Classification of Large-Scale High-Resolution SAR Images with Deep Transfer Learning

Zhongling Huang, Corneliu Octavian Dumitru, Zongxu Pan, *Member, IEEE* Bin Lei,
and Mihai Datcu, *Fellow, IEEE*

Abstract—The classification of large-scale high-resolution SAR land cover images acquired by satellites is a challenging task, facing several difficulties such as semantic annotation with expertise, changing data characteristics due to varying imaging parameters or regional target area differences, and complex scattering mechanisms being different from optical imaging. Given a large-scale SAR land cover dataset collected from TerraSAR-X images with a hierarchical three-level annotation of 150 categories and comprising more than 100,000 patches, three main challenges in automatically interpreting SAR images of highly imbalanced classes, geographic diversity, and label noise are addressed. In this letter, a deep transfer learning method is proposed based on a similarly annotated optical land cover dataset (NWPU-RESISC45). Besides, a top-2 smooth loss function with cost-sensitive parameters was introduced to tackle the label noise and imbalanced classes' problems. The proposed method shows high efficiency in transferring information from a similarly annotated remote sensing dataset, a robust performance on highly imbalanced classes, and is alleviating the over-fitting problem caused by label noise. What's more, the learned deep model has a good generalization for other SAR-specific tasks, such as MSTAR target recognition with a state-of-the-art classification accuracy of 99.46%.

Index Terms—TerraSAR-X, high-resolution SAR images, land cover classification, transfer learning, label noise

I. INTRODUCTION

SYNTHETIC Aperture Radar (SAR) land use and land cover (LULC) classification is an important step in SAR image interpretation. Recognizing different land covers in SAR images requires an expertise in interpreting the complex scattering characteristics under various circumstances such as different wavelengths, incidence angles and target areas which makes the annotation of large-scale SAR datasets time consuming. With the development of deep learning algorithms, hierarchical features can be extracted automatically. Unsupervised learning approaches, such as deep belief networks (DBNs) [1], and auto-encoders (AEs) [2], [3], are widely applied to SAR images for classification. With more detailed

texture information, high-resolution (HR) PolSAR images are used to learn spatial and polarimetric features jointly via deep convolutional neural networks (DCNNs) for supervised classification which achieves state-of-the-art performance [4].

However, many SAR LULC classification algorithms mainly focus on specific tasks which may not generalize well on other SAR tasks, for example, classifying vegetation and agricultural land covers in farmland areas [5], or training an urban-specific network with two scene images in urban areas [6]. Although a big volume of SAR data all over the world is acquired every day, it is challenging to automatically interpret land covers in SAR images with such diversity and to obtain a well-generalized model for SAR image understanding.

In this letter, a global dataset of multi-level annotated HR SAR images [7] has been experimented. These SAR images vary in geographic locations, incidence angles, radiometric calibration factors, and ascending or descending satellite orbit branches. With the difficulties of highly imbalanced categories, geographic diversity and noisy annotations, our contributions are as follows:

(1) A very deep residual network with transitive transfer learning method from natural images to remote sensing and then to SAR data is used to understand a large amount of land cover SAR images from all over the world.

(2) A cost-sensitive top-2 smooth combined loss function with cross-entropy loss is proposed to alleviate the imbalanced problem and model bias to mislabeled training samples.

(3) The proposed network is proved to have a good generalization on other SAR datasets and tasks.

In the following, we will firstly introduce the TerraSAR-X (TSX) annotated land cover dataset in Section II. Our transfer learning based method, as well as the experimental results will be presented in Section III and Section IV, followed by a conclusion in Section V.

II. DATASETS

The HR SAR annotated land cover dataset [7] exploited in this paper has been collected from the X-band TSX instrument [8]; we used horizontally polarized (HH), multi-look ground range detected (MGD) products which were radiometrically enhanced. The selected SAR images were taken in High Resolution Spotlight mode with a resolution of 2.9 m, acquired with an incidence angle between 20° and 50°, in both descending and ascending pass directions.

