, the CA-CFAR detector and the peak grouping algorithm were utilized for detection.

The detection results on the RD maps of the interfered signal, the clean reference signal and the recovered signals processed by the FCNs and other IM approaches are shown in Fig. 13(a)-(h). The center of the white rectangle indicates the point if the detector labels it as a target. Due to strong interferences, only the target with a Doppler speed of 3 m/s is detected in the RD map of interfered signal, and all the other five targets are missing (Fig. 13(a)). Although most targets can be detected in the RD map after IM with the FCN, the target at the distance of 20 m with a Doppler speed of 0 m/s is still missing (Fig. 13(c)), which is caused by the residual interferences near the missed target. By contrast, the RD maps after IM processing by the CNN, U-Net, and CV-FCN lead to the same detection results (see Fig. 13(d)-(f)). By using dilated convolutions and the proposed contrastive learning strategy, the interferences around the target at the distance of 20 m is further suppressed (Fig. 13(g)). Consequently, all the six targets can be detected, and the range/velocity estimation error

with all the recovered signals processed by the IM approaches is zero in this case. Although all the targets are detected in the RD map after IM processing with the ResNet-based approach, residual interferences can still be observed (Fig. 13(h)), which may cause false alarms in interference-dense scenarios.

VI. Measured Results

In this section, the measured radar signals collected using the PARSAX radar in the A13 highway with traffic flow in Delft [11] were used to test and verify the generalization of the proposed method. Due to the fact that the clean reference cannot be obtained with the radar system, which is generally the case in practice, the qualitative interference mitigation results on the radar signals collected on the highway, including the signal waveform in the time domain, the t-f diagram, and range profiles of beat signals, are shown in Fig. 14. As shown in Fig. 14(a) and (c), four large pulses (corresponds to four inclined thick lines in the t-f diagram) caused by strong interferences can be observed. Then, the interferences lead to increased noise floor in the range profiles of received beat signals, and the two weaker targets cannot be detected (see Fig. 14(b)). To overcome the missed detection of targets caused by the strong interferences, the proposed IM approaches were used to suppress the interferences in measured radar signals. Note that the networks were trained using the dataset including only simulated radar signals.

The t-f diagram of the recovered signal processed by the normal FCN is shown in Fig. 14(d). The interferences are entirely removed in the negative frequency, but there are still residual interferences mixed with the desired spectrum of targets in the positive frequency. In contrast, with the receptive field of the network gradually increasing, the DFCNs show a slightly better IM performance (Fig. 14(e)-(g)). Furthermore, with the contrastive learning, the interference in the positive frequency can be removed as shown in Fig. 14(h). After IM, three peaks of the targets can be clearly seen in the range profile (Fig. 14(i)). Moreover, the DFCN trained with contrastive learning shows a significantly lower noise floor.

The experimental results on measured radar signals collected in the real-world scenes have shown a good generalization performance of the proposed method. Moreover, we want to mention that the PARSAX radar in TU Delft uses an arbitrary waveform generator (AWG) to generate an FMCW waveform with a central frequency of 125 MHz and then it is up-converted by mixing with a synchronized stable local oscillator (LO) at a frequency 3.315 GHz for transmission. In our simulation, the center frequency of the FMCW waveform is set to 76.5 GHz. Although there exists a difference in radar parameters, the networks trained using only simulated data can still show good interference mitigation performance in the measured radar signals. Furthermore, one can see the effectiveness of the contrastive learning in helping the neural networks to remove the residual interferences and noise in measured radar signals.

VII. Conclusion

In this paper, the DL-based interference mitigation approach in the t-f domain is proposed for automotive FMCW radar.

The dilated convolution was used to build the network (i.e., DFCN). Meanwhile, the optimal setting of the dilated rate in each convolutional layer is proposed. The proposed DFCN(139) shows a significantly higher SINR than the normal FCN. Then, the contrastive loss function and strategy were designed. No feature extraction network and hyper-parameter are needed. The networks trained using contrastive learning show better performance than the one trained with only MSE. Besides, even though the frequency bands of simulated and measured signals are different, our method still shows good results on the measured data. The proposed dilated convolution and contrastive learning can be easily integrated into existing DL-based IM methods without increasing the network complexity. Therefore, the proposed IM approach can be better applied into the actual radar system.

In the future, the deformable convolution with dilated rate would be considered in the interference mitigation problems to realize a more flexible IM network.

References

[1] F. Wen, J. Shi, G. Gui, H. Gacanin, and O. A. Dobre, "3-D positioning method for anonymous UAV based on bistatic polarized MIMO radar," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 815–827, Jan. 2023.

[2] Y. Guo, X. Wang, X. Lan, and T. Su, "Traffic target location estimation based on tensor decomposition in intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 18, 2022, doi: 10.1109/TITS.2022.3165584.

[3] S. Lim, J. Jung, B.-H. Lee, J. Choi, and S.-C. Kim, "Radar sensor-based estimation of vehicle orientation for autonomous driving," *IEEE Sensors J.*, vol. 22, no. 22, pp. 21924–21932, Nov. 2022.

[4] W. Kim, H. Cho, J. Kim, B. Kim, and S. Lee, "Target classification using combined YOLO-SVM in high-resolution automotive FMCW radar," in *Proc. IEEE Radar Conf.*, Sep. 2020, pp. 1–5.

