ROLLABLE LATENT SPACE FOR AZIMUTH INVARIANT SAR TARGET RECOGNITION

Kazutoshi Sagi, Takahiro Toizumi, and Yuzo Senda

Data Science Research Laboratories, NEC corporation

ABSTRACT

This paper proposes rollable latent space (RLS) for an azimuth invariant synthetic aperture radar (SAR) target recognition. Scarce labeled data and limited viewing direction are critical issues in SAR target recognition.The RLS is a designed space in which rolling of latent features corresponds to 3D rotation of an object. Thus latent features of an arbitrary view can be inferred using those of different views. This characteristic further enables us to augment data from limited viewing in RLS. RLS-based classifiers with and without data augmentation and a conventional classifier trained with target front shots are evaluated over untrained target back shots. Results show that the RLS-based classifier with augmentation improves an accuracy by 30 % compared to the conventional classifier.

Index Terms— SAR, ATR, autoencoder, latent feature, view, disentangle

1. INTRODUCTION

Automatic target recognition (ATR) on synthetic aperture radar (SAR) imagery is under keen interest especially in social security and defense applications. SAR ATR was considered difficult since human operators are no good at interpreting SAR images and crafting good manual features for ATR [1]. After deep neural network (DNN) techniques are introduced, SAR ATR reaches a feasible level [2].

Several classification approaches have already been proposed in SAR ATR. Wilmanski et al. applied convolutional neural network (CNN) to SAR classification problem and showed promising results of high accuracy [3]. The method achieved 97 % in accuracy while manual feature based classifiers showed approximately 70 %. The above paper relies on Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset [4]. The dataset contains SAR images of 360 viewing directions of 10 military ground target vehicles from depression angles of $15°$ and $17°$. Usually all SAR images from depression angle of $17°$ were used to train classifiers, but that is unrealistic in actual applications. There are still challenges due to limited viewing direction and scarce labeled data.

To obtain intrinsic features invariant to viewing direction, Song et al. used a generative DNN framework to capture the features and the viewing direction of a target [5]. The framework consists of a constructor, a generator, and an interpreter. The constructor uses labels of training data to form a continuous target feature space independent of viewing direction. The generator and the interpreter learn correspondences between SAR images and sets of point in the feature space and viewing direction. Even though the feature space is continuous, its representation capability could be limited to variety of the labels.

To overcome the challenges in a different way, Avolio et al. proposed use of a SAR simulator to generate plenty of training data from computer-aided design (CAD) models [6]. A classifier can be trained with the generated data, but CAD models are necessary instead.

Several approaches have been proposed for 3D view synthesis in other computer vision applications such as robotics, rendering and modeling. One of them is to employ conditional autoencoder to disentangle latent vactors into two parts, i.e. pose and identity [7]. The autoencoder modifies the pose part to produce a different view of an object. The identity part contains individuality of an object and thus it can be used for recognition.

A purpose of this paper is to propose new disentangling using a novel autoencoder. The novel autoencoder network is developed for learning 3D rotations of objects in SAR images. The learned autoencoder provides a discriminative view-invariant feature space. This disentangling can handle any SAR images without labels or CAD models of targets.

2. PROPOSED METHOD

2.1. Rollable Latent Space

We propose rollable latent space (RLS) to cope with variation from viewing directions. In object recognition, a latent vector is expected to represent disentangled information of a certain object. In the RLS, a latent vector consists of multiple sub-vectors, each of which has a fixed length of structural and directional information. Number of sub-vectors corresponds to the number of equally spaced viewing directions. The sub-vector can be rolled to represent a 3D rotated image of its structural information.

An RLS encoder converts an image **X** seen from a direc-

tion θ_i to a latent vector $\mathbf{Z}$,

$$\mathbf{Z}(\theta_i) = Encoder(\mathbf{X}(\theta_i)). \quad (1)$$

$\mathbf{Z}$ from another direction θ_j can be approximated by simply applying roll function to $\mathbf{Z}$,

$$\mathbf{Z}(\theta_j) = Encoder(\mathbf{X}(\theta_j)) \quad (2)$$
$$\simeq Roll(\mathbf{Z}(\theta_i), j-i), \quad (3)$$

where $Roll(\mathbf{Z}, s)$ rolls all sub-vectors of $\mathbf{Z}$ by shift parameter s. In a simple RLS, $Roll(\mathbf{Z})$ can be written as

$$Roll(\mathbf{Z}_k, s) = \mathbf{R}^s \cdot \mathbf{Z}_k, \quad (4)$$

where $\mathbf{Z}_k$ is a sub-vector of $\mathbf{Z}$, and $\mathbf{R}$ is a cyclic permutation matrix. Roll function can be interpolative to support continuous rotation.

