Multiview Low-Rank Hybrid Dilated Network for SAR Target Recognition Using Limited Training Samples

ZHAOHUI XUE, (Member, IEEE), AND MENGXUE ZHANG
School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

ABSTRACT

Deep learning (DL) have been intensively exploited for synthetic aperture radar (SAR) target classification, and we have witnessed exciting performances provided by previous studies. Multiview-based DL methods exhibited huge potentials for SAR target classification since they can generate adequate sample images from few raw images. In order to further improve the generalization performance of the multiview-based framework with limited training samples, we propose a novel multiview low-rank hybrid dilated network (MLHDN) for SAR target recognition. Firstly, we design a parameter-sharing hybrid dilated convolution (HDC) to learn multiview features. Secondly, a composite low-rank bilinear pooling (CLRBP) is proposed to fuse multiview features and to reduce their dimensions, yielding class-oriented and compacted feature vectors which are distinctive and representative for classification. Finally, a Softmax layer is used as a classifier. Accordingly, MLHDN encompasses fewer parameters compared to existing multiview-based DL methods. Experimental results demonstrate that MLHDN can achieve state-of-the-art performances on the moving and stationary target acquisition and recognition (MSTAR) datasets for SAR target classification, yielding an accuracy of 96.13% with only 10 training samples per class.

INDEX TERMS

Synthetic aperture radar (SAR), target classification, multiview, convolutional neural networks (CNNs), hybrid dilated convolution (HDC), composite low-rank bilinear pooling (CLRBP).

I. INTRODUCTION

Synthetic aperture radar (SAR) has been widely used for Earth remote sensing due to its special imaging capabilities of day-and-night and all-weather [1]. With distinctive characteristics of high-resolution, multi-aspect, multi-dimension, and multi-polarization, SAR images are capable of conducting military and civilian surveillance tasks, such as automatic target recognition (ATR). In general, an entire computation scheme of SAR-ATR includes three exclusive stages: detection, discrimination, and classification [2].

Various existing methods implemented at the classification stage of SAR-ATR can be categorized as model-design-based and feature-learning-driven methods. The model-design-based methods make prediction based on the predefined target model and feature matching engineering [3], which are complex in model design. Whereas, the feature-learning-driven methods rely on target templates or feature representations of the target for classification, e.g., using support vector machines [4], ensemble learning [5], hidden Markov model [6], manifold learning [7], graph learning [8], sparse representation [9], template matching [10], compressive sensing [11], low-rank representation [12], which is more flexible and can be easily implemented.

Recently, deep learning (DL) has been intensively exploited for SAR target classification [13]. Extensive experiments have found that resorting to automatically learn data-driven features from SAR images, DL methods have achieved excellent classification performance in SAR-ATR. According to the issues they solved, most of those methods can be categorized into the following families.

Firstly, many studies aimed at addressing the overfitting problem due to limited training samples. A seminal
work [14] of A-ConvNets (a new all-convolutional networks) was designed to reduce the free parameters, and it only consists of sparsely connected layers, rather than fully connected layers which is commonly used in convolutional neural network (CNN). Other studies solved this issue by using various DL methods and learning strategies. For example, transfer learning was combined with stacked convolutional auto-encoders [15]; Convolutional highway unit architecture was formed by combining a modified convolutional highway layer, a max pooling layer, and a dropout layer [16]; Semi-supervised learning was combined with a deep convolutional generative adversarial networks [17], [18]; Deep memory convolutional neural networks was trained by transferring parameters in two steps [19].

Secondly, considering the issues of translation of target, randomness of speckle noise, lack of pose image in training data, different data augmentation operations were introduced into CNN [20]–[22]. Although data augmentation can increase data diversity and robustness of learned features, the augmentation rule is human-designed and need experiments to validate its effectiveness. Therefore, stacked auto-encoder was exploited by using a supervised constraint (a restriction based on Euclidean distance and a dropout step) [23].

Thirdly, the spatial and channel correlations between features are rarely exploited in CNN-based SAR target classification. To this end, various attention mechanisms were introduced into CNNs to improve the representational power of salient features. For example, a lightweight CNN [24] model was proposed by incorporating a “Squeeze-and-Excitation” block with a spatial attention model, which they term the “channel-wise and spatial attention” block.

Finally, the space-varying backscattering information collected by SAR sensor is often missed by traditional methods for SAR target classification. SAR target classification will be benefited from multi-aspect measurements, where the target and the associated SAR image are often highly sensitive to the viewing angles. Therefore, a novel multi-aspect-aware method is proposed to achieve this idea through the bidirectional long short-term memory recurrent neural networks [25]. Similarly, multiview CNN framework was employed to learn the classification features from the couple of multiview SAR images autonomously and effectively [26]–[28]. In addition, a region-based convolutional neural network was exploited to address the issue of large scaled variabilities of objects in large scene SAR images [29], [30].