This annotated semantic dataset covers urban and non-urban areas from all over the world, grouped into 46 collections in

TerraSAR-X	1-Settlements	2-Industrial areas	*3-Military facilities*	4-Public transportations	5-Agricultural land	6-Vegetation	7-Bare ground	8-Water bodies	*9-Unclassified*
Number	31696	4027	*285*	6442	7804	6671	4194	9847	*14818*
Example									
NWPU-RESISC45	baseball diamond(123), basketball court(123), church(121), commercial area(122), dense residential area(114), ground track field(123), medium residential(113), mobile home park(11), palace(12), parking lot(131), sparse residential(111), stadium(14), tennis court(123)	industrial areas(2), storage tank(214), thermal power station(21),	None	airplane(421), airport(41), bridge(45), freeway(432), harbor(46), intersection(43), overpass(32), railway(44), railway station(44), roundabout(434), runway(431), ship(472)	circular farmland(5), rectangular farmland(5), terrace(5)	chaparral, forest(61,62,63,64)	desert(74), mountain(77), meadow(71)	beach(86), lake(82), river(81), sea ice(84)	snow berg, cloud, wetland, golf course, island
Example			None						

Fig. 1. Nine kinds of land cover of our annotated level-1 TSX dataset with imbalanced classes. Statistically, 40 categories in NWPU-RESISC45 are semantically similar to the TSX dataset. The number in brackets indicates the assigned label of TSX, which can be indexed in TABLE A.I., Appendix [7]. Due to the small number of patches in "Military facilities" and indistinct label "Unclassified", only seven classes are applied in our experiments.

Fig. 2. The high intra-class diversity of *"low-density residential areas"* in two different cities.

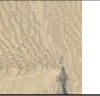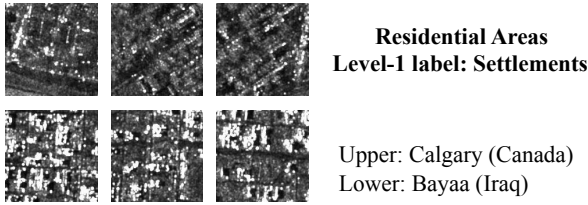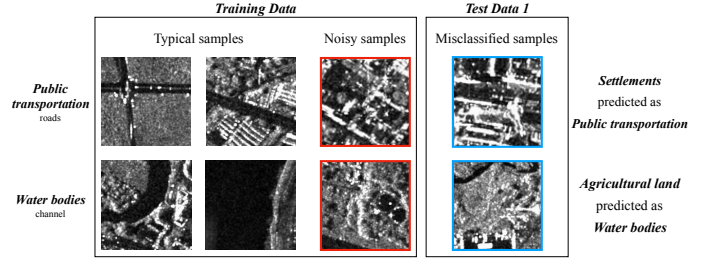Fig. 3. Examples of label noise and model bias due to the noisy labels.

total. With a pixel spacing of 1.25 m, each SAR image was tiled into non-overlapping patches with a size of $160{\times}160$ pixels, resulting in $200{\times}200$ m^2 on the ground so that a patch covered a typical object on the ground. A single label was assigned to each patch based on the dominating class. With three-level hierarchical annotation, only level-1 labels are used in this paper.

Automatically interpreting the worldwide large-scale SAR dataset is challenging due to the following reasons:

1) Highly imbalanced classes: As shown in Fig. 1, the imbalanced classes result in a big difference of data volume for each category in level-1 annotation case, mainly due to the fact that the dataset focuses on urban areas. This imbalance problem appears more serious in level-2 and level-3 annotation.

2) Geographic diversity: Collected world-widely, different architectural styles of cities bring out a large geographic diversity of images. The typical regional characteristics result in high intra-class diversity and inter-class similarity as shown in Fig. 2. Besides, the acquisition period of these images ranges from 2007 to 2012. During this period of time, some non-constant parameters such as the radiometric calibration factor, will result in different characteristics of the SAR images.

3) Label noise: By covering an area of $200{\times}200$ m^2 on the ground, sometimes a patch contains more than one land cover

type. Even though the annotation was based on the dominating class, there still existed some incomplete annotations or some semantic biases [9] which may lead the trained model to over-fitting when using these misleading samples. Fig. 3 gives an example of label noise in our dataset.