To let *Encoder* learn this disentanglement, an autoencoder approach is employed. An RLS autoencoder is defined as

$$\mathbf{X}(\theta_j) \simeq Decoder(Roll(Encoder(\mathbf{X}_i), j-i)). \quad (5)$$

By feeding images of known objects to the autoencoder with varying s, *Encoder* and *Decoder* are obtained.

2.2. Feature Augmentation in RLS

To train a classifier, data augmentation is necessary especially for scarce labeled images like SAR. In addition to ordinary augmentation techniques, feature-level augmentation can be applied easily in RLS. A simple rolling function can produce a number of augmented feature vectors when using a latent vector $\mathbf{Z}$ in RLS as a feature vector. *Classifier* is trained to fulfill

$$\mathbf{Y}_i = Classifier(Roll(Encoder(\mathbf{X}_i), s_{rand})), \quad (6)$$

where $\mathbf{Y}_i$ is the label of $\mathbf{X}_i$, when s_{rand} is randomly given for feature augmentation.

2.3. Implementation

Figure 1 shows an entire structure of RLS-based autoencoder and classifier used in this paper. To promote disentanglement, variational autoencoder (VAE) is employed in conjunction with CNN [8]. In this paper a deep convolutional encoder-decoder architecture proposed by Yang et al. [7] is applied to the encoder and the decoder. The encoder network consists of three 5×5 convolution layers with stride 2 and 2-pixel padding and ReLU activation, and two fully connected layers. Note that employing CNN architecture reduces an error of the target shift of few pixels because of its convolutions and strides. The decoder takes symmetric architecture to the encoder. Fixed stride-2 convolution and upsampling are employed instead of max-pooling and unpooling. The classifier recieves the feature vector $\mathbf{Z}$ as an input. The classifier structure employed here simply consists of two additional fully connected layers with nodes of 120 and 5, respectively.
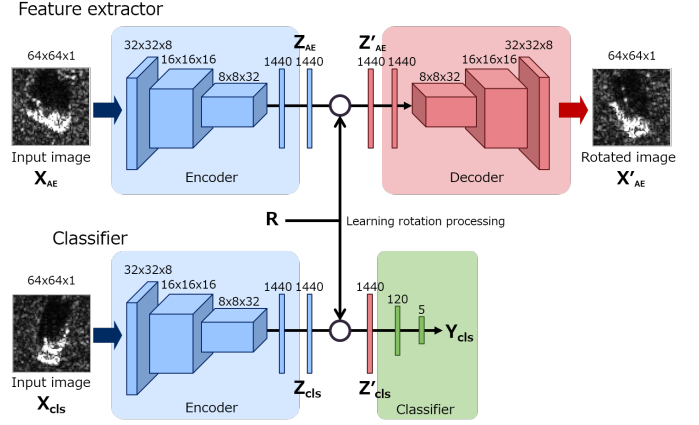
Fig. 1. Proposed network architecture for learning a RLS-based classifier.

3. EXPERIMENT

3.1. Setup

For the sake of demonstrating RLS potential, we perform an experiment based on realistic SAR ATR operation. The experiment procedure is following below,

1. RLS training using omnidirectional SAR images,

2. classifier training using SAR target front shots,

3. classifier testing for SAR target back shots.

MSTAR SAR images of depression angle $17°$ are used in this paper. The $17°$ depression angle dataset is separated into three sub groups. Figure 2 presents examples of MSTAR SAR images used in this paper. All SAR chips are cropped into $64×64$ pixels from $128×128$ pixels in order to exclude a contribution of background scattering to classification performance.

For RLS training phase, five vehicle classes are used. These five vehicle classes are selected to have variation in vehicle type in order to let RLS be generalized. Each class contains SAR images observed from entire azimuth angles. Table 1 describes the detail of selected classes.