Recent researches have proved that DL-based methods are effective for SAR target classification, but they usually require large training samples to ensure the accuracy. However, it would be much difficult and expensive to obtain enough training samples in practical applications. Multiview-based DL methods can generate sufficient sample images from few raw image based on the diversity of the aspect-angle, which is a natural choice to tackle the sample limitation issue of SAR target classification. Based on the above observations, we focus on exploiting a novel multiview-based DL method for SAR target classification using limited training samples.

In this paper, we propose a novel composite low-rank bilinear pooling (CLRBP) to effectively leverage multiview information for SAR target classification. On the one hand, CLRBP can map high-dimensional bilinear features into low-dimensional and compact features by using low-rank representation. On the other hand, CLRBP exerts a composite decomposition to model similar and exclusive information from multiview features, yielding class-oriented feature vectors, which is more representative and discriminative for classification. In this context, a novel multiview low-rank hybrid dilated network (MLHDN), which is composed of a parameter-sharing hybrid dilated convolution (HDC), a CLRBP, and a Softmax layer, is proposed to realize SAR target classification.

It is worth noting that some recent studies including [25]–[28] focused on multiview-based DL for SAR classification. However, the major conceptual differences between those existing methods and ours are twofold: 1) We exploit bilinear pooling to leverage multiview features and low-rank representation to reduce their dimensions, whereas existing methods simply employed concatenation or summation to fuse multiview features. 2) We introduce unpadded dilated convolutional layers to enlarge the receptive field and aggregate global information, whereas previous studies mainly used standard convolutional kernel, and none of pooling layer is used in our network. Moreover, although CLRBP is inspired by [31], there exists several salient distinctions: firstly, CLRBP consists of two independent projection matrices as we decompose the features into two disparate parts; secondly, one projection matrix in CLRBP is constrained as sharing among different views to learn similar features using fewer parameters; thirdly, a learnable weight is employed to module the balance between similar and exclusive features for each class.

In this context, the main contributions of our work can be summarized as follows.

1) We propose a novel fusing method, namely CLRBP, which can reduce the dimensions of multiview features by using two low-rank projections for the similar and exclusive information respectively, which has fewer learnable parameters based on a composite decomposition, thus reducing the risk of over-fitting under the circumstance of limited training samples. In addition, a parameter-sharing HDC is introduced to extract more informative features since it can increase the receptive field and aggregate global information.

2) We propose a novel end-to-end network MLHDN, integrating a parameter-sharing HDC, a CLRBP, and a Softmax layer to realize SAR target classification with limited training samples. As we know, this network is unique in the related literatures. The experimental results demonstrate that MLHDN is extendable to the number of views. In addition, MLHDN is superior
The proposed end-to-end DL architecture (MLHDN) for SAR target classification is graphically illustrated in Fig. 1. Firstly, the input multiview images are rotated by their specific aspect angles. Secondly, a parameter-sharing HDC is introduced to learn multiview features. Thirdly, a CLRBP is proposed to learn representative and discriminative feature vectors. Finally, a Softmax activation is adopted to yield the prediction probability for each class.

### A. MULTIVIEW SAR IMAGE ACQUISITION

In general, multiple SAR images of different aspects of a target can provide more information than a single SAR image, which contributes to SAR-ATR [32].

Provided that a SAR sensor of an airborne radar system can acquire a series of images for a ground target with diverse aspect angles during a flight. However, there is a tough tradeoff between recognition accuracy and task feasibility in terms of time. Accordingly, only a few SAR images can be obtained, and these images recording a target of diverse aspect angles, make up a SAR image set. In order to alleviate this issue, the author in [26] proposed a novel multiview SAR image generation method, which can generate adequate multiview SAR images from a SAR image set based on the image combinations, i.e., given the view interval angle and the number of views, a very large number of view combinations of SAR images per class can be obtained.

For example, given a view interval angle of 360° and a view number of 4, the theoretically largest combination number is $A_4^4$ if considering the order and $C_4^4$ without considering the order when using 10 images per class. However, with a smaller view interval angle like 45°, this combination number would decline rapidly because many of them are not within a same interval.

Furthermore, considering the time complexity, a random selection worked on all generated image combinations is needed [26]. To be specific, a sample-balanced strategy is adopted in our experiments, i.e., supposing the total number is set to 2000 for each class, 200 image combinations are randomly selected for each sample (a image combination belongs to a sample if this sample locates at the first position).

