# META METRIC LEARNING FOR HIGHLY IMBALANCED AERIAL SCENE CLASSIFICATION

*Jian Guan*[1]*, Jiabei Liu*[1]*, Jianguo Sun*[1*]*, Pengming Feng*[2]*, Tong Shuai*[3]*, and Wenwu Wang*[4]

[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China
[2]State Key Laboratory of Space-Ground Integrated Information Technology, Beijing, 100095, China
[3]CETC Key Laboratory of Aerospace Information Applications, Shijiazhuang, 050081, China
[4]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

## ABSTRACT

Class imbalance is an important factor that affects the performance of deep learning models used for remote sensing scene classification. In this paper, we propose a random fine-tuning meta metric learning model (RF-MML) to address this problem. Derived from episodic training in meta metric learning, a novel strategy is proposed to train the model, which consists of two phases, i.e., random episodic training and all classes fine-tuning. By introducing randomness into the episodic training and integrating it with fine-tuning for all classes, the few-shot meta-learning paradigm can be successfully applied to class imbalanced data to improve the classification performance. Experiments are conducted to demonstrate the effectiveness of the proposed model on class imbalanced datasets, and the results show the superiority of our model, as compared with other state-of-the-art methods.

***Index Terms*—** Remote sensing, scene classification, class imbalance, meta-learning, metric learning

## 1. INTRODUCTION

Scene image analysis is an important research topic in the field of remote sensing. Recently, as the amount of accessible remote sensing image data increases substantially, advanced scene image analysis, e.g., scene classification, has attracted increasing research interests. Remote sensing scene classification aims to mark aerial images automatically with specific semantic categories, which is a fundamental problem in understanding high-resolution remote sensing images [1].

Recently, deep learning methods have been applied to address scene classification problems in remote sensing with promising performance [2, 3]. The training datasets used in these methods are generated by human annotations, hence, in theory, a balanced distribution can be enforced for each class [4, 5, 6]. In practice, however, due to the difficulty in data acquisition, scene images are usually unbalanced across classes. Using such data can lead to poor performance on the minority classes [7].

Despite of its significance, little research has been conducted to address this problem. However, several existing ideas for other applications can be related to this problem. For example, the re-sampling based methods [7, 8] aim to ensure equal distribution of each class, by down-sampling the majority classes or over-sampling the minority classes. In the re-weighting based method [9], additional weights are applied in the loss function for each class with inverse class frequency. The re-sampling based methods can lose information on the majority classes or quickly over-fit the minority classes, whereas the re-weighting based methods can alleviate such problem only to some extent.

The impact of class imbalance on the result using convolutional neural networks (CNNs) has been studied in [10], where the number of training samples is adjusted according to scene complexity which can be limited when only few samples are available in minority classes. In the work [11], a competence estimation and selection method is designed for scenes with multi-source heterogeneous data, and to improve the performance of the data fusion system with the presence of class imbalance. However, the effectiveness of those methods is still limited for highly imbalanced data [12].

In this paper, we propose a random fine-tuning meta metric learning model (RF-MML) to address the class imbalanced problem for aerial scene classification. Meta metric learning [13, 14] is originally proposed to solve the few-shot learning problem, which learns from abundant training examples in the base classes to recognize novel classes with a limited amount of labeled examples. Inspired by this, our RF-MML can learn generalized knowledge from data-rich majority classes, and transfer it to data-poor minority classes. However, directly employing this paradigm to the class imbalanced classification will encounter catastrophic forgetting problem (i.e., forgetting the initial categories on which it was trained) [15], thus leading to poor performance on the majority classes. Hence, in our work, a new training procedure consists of two phases is introduced to improve the meta model's

[*] Corresponding author

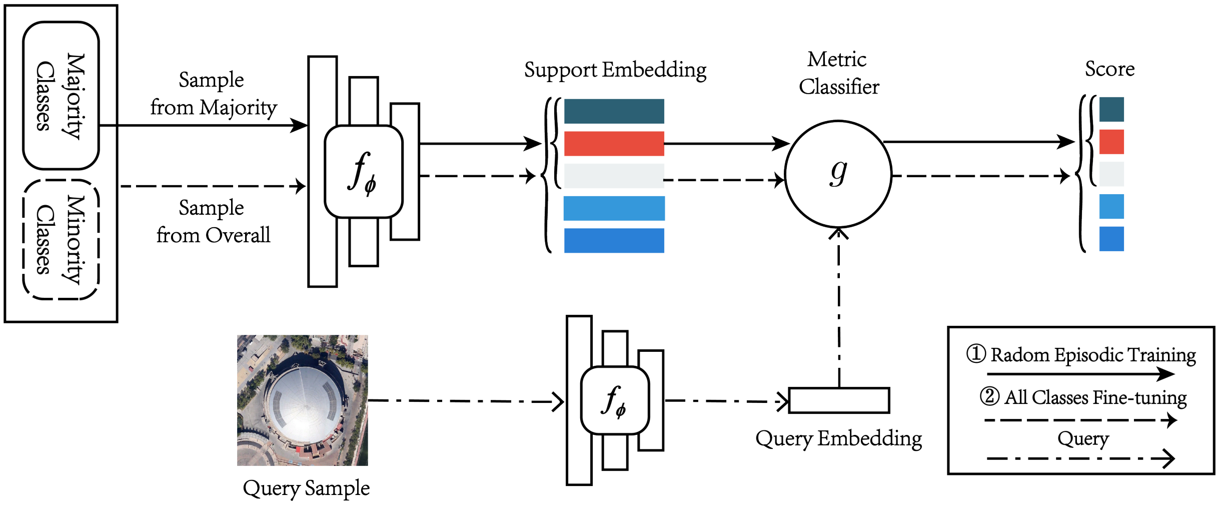

**Fig. 1**: Framework of the proposed metric based meta-learning model, where the training procedure includes two phases: random episodic training and all classes fine-tuning. Here, query samples are embedded and classified in the same way in both phases. $f_\phi$ is an embedding model, and $g$ denotes a metric based linear classier.