Due to the low number of only 285 patches from two collections of *"Military facilities"* and the unlabeled patches of *"Unclassified"*, we only used the remaining seven land cover classes in this paper.

Another remote sensing LULC dataset used in this paper is NWPU-RESISC45 [10], which has a land cover annotation similar to the seven classes of TSX dataset used in our experiment, shown in Fig. 1, containing 31,500 images covering 45 scene classes.

III. PROPOSED METHOD

Facing the difficulties mentioned above, we propose a deep transfer learning based method in this part to tackle the SAR land cover classification issue.

A. Transfer Learning with Deep Residual Networks

ResNet-18 [11] shown in Fig. 4, containing four residual blocks repeated twice, with two convolution layers in each
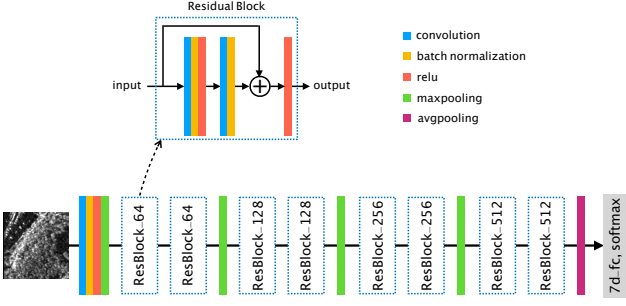
Fig. 4. Residual network architecture applied in our experiments.

block, followed by an average pooling layer and a fully-connected layer related to the class number of the task, was applied in our experiments. Such a very deep well-trained neural network (with a data volume of about 43.7 MB) has a great ability in extracting hierarchical features but is difficult to train from scratch. Transfer learning provides a better initialization to the parameters in deep layers, making the optimization easier and speeding up the training process. The ImageNet pre-trained model is popular as a transferred source. However, the large difference lying between SAR land cover patches and natural object images may harm the transferability, making it necessary to find a bridge to narrow the gap [12].

Rather than using the ImageNet pre-trained model as most remote sensing applications do, we proposed a transitive transfer learning method with the help of the NWPU-RESISC45 dataset [10] to build a bridge linking the natural and SAR land cover images. ResNet-18 pre-trained on ImageNet [11] was applied as our initial model and then the convolution layers were fine-tuned with the NWPU-RESISC45 dataset, obtaining the remote sensing enhanced model as the transfer source for our applications.

B. Imbalanced Classification

When we trained a model with a dataset of highly imbalanced classes, we applied some techniques such as sampling and cost-sensitive learning [13]. For evaluation, we considered some metrics beyond accuracy, like precision, recall and F1-score.

Each time when we input a batch of samples to the network, over-sampling of the minority classes and under-sampling of the majority classes were applied to ensure a balanced distribution in that batch. Moreover, the training samples were weighted according to the corresponding proportions of the dataset in order to penalize the misclassification of the minority classes more heavily than for the majority classes.

C. Top-2 Smooth Loss Function

Here, we introduced the top-2 smooth loss function [14] to reduce the label noise in our dataset. With the classification of seven types of land cover in level-1, we mainly focused on the top-2 predictions $\bar{\mathbf{y}} \in \mathbb{R}^2$. We define the label space $\mathcal{Y} = \{1, 2, ..., n\}$ and the ground truth label $y \in \mathcal{Y}$, the top-2 predictions $\bar{\mathbf{y}} \in \mathcal{Y}^{(2)}$. The subset of $\mathcal{Y}^{(2)}$ where the ground

truth y is included, is written as $\mathcal{Y}_y^{(2)}$. The top-2 smooth loss function with temperature can be defined as:

$$
\begin{aligned}
L_{2,\tau}(\mathbf{s}, y) = & \tau \log[\sum_{\bar{\mathbf{y}} \in \mathcal{Y}^{(2)}} \exp(\frac{1}{\tau}(\Delta(\bar{\mathbf{y}}, y) + \frac{1}{2}\sum_{j \in \bar{\mathbf{y}}} s_j))] \\
& - \tau \log[\sum_{\bar{\mathbf{y}} \in \mathcal{Y}_y^{(2)}} \exp(\frac{1}{2\tau}\sum_{j \in \bar{\mathbf{y}}} s_j)].
\end{aligned}
\tag{1}
$$

where $\mathbf{s} \in \mathbb{R}^n$ denotes a vector of scores of n labels, τ the temperature parameter and $\Delta(\bar{\mathbf{y}}, y) \triangleq \mathbb{1}(y \notin \bar{\mathbf{y}})$ denotes the margin. The top-2 smooth loss function takes two most likely predictions into consideration, making it possible to prevent over-fitting with label noise when learning only with the cross-entropy loss of a single label.