### B. HYBRID DILATED CONVOLUTION

Generally, HDC is composed of several 2D dilated (or atrous) convolutional layers which is constructed by inserting zero holes between each pixel in the convolutional kernel [33]. Compared to a standard convolutional layer, a dilated convolutional layer can increase the resolution of intermediate feature maps. However, there is a gridding issue caused by a dilated convolutional layer since the center pixel can only perceive information in a checkerboard fashion, losing a large portion of low-level information. In order to alleviate this problem, the authors in [34] proposed a solution by manually designing a group of dilated convolutional layers with different dilation rates to ensure that the final receptive field could fully cover a square region without any holes. Supposing $n$ convolutional layers with a same kernel size of $K \times K$ that having dilation rates of $r = [r_1, r_2, \ldots, r_n]$, a constraint is defined as follow

$$M_i = \max \{M_{i+1} - 2r_i, 2r_i - M_{i+1}, r_i\}, \quad i = 1, 2, \ldots, n-1$$

$$M_n = r_n, \quad (1)$$

where $M_i$ is the maximal distance between two nonzero values, and the ultimate goal is to let $M_2 \leq K$. For example, a kernel size $K = 3$ and $r = [1, 3, 9, 27]$. Then, we have $M_1 = 1, M_2 = 3, M_3 = 7, \text{ and } M_4 = 27$. Obviously, $M_2$ satisfies this goal, and these four values are also the theoretical maximum for a four-layer architecture.

Based on this criterion, a parameter-sharing HDC is designed as a multiview feature extractor of our network. Compared to other extractors, HDC can effective enlarge the receptive field and aggregate global information so that it can obtain more representative features.

### C. COMPOSITE LOW-RANK BILINEAR POOLING

We propose a CLRBP to obtain class-oriented feature vectors from the high-dimensional bilinear pooling feature maps.
Firstly, a bilinear pooling [35] is used to calculate orderless bilinear representations from the features extracted by HDC. The bilinear pooling operation can be formulated as
\[ B_i = F_i^T F_i, \quad i = 1, 2, \ldots, N, \] where \( F_i = [f^1, f^2, \ldots, f^V] \in \mathbb{R}^{k \times d} \) is a intermediate feature matrix from the \( i \)-th multiple images set with \( V \) views, and \( B_i \in \mathbb{R}^{d \times d} \) is the \( i \)-th bilinear pooling feature map.

Since \( B_i \) has high dimensions, we attempt to reduce its dimensions and acquire a compact feature vector from \( B_i \), which can be formulated as
\[ S_i = vec(B_i)P, \]
where \( S_i \in \mathbb{R}^M \) is a compact feature vector \((M \ll d \times d)\) and \( P \) is a projection matrix. However, it may be unsuitable to directly learn \( P \) for \( B_i \), since \( B_i \) may contain redundant information between different views. On the one hand, for different features \( f \in F_i \), they may encompass similar information since they belong to an identical target. On the other hand, they may also encompass exclusive information since they come from diverse views with different aspect angles. In this circumstance, we propose to formulate \( S_i \) as
\[ S_i = vec(B_i^f)P^s + vec(B_i^e)P^e, \]
where \( B_i^f \) represents similar features among images which are irrelevant to the view, and \( B_i^e \) represents exclusive features belonging to different images. The addition operation is used to fuse the two features.

However, it is difficult to directly fetch \( B_i^f \) and \( B_i^e \) because they are closely included in \( B_i \). Therefore, we only decompose \( P \) to approximate \( S_i \). In this circumstance, \( S_i \) can be estimated as
\[ S_i = vec(B_i^f)P^s + vec(B_i^e)P^e, \]
where \( P^s \) is constrained as the shared parameters between different views, and \( vec(B_i^f)P^s \) is the estimation of \( vec(B_i^f)P^s \). While \( P^e \) is unconstrained, and \( vec(B_i^e)P^e \) is the estimation of \( vec(B_i^e)P^e \). Besides, a parameter \( \lambda \) is employed to modulate the balance. Thus, \( S_i \) is calculated as
\[ S_i = vec(B_i)^{(\lambda P^s + (1 - \lambda)P^e)} = vec(B_i)(\lambda P^s + P^e - \lambda P^e). \]

Finally, we impose a low-rank constraint when learning \( P^s \) and \( P^e \) based on the eigen-decomposition used in [31]. The projection matrices \( P^s \) and \( P^e \) will have fewer learnable parameters under the low-rank constraint. Meanwhile, fewer parameters can reduce the risk of overfitting when training samples are limited. Therefore, it is beneficial to improve the generalization performance. Consequently, \( S_{ij} \) in \( S_i \) is eventually formulated as
\[ S_{ij} = vec(B_i)(vec(\lambda U_{ij}^+ U_{ij}^+ - \lambda U_{ij-} U_{ij-}^T + V_{ij}^T V_{ij}^T - V_{ij-} V_{ij-}^T + \lambda V_{ij-} V_{ij-}^T - \lambda V_{ij+} V_{ij+}^T)), \quad j = 1, 2, \ldots, M, \]
where \( U_{ij}^+ \) and \( U_{ij-} \) \( V_{ij}^+ \) and \( V_{ij-} \) are two ranks related to \( P^s \) and \( P^e \), we can control their respective learnable parameters by subtly modulating the corresponding ranks. Where \( r_s/2 \) is the second dimension of \( U_{ij}^+ \) and \( U_{ij-} \), i.e., \( r_s \) controls the low-rank level of \( P^s \). Similarly, \( r_e/2 \) is the second dimension of \( V_{ij}^+ \) and \( V_{ij-} \) and \( r_e \) controls the low-rank level of \( P^e \). With a smaller \( r_s \) or \( r_e \), the corresponding projection matrix would have a lower degree of the freedom, contributing to more compact feature learning.