[5] M. Umehira, Y. Watanabe, X. Wang, S. Takeda, and H. Kuroda, "Inter-radar interference in automotive FMCW radars and its mitigation challenges," in *Proc. IEEE Int. Symp. Radio-Frequency Integr. Technol. (RFIT)*, Sep. 2020, pp. 220–222.

[6] C. Aydogdu, H. Wymeersch, O. Eriksson, H. Herbertsson, and M. Rydström, "Synchronization-free RadChat for automotive radar interference mitigation," *Sustainability*, vol. 13, no. 12, p. 6891, Jun. 2021.

[7] F. Norouzian, A. Pirkani, E. Hoare, M. Cherniakov, and M. Gashinova, "Automotive radar waveform parameters randomisation for interference level reduction," in *Proc. IEEE Radar Conf.*, Sep. 2020, pp. 1–5.

[8] J. Wang, "CFAR-based interference mitigation for FMCW automotive radar systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12229–12238, Aug. 2022.

[9] J. Mun, S. Ha, and J. Lee, "Automotive radar signal interference mitigation using RNN with self attention," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3802–3806.

[10] N.-C. Ristea, A. Anghel, and R. T. Ionescu, "Fully convolutional neural networks for automotive radar interference mitigation," 2020, *arXiv:2007.11102*.

[11] J. Wang, R. Li, Y. He, and Y. Yang, "Prior-guided deep interference mitigation for FMCW radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5118316.

[12] J. Rock, M. Toth, E. Messner, P. Meissner, and F. Pernkopf, "Complex signal denoising and interference mitigation for automotive radar using convolutional neural networks," in *Proc. 22th Int. Conf. Inf. Fusion (FUSION)*, 2019, pp. 1–8.

[13] M. L. L. de Oliveira and M. J. G. Bekooij, "Deep convolutional autoencoder applied for noise reduction in range-Doppler maps of FMCW radars," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 630–635.

[14] J. Fuchs, A. Dubey, M. Lübke, R. Weigel, and F. Lurz, "Automotive radar interference mitigation using a convolutional autoencoder," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 315–320.

[15] Y. He, X. Li, R. Li, J. Wang, and X. Jing, "A deep-learning method for radar micro-Doppler spectrogram restoration," *Sensors*, vol. 20, no. 17, p. 5007, Sep. 2020.

[16] W. Fan et al., "Interference mitigation for synthetic aperture radar based on deep residual network," *Remote Sens.*, vol. 11, no. 14, p. 1654, Jul. 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[18] J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[19] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[20] H. Wu et al., "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10546–10555.

[21] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 319–345.

[22] J. Wang, M. Ding, and A. Yarovoy, "Matrix-pencil approach-based interference mitigation for FMCW radar systems," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 11, pp. 5099–5115, Nov. 2021.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[24] P. Wang et al., "Understanding convolution for semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1451–1460.

[25] X. Jia, Y. Peng, J. Li, Y. Xin, B. Ge, and S. Liu, "Pyramid dilated convolutional neural network for image denoising," *J. Electron. Imag.*, vol. 31, no. 2, Apr. 2022, Art. no. 023024.

[26] C. Liu, Z. Shang, and A. Qin, "A multiscale image denoising algorithm based on dilated residual convolution network," 2018, *arXiv:1812.09131*.

[27] Z. He, X. Luo, Y. Zhong, C. Jiang, and L. Zhao, "Information extraction method based on dilated convolution and character-enhanced word embedding," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery (CyberC)*, Oct. 2020, pp. 138–143.

[28] J. Ma, "Scribble-supervised ROI extraction using residual dense dilated network for remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Jul. 2021, pp. 2791–2794.

[29] L. Qian, Y. Lian, Q. Wei, S. Wu, and J. Zhang, "ODCN: Optimized dilated convolution network for 3D shape segmentation," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2019, pp. 378–389.

[30] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[31] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[32] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
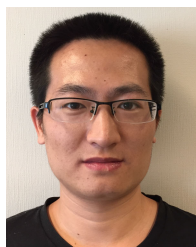
Jianping Wang (Member, IEEE) received the Ph.D. degree in electrical engineering from the Delft University of Technology, The Netherlands, in 2018.

From August 2012 to April, 2013, he was a Research Associate with the University of New South Wales, Australia, on frequency modulated continuous wave synthetic aperture radar signal processing for formation flying satellites. He is currently a Post-Doctoral Researcher with the Group of Microwave Sensing, Signals and Systems (MS3), Delft University of Technology. His research interests include microwave imaging, signal processing, and antenna array design.

Dr. Wang has served as a reviewer for many IEEE journals. He was a TPC Member of the IET International Radar Conference, Nanjing, China, in 2018. He was a finalist of the Best Student Paper Awards from the International Workshop on Advanced Ground Penetrating Radar (IWAGPR), Edinburgh, U.K., in 2017, and the International Conference on Radar, Brisbane, Australia, in 2018.

Runlong Li received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2020, where he is currently pursuing the M.S. degree in information and communication engineering.

His current research interests include radar signal processing, deep learning, and autonomous driving.

Xinqi Zhang received the B.S. degree in communication engineering from Henan Normal University, Henan, China, in 2020. He is currently pursuing the M.S. degree in electronic information with the Beijing University of Posts and Telecommunications. His research interests include radar signal processing and person identification.

Yuan He (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the National University of Defense Technology, Changsha, China, in 2007 and 2010, respectively, and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2014. He is an Associate Professor at the Beijing University of Posts and Telecommunications, Beijing, China. His main research interests are radar signal processing, wireless communication, and electromagnetic computation.