Table 1. Data set used to train Encoder and RLS

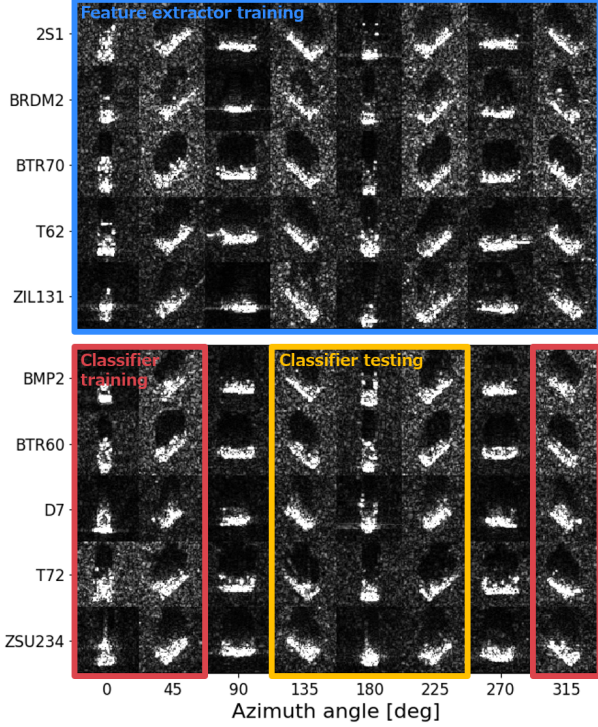| Class	Azimuth coverage	Size
2S1	$0° \sim 360°$	299
BRDM2	$0° \sim 360°$	298
BTR70	$0° \sim 360°$	233
T62	$0° \sim 360°$	299
ZIL131	$0° \sim 360°$	299
Total	–	1428

Fig. 2. MSTAR SAR sample images of depression angle of 17°. Images taken from eight different azimuth angles are shown. Five vehicle classes shown in the top panel (in the blue box) are used to train our auto-encoder. The other five vehicles are used in classification training (in the red box) and testing (in the yellow box).

The other five vehicle classes are used for a performance evaluation of RLS in classification with and without data augmentation in RLS. Those five vehicles are further separated into two groups by their azimuth angles for classifier training

Table 2. Data set used to classifier training and testing

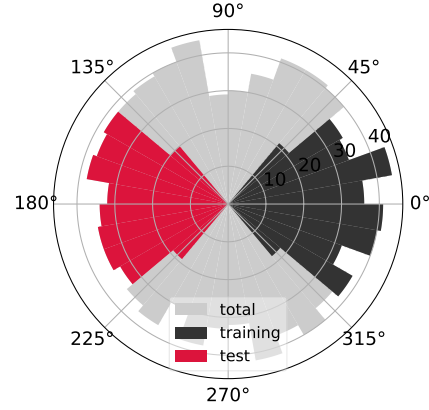|	Class	Azimuth coverage	Size
	BMP2-C21	$-45°(315°) \sim 45°$	59
	BTR60	$-45°(315°) \sim 45°$	66
Train	D7	$-45°(315°) \sim 45°$	77
	T72-SN132	$-45°(315°) \sim 45°$	59
	ZSU234	$-45°(315°) \sim 45°$	78
	Total	—	339
	BMP2-C21	$135° \sim 225°$	58
	BTR60	$135° \sim 225°$	57
Test	D7	$135° \sim 225°$	71
	T72-SN132	$135° \sim 225°$	57
	ZSU234	$135° \sim 225°$	78
	Total	—	321

Fig. 3. Azimuthal distribution of dataset used in training (dark gray area) and testing (red area) classifier. Light gray area indicates a distribution of all five vehicle images listed in Tab. 2.

and testing. Images within an azimuth angle range between $-45°$ and $45°$ are used in classifier training. Images within an azimuth angle range between $135°$ and $225°$ are used in classifier testing. Table 2 summarizes the details of data sets used in classifier training and testing. Figure 3 illustrates the azimuthal distribution of training and testing dataset.

3.2. Baseline

A CNN architecture modified from Wilmanski et al. [3] is selected as a baseline to compare with the proposed method. The differences between the baseline and the original Wilmanski's network are a size of input image and kernel sizes of three convolutional layers. The original Wilmanski's CNN uses 128×128 MSTAR chips as input images. We modified convolutional layers to fit with the input size of 64×64 pixels.