Once these two hyperparameters of \( r_s \) and \( r_e \) are determined, the dimensions of the learnable parameters is fixed, then they are randomly initialized and learned from data during training. Finally, we can obtain a compact feature vector \( S_i \) from \( F_i \).

\section{D. Multiview Low-Rank Hybrid Dilated Network}

With the above major components at hand, we now can build our network. Firstly, the SAR images are spatially downsampled with a fixed size of \( 83 \times 83 \), and they are rotated by their aspect angles to reduce the sensitivity of the orientation difference on classification accuracy [26]. Then, the designed parameter-sharing HDC is introduced to learn multiview features. The parameter-sharing HDC consists of 4 unpaded dilated convolutional layers where the kernel size is set to \( 3 \times 3 \) for each layer. In addition, the filters and dilation rates are empirically set to \([16, 32, 64, 128]\) and \([1, 3, 9, 27]\), respectively, for each layer.

Note that, the setting of filters follows the previous network [14], and the setting of dilation rates aims to maximize the dilation rate of each layer. Next, bilinear pooling is applied to the multiview features, and the obtained bilinear pooling feature maps are further projected into class-oriented feature vectors via the proposed CLRBP, where \( M \) is set to 10 which is equal to the number of classes, and the \( r_s \) and \( r_e \) are experimentally set to 4 and 2, respectively. Finally, a Softmax activation is applied to the feature vectors to yield the prediction probability for each class. The whole architecture of the proposed MLHDN is reported in Table 1.

It’s worth noting that the weights of network parameters are randomly initialized, a cross entropy loss is employed as the loss function, and Adam is adopted as the optimizer. In the training phase, a minibatch size is set to 32, an initial learning rate is empirically set to 0.001, and an epoch is set to 10.

\section{III. EXPERIMENTAL RESULTS}

\subsection{A. SAR DATA SETS}

The moving and stationary target acquisition and recognition (MSTAR) data sets [36] are utilized to evaluate the classification performance of our proposed method. MSTAR data sets contain hundreds of thousands of X-band SAR images of ground targets including different target types, aspect angles, depression angles, serial numbers, and articulation. The publicly released data sets include ten different categories of ground targets: BMP-2, BRDM-2, BTR-60, and
BTR-70 (armored personnel carrier); T-62 and T-72 (tank); 2S1 (rocket launcher); ZSU-234 (air defense unit); ZIL-131 (truck); D7 (bulldozer). The targets have been measured over the full 360° azimuth angles with 1°-5° increments. Those targets are presented with 0.3 m resolution, and most of them have 128 × 128 pixels. Fig. 2 visually depicts those ten types of targets.

B. EXPERIMENTAL SETTINGS

- For training and test image sets preparation, we follow a multiview data formation strategy used in [26]. Note that the training interval in our experiments is set to 360°, and we take the order into consideration to create enough image combinations using a few training samples, and 2000 train image combinations per class are selected as mentioned above. As for the generation process of the test image combinations, we do not consider the order, and the test interval is analyzed in the following experiments. Similarly, 2000 test image combinations per class are randomly selected from a large set. The training and test samples are shown in Table 2.

- For performance comparison, we implement five DL-based methods including multi-layer perceptron (MLP) [36], deep convolutional neural networks (DCNN), a new all-convolutional networks (A-ConvNet) [14], convolutional highway unit network (CHU-CNN) [16], and four-view deep CNNs (4-VDCNN) [26]. To better evaluate the proposed method, we also implement MLHDN with 1-view for comparison, which only needs one image as the input. Note that, all those six methods use the aspect angle information to rotate the input images.

- Experimental results are quantitatively evaluated using classification accuracy, parameter size, and computational time. It should be noted that all the implementations were carried out using TensorFlow 1.12.0 and Keras 2.2.4 on the Ubuntu 16.04 LTS desktop PC equipped with GeForce GTX 1070 (8GB), Intel Xeon E3 CPU (at 3.3GHz x 8), and 32 GB of RAM. Moreover, we have conducted ten independent runs in each experiment, and reported the averaged accuracies with standard deviations.