In the proposed method, the top-2 smooth loss was combined with the cross-entropy loss with a trade-off λ rather than totally replacing it as [14] did, because $k = 2$ is not much smaller than $n = 7$ in our case. The top-2 loss was used as a constraint to alleviate model bias towards the single misleading label when there was more than one kind of land cover on the ground. Finally, adding an L2 regularization term and the loss weight w_y, the combined loss function can be written as:

$$
\begin{aligned}
L_{2,\tau,\mu,\lambda}(\mathbf{s}, y) = & (1-\lambda)(-\sum_i s_i \log \frac{\exp(s_i)}{\sum_{j=1}^{C} \exp(s_j)}) \\
& + w_y \lambda L_{2,\tau}(\mathbf{s}, y) + \mu \sum_i \|\beta_i\|^2,
\end{aligned}
\tag{2}
$$

where β_i denotes the weights in each layer, and λ, τ and μ are three hyper-parameters which control the trade-off between the top-2 and cross-entropy losses, temperature tuning, and the penalty of the regularization term. The cost weight w_y related to the class y is given by:

$$
w_y = \frac{1}{n-1}(1 - \frac{N_y}{\sum_{i=1}^{n} N_i}),
\tag{3}
$$

where n denotes the class number and N_i denotes the number of samples in the ith class.

In addition, we use bagging [15], training five sub-models with re-sampling data and voting the predictions of all sub-models in order to make the trained model more robust and general.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

We firstly introduce our experimental settings in Section IV-A. In Section IV-B and IV-C we analyze the effectiveness of different transfer sources and the combined top-2 smooth loss. Finally, we use three examples to demonstrate the good generalization of the proposed SAR specific deep model in IV-D.

A. Experimental Settings

On the one hand, we randomly chose 200 patches in each category from collection 01 to 36 as *Test Data 1* and the remaining as *Training Data*. On the other hand, patches of the
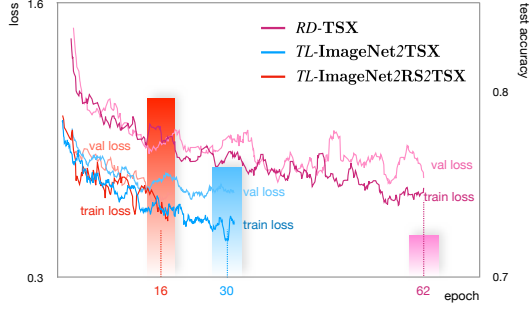
Fig. 5. The lines denote the loss values on training and validation data before early stopping, and the bars denote the validation accuracy of three experimental settings.
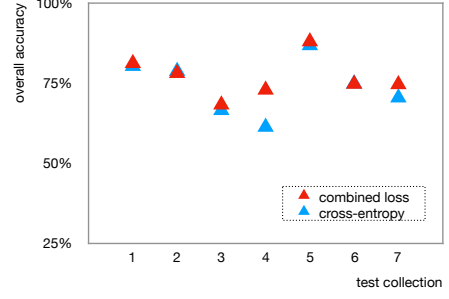

Fig. 6. The classification results on seven new collections of *Test Data 2*, learning with cross-entropy, and the combined loss shown as blue/red dots, respectively.

full scene images from another seven collections were used as *Test Data 2* which had a different distribution with *Training Data* and *Test Data 1*. The details of training and test data settings are presented in Table I.

The TSX detected images are represented by 16 bits spanning a large dynamic range. In our experiments, the 16 bit data after log transform was fed into the network to prevent the gradients from vanishing during training. When transferring the pre-trained model, only the last fully-connected layer was adapted to fit the new task.