3.3. Results

Figure 4 shows classification accuracies of proposed approaches with and without data augmentation in RLS. The accuracies are averaged over 20 calculations with different initial weights. Errors are given as 2 standard deviations (2σ). Figure 4 also plots an accuracy of the baseline CNN-based classifier. The proposed RLS-based classifier with data augmentation achieves the highest accuracy of 70.2 % ($\pm$4.04 %). The second highest accuracy of 56.2 % ($\pm$3.51 %) is for the proposed RLS-based classifier without data augmentation, while the accuracy of the baseline is 38.7 % ($\pm$1.03 %). The proposed method with augmentation outperforms the others. This is because that RLS has successfully obtained an azimuth rotation invariant feature space.
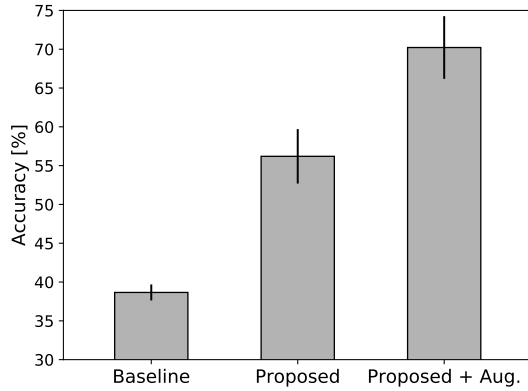
Fig. 4. Mean accuracy for MSTAR five class classification over 20 calculations and its 2σ value as an error. The baseline CNN (left), the proposed methods w/o (middle) and w/ (right) data augmentation in RLS are presented here.

Fig. 5. A confusion matrix of the classification result obtained by the proposed method with data augmentation in RLS. Color shows a percentage value.

Figure 5 is a confusion matrix of the classification result obtained by the proposed method with augmentation. The most confused class is BMP2. Especially BMP2 and BTR60 misrecognize each other. This is due to similarity in **Z**, which shows there is still scope for improvements in learning RLS.

4. CONCLUSIONS

This paper has proposed RLS for an azimuth invariant target recognition in SAR images using an autoencoder architecture and a cyclic permutation matrix. In RLS, rolling of latent features corresponds to 3D rotation of an object. Latent features of an arbitrary view can be approximated by changing the cyclic permutation matrices, which enables us to augment data from limited viewing. Evaluations have been performed using MSTAR images. In the evaluations, we trained an RLS-based classifiers with and without augmentation and a conventional classifier with target front shots, then compared thier accuracies over untrained target back shots. It is confirmed that the RLS-based classifiers outperformed the conventional classifier. The RLS-based classifier with augmentation achieved the highest accuracy of 70.2 %, which improves an accuracy by 30 % compared to the conventional classifier.

5. REFERENCES

[1] 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.

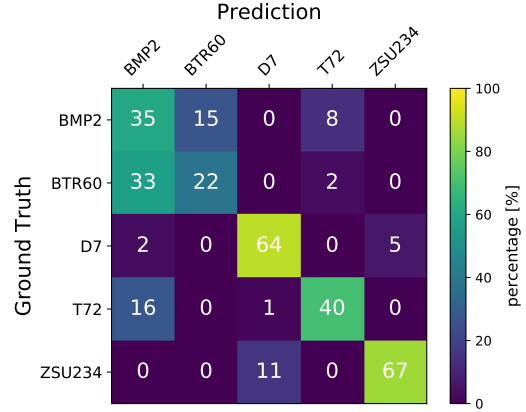[2] Chee Seng Chan John E. Ball, Derek T. Anderson, "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community," *Journal of Applied Remote Sensing*, vol. 11, pp. 11 – 11 – 54, 2017.

[3] Michael Wilmanski, Chris Kreucher, and Jim Lauer, "Modern approaches in deep learning for sar atr," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2016, pp. 98430N–98430N.

[4] Timothy D Ross, Steven W Worrell, Vincent J Velten, John C Mossing, and Michael Lee Bryant, "Standard sar atr evaluation experiments using the mstar public release data set," in *Algorithms for Synthetic Aperture Radar Imagery V*. International Society for Optics and Photonics, 1998, vol. 3370, pp. 566–574.

[5] Qian Song, Feng Xu, and Ya-Qiu Jin, "Deep sar image generative neural network and auto-construction of target feature space," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2017.

[6] C. Avolio, M. . M. Armenta, A. J. Lucena, M. J. F. Surez, P. V. Martn-Mateo, F. L. Gonzlez, B. L. Verdoy, A. Bucarelli, and M. Costantini, "Automatic recognition of targets on very high resolution sar images," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2017, pp. 2271–2274.

[7] Jimei Yang, Scott E Reed, Ming-Hsuan Yang, and Honglak Lee, "Weakly-supervised disentangling with recurrent transformations for 3d view synthesis," in *Advances in Neural Information Processing Systems*, 2015, pp. 1099–1107.

[8] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.