C. EXPERIMENT 1—HYPERPARAMETER ANALYSIS

In this experiment, we analyze the impacts of \( r_s \) and \( r_e \) on classification accuracy. To this end, \( r_s \) and \( r_e \) are respectively set to a same range of \( \{2, 4, 6, 8, 10\} \). The experimental results are reported in Table 3, where we found that the proposed method is insensitive to the two hyperparameters, and when \( r_s = 4 \) and \( r_e = 2 \), MLHDN achieves the highest accuracy with an OA of 96.13%. However, with a larger \( r_s \) or \( r_e \), the corresponding learnable PS or PE becomes complex.

### TABLE 1. The architecture of MLHDN.

| # Layer | Operation | Input size | Output size | Parameter size |
|---------|-----------|------------|-------------|----------------|
| 1       | Dilated Conv.16@3×3, Rate = 1 / Relu | 83 × 83 × 1 × 4 | 81 × 81 × 16 × 4 | 160 |
| 2       | Dilated Conv.32@3×3, Rate = 3 / Relu | 81 × 81 × 16 × 4 | 75 × 75 × 32 × 4 | 4640 |
| 3       | Dilated Conv.64@3×3, Rate = 9 / Relu | 75 × 75 × 32 × 4 | 57 × 57 × 64 × 4 | 18496 |
| 4       | Dilated Conv.128@3×3, Rate = 27 / Relu | 57 × 57 × 64 × 4 | 3 × 3 × 128 × 4 | 73856 |
| 5       | Concatenation | 3 × 3 × 128 × 4 | 9 × 512 | - |
| 6       | CLRBF | 9 × 512 | 1 × 10 | 10250 |
| 7       | Softmax activation | 1 × 10 | 1 × 10 | 10 |

### TABLE 2. Details of training and test data sets.

| Class | Serial No. | Depression | # Raw images | # Generated image combinations | Test | # Generated image combinations |
|-------|------------|------------|--------------|-------------------------------|------|-------------------------------|
| BMP-2 | 9563       | 17°        | 10           | 2k                            | 15°  | 196                           | 2k |
| BRDM-2| E71        | 17°        | 10           | 2k                            | 15°  | 274                           | 2k |
| BTR-60| k10yr7532  | 17°        | 10           | 2k                            | 15°  | 195                           | 2k |
| BTR-70| c71        | 17°        | 10           | 2k                            | 15°  | 196                           | 2k |
| T-62  | A31        | 17°        | 10           | 2k                            | 15°  | 273                           | 2k |
| T-72  | i32        | 17°        | 10           | 2k                            | 15°  | 196                           | 2k |
| 2S1   | b01        | 17°        | 10           | 2k                            | 15°  | 274                           | 2k |
| ZSU-234| d08       | 17°        | 10           | 2k                            | 15°  | 274                           | 2k |
| ZIL-131| E12       | 17°        | 10           | 2k                            | 15°  | 274                           | 2k |
| D7    | 92v13015   | 17°        | 10           | 2k                            | 15°  | 274                           | 2k |

### FIGURE 2. Examples of ten types of targets. Each class contains optical image (top) and the corresponding SAR image (bottom).
TABLE 3. The impacts of $r_e$ and $r_s$ on classification accuracy.

| $r_s$ | $r_e$  | 2       | 4       | 6       | 8       | 10      |
|-------|--------|---------|---------|---------|---------|---------|
| 2     | 95.96±2.16 | 95.71±2.34 | 95.83±1.64 | 95.97±2.54 | 95.31±2.14 |
| 4     | 94.13±2.39 | 95.48±2.05 | 95.75±1.69 | 95.90±2.15 | 95.37±2.43 |
| 6     | 95.96±2.09 | 95.66±2.20 | 95.26±1.41 | 95.91±2.25 | 95.75±2.44 |
| 8     | 95.63±2.30 | 95.52±2.44 | 95.37±2.52 | 95.35±2.05 | 95.61±2.16 |
| 10    | 95.45±2.36 | 95.49±1.75 | 95.10±2.36 | 95.91±2.03 | 95.06±2.28 |

TABLE 4. The contribution of parameter-sharing HDC to MLHDN.