The hyper-parameters of λ, τ and μ which control the trade-off between the top-2 and cross-entropy losses, the temperature tuning, and the penalty of the regularization term, were empirically set to 0.2, 1.0 and 0.25, respectively. Adam optimization algorithm [16] was applied and the initial learning rate was set to 10^{-5} when using the top-2 loss, and to 10^{-4} otherwise.

B. The Transitive Transferring Model

We compared the results of overall classification accuracy in *Test Data 1* and the average stopping time of training a ResNet-18 network for the following cases to demonstrate the effectiveness of transferring from proper source data to SAR land cover classification:

*1) RD-**TSX***: By randomly initializing all layers of ResNet-18, the network was trained with the TSX dataset from scratch.

*2) TL-**ImageNet2TSX***: By using the pre-trained model of ImageNet, the network was fine-tuned with the TSX dataset with only the last layer related to the classes trained from scratch.

*3) TL-**ImageNet2RS2TSX***: The hierarchical convolution layers of ImageNet pre-trained model were fine-tuned with NWPU-RESISC45 optical remote sensing images for 45 types of land cover classification and then transferred to the TSX dataset.

Fig. 5 illustrates that training such a very deep network like ResNet-18 with the TSX dataset from scratch is difficult and time-consuming. Consequently, it is necessary to find a better initialization of millions of parameters. Although the ImageNet pre-trained model has been widely used for remote sensing applications, it is controversial to directly transfer the natural images to SAR data due to the huge differences between them.

Here, we took advantage of the remote sensing images with similar categories to our SAR land cover dataset to narrow the gap between natural images and SAR data. We see a performance improvement both in test accuracy and training time.

C. Combined with Top-2 Loss

Besides applying some strategies such as Dropout and L2 regularization to avoid over-fitting during training, we found that label noise may play an important role with a negative influence on the trained model. To address this prediction bias issue, we introduced a top-2 smooth loss function to constrain the cross-entropy from introducing a bias to the noisy labels. We use different metrics of top-1, top-2 overall accuracy, and also F1-score to evaluate the results which is defined by $F_1 = 2/(p^{-1} + r^{-1})$, where $p = \frac{TP}{TP+FP}$, $r = \frac{TP}{TP+FN}$, and TP, FP, FN denote the true positive, false positive and false negative samples, respectively.

The top-2 smooth loss focuses on two most likely land cover classes, thus penalizing the predictions only when neither of the top-2 likely labels are the ground truth. Table II shows the training time of using different loss functions and the test results of F1-score, top-1, and top-2 accuracies both on *Test Data 1* and *2*. The top-2 accuracy is the percentage of time that the classifier gives the correct class in the top-2 probabilities. We found that training with top-2 smooth loss only would lead to a fast convergence of the deep network to speed up training. However, the top-2 smooth loss was too relaxing for the optimization resulting in much smaller gradients after a period of time and weights being difficult to update. As a consequence, we combined the top-2 smooth loss with the cross-entropy as a constraint to achieve a faster convergence with nine epochs compared with 16 epochs of learning with cross-entropy. With the improvement in F1-score and accuracy, together with the obvious speed-up in training time, the combined loss performed considerably better.

D. Generalization to Other Data and Tasks

We could demonstrate the good generalization of our SAR deep ResNet model by testing two tasks: (1) land cover classification on new collections of *Test Data 2*; (2) MSTAR [17] target recognition without data augmentation, as shown in Table III.