| # Layer | MLHDN$^1$ | MLHDN$^2$ | MLHDN$^3$ | MLHDN (ours)$^4$ |
|---------|-----------|-----------|-----------|------------------|
| 1       | Conv16@3×3 / ReLu | Conv16@3×3 / ReLu | Conv16@3×3 / ReLu | Dilated Conv16@3×3, Rate = 3 / ReLu |
| 2       | Max pooling 2×2 | Average pooling 2×2 | Conv16@3×3, Stride = 2 / ReLu | Dilated Conv32@3×3, Rate = 3 / ReLu |
| 3       | Conv32@3×3 / ReLu | Conv32@3×3 / ReLu | Conv32@3×3 / ReLu | Dilated Conv64@3×3, Rate = 9 / ReLu |
| 4       | Max pooling 2×2 | Average pooling 2×2 | Conv32@3×3, Stride = 2 / ReLu | Dilated Conv128@3×3, Rate = 27 / ReLu |
| 5       | Conv64@3×3 / ReLu | Conv64@3×3 / ReLu | Conv64@3×3 / ReLu | - |
| 6       | Max pooling 2×2 | Average pooling 2×2 | Conv64@3×3, Stride = 2 / ReLu | - |
| 7       | Conv128@3×3 / ReLu | Conv128@3×3 / ReLu | Conv128@3×3 / ReLu | - |
| 8       | Max pooling 2×2 | Average pooling 2×2 | Conv128@3×3, Stride = 2 / ReLu | - |
| 9       | Concatenation | CLRB | Softmax activation | - |
| 10      |           |           |           | - |
| 11      |           |           |           | - |

| Class   | ZS1 | BMP2 | BRDM, 2 | BTR60 | BTR70 | D7  | T52  | T72  | ZIL131 | ZSU, 23.4 | OA (%) | κ = 100 |
|---------|-----|------|---------|-------|-------|-----|------|------|--------|---------|--------|---------|
| 2       | 85.96±10.96 | 84.39±10.39 | 87.78±10.16 | 99.64±1.08 |
| 4       | 54.39±13.71 | 45.37±13.23 | 59.35±15.28 | 86.09±13.85 |
| 6       | 73.43±9.46  | 82.23±10.14 | 72.93±10.25 | 83.53±6.71 | 99.86±1.30 |
| 8       | 66.97±11.31 | 73.76±11.18 | 70.48±7.80  | 89.13±11.01 | 99.86±1.30 |
| 10      | 94.65±4.73  | 93.94±4.87  | 94.39±9.81  | 99.98±0.05  | 99.86±1.30 |
| 11      | 79.47±8.81  | 84.75±11.81 | 96.11±5.08  | 98.33±2.45  | 99.86±1.30 |
| 2       | 81.06±3.29  | 82.92±3.96  | 84.58±4.98  | 96.13±2.39  | 99.86±1.30 |

$^1$ A variant of MLHDN by replacing HDC with standard convolution and max pooling.
$^2$ A variant of MLHDN by replacing HDC with standard convolution and average pooling.
$^3$ A variant of MLHDN by replacing HDC with two standard convolutions (Stride = 1 and Stride = 2).
$^4$ The proposed method.

These four methods have the same number of parameter size.

which may bring poor generalization performance. Considering this, relatively small values are suitable for these two hyperparameters when using limited training samples. Thus, in the remaining experiments, the optimal values for $r_s$ and $r_e$ are set to 4 and 2, respectively.

In addition, the impact of number of views on classification accuracy is analyzed in this experiment. As we can see from Fig. 3, the accuracy becomes higher with the increment of the number of views. For instance, the proposed method obtains an OA of 97.23% with the number of views reaches to 8. Note that, 4-VDCNN was designed for four views, thus the number of views is experimentally set to 4 in the following experiments for a fair comparison. Therefore, this result demonstrates that MLHDN has good capacity of extendibility in terms of number of views.

D. EXPERIMENT 2—ABLATION ANALYSIS

In this experiment, we evaluate the contribution of different ingredients to the classification performance of MLHDN. Firstly, we analyze the impact of parameter-sharing HDC on OA. To this end, we implement three variants of MLHDN: replacing HDC with standard convolution and max pooling, replacing HDC with standard convolution and average pooling, and replacing HDC with two standard convolutions (stride = 1 and stride = 2). As shown in Table 4, the proposed method achieves the best classification performance among different variants, with around of 11-15% improvements in terms of OA. As for class-specific accuracies, our method obtains the best results for different classes. We also
observe that traditional CNN-based methods with several interleaved standard convolutional layers and max or average pooling layers exhibit similar performances with OAs lower than 85%. This result demonstrates the superiority of HDC compared to traditional interleaved convolution and pooling architectures.

Secondly, we evaluate the effects of CLRBP on OA. To fulfill this objective, we implement three variants of MLHDN: replacing CLRBP + Softmax activation with FC, replacing CLRBP + Softmax activation with BP + FC, and replacing CLRBP with LRBP. As reported in Table 5, the proposed method obtains the highest OA, with 0.35-1.27% improvements compared to other counterparts. Note that, although these three variants can realize comparative performance, proposed CLRBP + Softmax activation obtains the best performance with the fewest number of parameter size. We also find that MLHDN achieves the best class-specific accuracies for six of ten classes. Especially, by employing BP, LRBP, and CLRBP, MLHDN obtains stepwise improvements in terms of OA, which certifies the effectiveness of the proposed CLRBP operation. In addition, when comparing HDC with CLRBP, we can see that they jointly contribute to the performance improvements of MLHDN. However, the contribution of HDC is more significant than that of CLRBP, i.e., HDC contributes to the improvement from 84.58% to 94.86%, and CLRBP leads to a further improvement from 94.86% to 96.13%. Note that, there are two reasons to explain the superiority of MLHDN compared to other variants when using limited training samples. Firstly, the parameter-sharing HDC can effectively aggregate global information without resolution loss while the resolution is a crucial factor for accurate target recognition [10]. Secondly, the CLRBP can extract more compact features by utilizing the low-rank property intrinsically existed in multiview SAR images.

### E. EXPERIMENT 3—VIEW INTERVAL ANGLE ANALYSIS

In this experiment, we test the impact of view interval angle on OA. Due to limited samples, the training view interval angle is set to 360° in order to generate enough multiview images. In contrast, the test view interval angle may be fixed by an actual reconnaissance task. As shown in Fig. 4, we also provide the results of other compared methods, and the proposed method obtains the highest OAs in different cases. It’s interesting to note that MLHDN (1-view) also obtains higher accuracy than the other methods. According to the results, when the time of a reconnaissance task is sufficient, which...
is similar with our experiment under a view interval angle of 135°, the proposed method achieves an OA of 98.13%. This is due to the fact that the images with a large view interval angle may contain more diverse information. Whereas, when the time is extremely limited, which is similar with our experiment under a view interval angle of 15°, MLHDN obtains a tolerable OA of 94.67%. Note that, in the following experiments, the view interval angle is set to 45°, which is same to that of 4-VDCNN for a fair comparison. This result demonstrates that MLHDN is robust to the view interval angle.

In general, the virtual aspect angles ought to be estimated, and the estimation errors of the target aspect angle may impact the recognition performance. As shown in Fig. 5, the performances of different methods decrease as the increment of standard deviation. This can be explained that when the aspect-angle estimation error becomes larger, the information induced from those wrongly estimated aspect angles deteriorates the training performance, thus leading to the decrease of prediction accuracy. It’s worth noting that, MLHDN still yields better performance than the other methods. For instance, MLHDN obtains an OA of 90.68% when the standard deviation reaches to 10, which is still acceptable for actual application. This result demonstrates that MLHDN is robust to the aspect-angle estimation error.

In this experiment, we evaluate the generalization performance of the proposed method under limited training samples per class. To this end, we randomly choose different numbers of training samples per class according to a range of {5, 6, ..., 15}. Fig. 6 plots the evolution of classification accuracy with the number of training samples per class. As shown in the figure, the proposed method achieves the best result in different cases. Especially, MLHDN obtains a tolerable OA of 88.28% with only 5 training samples per class, which is 6.93-40.90% higher than others. In addition, MLHDN obtains a remarkable OA of 98.67% when using 15 samples per class for training, which is 2.88-12.62% higher than others. Therefore, the result demonstrates that MLHDN has better generalization performance than the other counterparts under limited training samples. Note that, the performance of 4-VDCNN is not good with limited training samples. However, the OAs of 4-VDCNN greatly increase as the number of training samples becomes larger. When the number reaches to 15, 4-VDCNN can still provide competitive results. In addition, the other traditional methods also yield good results since they resort to the aspect-angle information.
of our method is 627s, which is a little higher than that of 4-VDCNN. Whereas, the test time of our method is 144s, which is almost the same to that of 4-VDCNN.

IV. CONCLUSION

Focusing on the limitations of existing multiview-based SAR target classification methods when facing with limited training samples, this paper presented a novel multiview low-rank hybrid dilated network. In this context, the superiority of MLHDN compared to others has been clearly verified.

Currently, we exploited a parameter-sharing HDC and a CLRBP for multiview-based SAR target classification with limited training samples. In the future, we will extend our method to few-shot learning which only needs very few training samples.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, “A tutorial on synthetic aperture radar,” IEEE Geosci. Remote Sens. Mag., vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, “Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review,” IEEE Access, vol. 4, pp. 6014–6058, 2016.

[3] L. C. Potter and R. L. Moses, “Attributed scattering centers for SAR ATR,” IEEE Trans. Aerosp. Electron. Syst., vol. 37, no. 2, pp. 643–654, Apr. 2001.

[4] U. Srinivas, V. Monga, and R. G. Raj, “SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 48, no. 2, pp. 969–980, Apr. 2012.

[5] Y. Sun, Z. Liu, S. Todorovic, and J. Li, “Adaptive boosting for SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 49, no. 1, pp. 112–125, Jan. 2007.

[6] G. Dong, N. Wang, and G. Kuang, “Sparse representation of monogenic signal: With application to target recognition in SAR images,” IEEE Image Process. Lett., vol. 27, no. 12, pp. 5983–5995, 2014.

[7] D. Papson and R. M. Narayanan, “Classification via the shadow region metric center scaling embedding for SAR ATR,” IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 1, pp. 180–192, Jan. 2014.