TABLE I
TRAINING AND TEST DATASET

Category	Settlements	Industrial areas	Public transportations	Agricultural land	Natural vegetation	Bare ground	Water bodies	Total
Training Data	24,930	2,979	4,485	6,029	4,911	2,240	6,826	**52,400**
Test Data 1	200	200	200	200	200	200	200	**1,400**
Test Data 2	3,155	563	928	925	1,383	1,319	2,133	**10,406**

TABLE II
CLASSIFICATION RESULT ON TEST DATA 1 AND TEST DATA 2

	Loss / Eval	ce loss	top-2 loss	combined loss
Test Data 1	F1 score	0.8318	0.7954	**0.8327**
	top-1 acc	83.14%	79.5%	**83.21%**
	top-2 acc	**95.36%**	93.29%	95.00%
Test Data 2	F1 score	0.709	0.684	**0.72**
	top-1 acc	74.09%	73.24%	**76.78%**
	top-2 acc	91.26%	91.25%	**92.62%**
training time		16 epochs	**6 epochs**	9 epochs

TABLE III
TRANSFER TO NEW TASKS ON SAR DATA

Dataset	Task	Classes	Transfer Result
Test Data 2	scene classification	7	76.78%
MSTAR [17]	target recognition	10	99.46%

Fig. 6 shows the test results of each new collection in *Test Data 2* which has a different distribution with the training data. In most new collections, the performance of the combined loss was better than cross-entropy (2.69% higher on average) which demonstrates that the trained model with combined loss reduces over-fitting due to label noise and improves generalization on new dataset.

Furthermore, when we transferred the proposed model to an MSTAR [17] 10-class recognition task, we achieved a state-of-the-art overall accuracy of 99.46%, compared with 98.02% using an ImageNet pre-trained model, and 99.09% by other CNN-based methods [18].

V. CONCLUSION

In this letter, a deep transfer learning method combining the top-2 smooth loss was proposed to solve the land cover classification problem for a large-scale HR TSX dataset, with highly imbalanced classes, geographic diversity and noisy labels. Rather than applying the ImageNet pre-trained model of ResNet-18 to SAR images directly, we used an optical remote sensing land cover dataset to narrow the gap between SAR and natural images which results in a significant improvement on feature transferability. The proposed combined loss function is successful in reducing model bias to noisy labels and our proposed deep ResNet model shows a good generalization which allows it to be exploited for other SAR-specific tasks.

ACKNOWLEDGMENT

REFERENCES

[1] Q. Lv, Y. Dou, X. Niu, J. Xu, J. Xu, and F. Xia, "Urban land use and land cover classification using remotely sensed SAR data through deep belief networks," *Journal of Sensors*, vol. 2015, 2015.

[2] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sept 2016.

[3] Z. Huang, Z. 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, 2017.

[4] W. Wu, H. Li, L. Zhang, X. Li, and H. Guo, "High-resolution Pol-SAR scene classification with pretrained deep convnets and manifold polarimetric parameters," *IEEE Trans. on Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6159–6168, Oct 2018.

[5] C. Hütt, W. Koppe, Y. Miao, and G. Bareth, "Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images," *Remote Sens.*, vol. 8, no. 8, 2016.

[6] J. Zhao, W. Guo, B. Liu, Z. Zhang, W. Yu, and S. Cui, "Preliminary exploration of SAR image land cover classification with noisy labels," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul 2017, pp. 3274–3277.

[7] C. O. Dumitru, G. Schwarz, and M. Datcu, "Land cover semantic annotation derived from high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2215–2232, 2016.

[8] "TerraSAR-X basic product specification document, issue 1.9," https://sss.terrasar-x.dlr.de/docs/TX-GS-DD-3302.pdf, accessed December 10, 2018.

[9] A. M. M. de Oca, R. Bahmanyar, N. Nistor, and M. Datcu, "Earth observation image semantic bias: A collaborative user annotation approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2462–2477, Jun 2017.

[10] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct 2017.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[12] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in sar target recognition based on deep cnns," *IEEE Trans. on Geosci. Remote Sens.*, pp. 1–13, 2019.

[13] M. Kukar, I. Kononenko *et al.*, "Cost-sensitive learning with neural networks." in *ECAI*, 1998, pp. 445–449.

[14] L. Berrada, A. Zisserman, and M. P. Kumar, "Smooth loss functions for deep top-k classification," in *Proc. IEEE International Conference on Learning Representations (ICLR)*, 2018.

[15] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Mach. Learn.*, vol. 36, no. 1, pp. 105–139, Jul 1999.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[17] "The air force moving and stationary target recognition database," https://www.sdms.afrl.af.mil/datasets/mstar/, accessed 3 February, 2016.

[18] 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, 2016.