[8] S. Papson and R. M. Narayanan, “Classification using discriminative graphical models,” IEEE Trans. Aerosp. Electron. Syst., vol. 48, no. 2, pp. 969–980, Apr. 2012.

[9] Y. Huang, J. Peia, J. Yang, B. Wang, and X. Liu, “Neighborhood geometric center scaling embedding for SAR ATR,” IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 1, pp. 180–192, Jan. 2014.

[10] G. Dong, N. Wang, and G. Kuang, “Sparse representation of monogenic signal: With application to target recognition in SAR images,” IEEE Signal Process. Lett., vol. 21, no. 8, pp. 952–956, Aug. 2014.

[11] B. Ding, G. Wen, C. Ma, and X. Yang, “An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features,” IEEE Trans. Image Process., vol. 27, no. 12, pp. 5983–5995, Dec. 2018.

[12] L. Jin, J. Chen, and X. Peng, “Joint classification of complementary features based on multitask compressive sensing with application to synthetic aperture radar automatic target recognition,” J. Electron. Imag., vol. 27, no. 5, Art. no. 053034, 2018.

[13] Y. Huang, G. Liao, Z. Zhang, X. Xiang, J. Li, and A. Nehorai, “SAR automatic target recognition using joint low-rank and sparse multiview denoising,” IEEE Geosci. Remote Sens. Lett., vol. 15, no. 10, pp. 1570–1574, Oct. 2018.
X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, “Deep learning in remote sensing: A comprehensive review and list of resources,” IEEE Geosci. Remote Sens. Mag., vol. 5, no. 4, pp. 8–36, Dec. 2017.

S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, “Target classification using the deep convolutional networks for SAR images,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

Z. L. Huang, Z. X. Pan, and B. Lei, “Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data,” Remote Sens., vol. 9, no. 9, p. 21, Sep. 2017.

Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, “Deep convolutional highway unit network for SAR target classification with limited labeled training data,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 7, pp. 1091–1095, Jul. 2017.

F. Gao, Y. Yang, J. Wang, J. P. Sun, E. F. Yang, and H. Y. Zhou, “A deep convolutional generative adversarial networks (DCGAN)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images,” Remote Sens., vol. 10, no. 6, p. 21, Jun. 2018.

F. Gao, F. Ma, J. Wang, J. P. Sun, E. F. Yang, and H. Y. Zhou, “Semi-supervised generative adversarial nets with multiple generators for SAR image recognition,” Sensors, vol. 18, no. 8, p. 19, Aug. 2018.

R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, “SAR targets classification based on deep memory convolution neural networks and transfer parameters,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 8, pp. 2834–2846, Aug. 2018.

J. Ding, B. Chen, H. Liu, and M. Huang, “Convolutional neural network with data augmentation for SAR target recognition,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 3, pp. 364–368, Mar. 2016.

K. Du, Y. Deng, R. Wang, T. Zhao, and N. Li, “SAR ATR based on displacement- and rotation-insensitive CNN,” Remote Sens. Lett., vol. 7, no. 9, pp. 895–904, 2016.

S. A. Wagner, “SAR ATR by a combination of convolutional neural network and support vector machines,” IEEE Trans. Aerosp. Electron. Syst., vol. 52, no. 6, pp. 2861–2872, Dec. 2016.

S. Deng, L. Du, C. Li, J. Ding, and H. Liu, “SAR automatic target recognition based on Euclidean distance restricted autoencoder,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 7, pp. 3323–3333, Jul. 2017.

J. Q. Shao, C. W. Qu, J. W. Li, and S. J. Peng, “A lightweight convolutional neural network based on visual attention for SAR image target classification,” Sensors, vol. 18, no. 9, p. 21, Sep. 2018.

Z. Cui, S. Dang, Z. Cao, S. Wang, and N. Liu, “Multi-scale object detection in remote sensing imagery with convolutional neural networks,” ISPRS J. Photogramm. Remote Sens., vol. 145, pp. 3–22, Nov. 2018.

S. Kong and C. Fowlkes, “Low-rank bilinear pooling for fine-grained classification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 365–374.

G. F. Brendel and L. L. Horowitz, “Benefits of aspect diversity for SAR ATR: Fundamental and experimental results,” Proc. SPIE, vol. 4053, pp. 567–578, Aug. 2000.

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018.

P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, “Understanding convolution for semantic segmentation,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2018, pp. 1451–1460.

T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Moshing, and M. L. Bryant, “Standard SAR ATR evaluation experiments using the MSTAR public release data set,” Proc. SPIE, vol. 3370, pp. 566–573, Sep. 1998.

D. A. E. Morgan, “Deep convolutional neural networks for ATR from SAR imagery,” Proc. SPIE, vol. 9475, May 2015, Art. no. 94750F.