performance. Firstly, random episodic training on majority classes is implemented to learn generalized knowledge across data distribution. Then, all classes are fine-tuned to further improve the performance for multi-category classification. A number of experiments are performed on class imbalanced datasets to demonstrate the improved performance of our proposed model as compared with other state-of-the-art methods.

The remainder of the paper is organized as follows: Section 2 presents the proposed method in details; Section 3 shows the experimental results; and Section 4 summarizes the paper and draws the conclusion.

## 2. METHODOLOGY

In this section, our proposed RF-MML is introduced, which consists of a feature embedding network $f_\phi$ and a metric based linear classifier $g$, where $\phi$ denotes the learnable parameters of the network. The overall framework and training procedure are given in Fig. 1. In this model, a novel meta-learning procedure is introduced for model training, which includes two phases: *random episodic training* (RET) and *all classes fine-tuning* (ACF). The details are given as follows.

### 2.1. Random Episodic Training

Motivated by episode-based meta-learning [13, 14], we introduce a novel random episodic training strategy for model training. Different from [13, 14], where the form of episodes is fixed along training, our RET can randomly generate different forms for each episode, hence can improve the generalization across different multi-class classification.

For aerial scene classification, assume a highly imbalanced training set $\mathcal{D}^{train} = \{(x_t, y_t)\}_{t=1}^{T}$, where $x_t$ denotes an aerial scene image and with its corresponding label $y_t$. The overall set of categories is denoted as $\mathcal{C}^{train}$. Here, we divide $\mathcal{C}^{train}$ into two subsets: $\mathcal{C}^{maj}$ and $\mathcal{C}^{min}$, where $\mathcal{C}^{maj}$ contains sufficient samples in each class, whereas $\mathcal{C}^{min}$ contains very limited samples in each class.

In RET, the model is trained by $K$-ways, $N$-shot classification tasks $\mathcal{T}_i$ on $\mathcal{C}^{maj}$, where $K$ is the number of classes, and $N$ denotes the number of training samples in each class. Rather than being fixed as in [13, 14], $K$ is generated from a range of $[2, |\mathcal{C}^{maj}|]$ and $N$ is generated from a range of $[1, N^{max})$, where $|\mathcal{C}^{maj}|$ is the number of elements in class set $\mathcal{C}^{maj}$, and $N^{max}$ denotes the number of samples available in minority classes. After $K$ and $N$ are generated, we randomly select $K$ classes from $\mathcal{C}^{maj}$ to construct the class set $\mathcal{C}^i$. Then, $N$ samples are selected from each class in $\mathcal{C}^i$. As a result, we can obtain a support set $\mathcal{D}_i^{support}$ with $N \times K$ samples, defined as

$$\mathcal{D}_i^{support} = \{(x_n, y_n) | n = 1, \cdots, N \times K, y_n \in \mathcal{C}^i\} \quad (1)$$

In addition, the query set $\mathcal{D}_i^{query}$ can be constructed by $M$ samples selected from $\mathcal{C}^i$, as:

$$\mathcal{D}_i^{query} = \{(x_n, y_n) | n = 1, \cdots, M, y_n \in \mathcal{C}^i\} \quad (2)$$

After that, both $\mathcal{D}_i^{support}$ and $\mathcal{D}_i^{query}$ are mapped into feature space by an embedding network $f_\phi$. Finally, the query embedding features are classified by a metric based linear classier $g$ according to support embeddings. The loss is calculated from the results of classification score, and $f_\phi$ is optimized by backpropagation. The details of the loss and optimization will be given in Sec.2.3.

### 2.2. All Classes Fine-Tuning

To further improve the performance for class imbalanced data, a novel training phase named all classes fine-tuning is integrated with the meta-learning paradigm. Episodic training is performed on all categories once RET is finished, as a result, the model is adapted to the specific multi-classes classification task without forgetting the initial majority classes and thus achieves better results.

Here in ACF, the embedding network $f_\phi$ is trained episodically by all categories in $\mathcal{C}^{train}$. More specifically, the embedding network $f_\phi$ is trained in $K$-ways $N$-shot classification tasks, where the $K$ is fixed to $|\mathcal{C}^{train}|$ and $N$ is randomly generated in the range of $[1, N^{max})$. As for the query set, we select the same number of samples from each category to form a balanced query set. The loss calculation and model optimization are the same as that in RET.

### 2.3. Linear Classifier and Loss Function

In our proposed model, the embedding network $f_\phi$ can incorporate any derivable metric based linear classifiers, such as nearest class prototype [14] and support vector machines (SVMs) [16]. Here, to demonstrate the validation of our model, we choose nearest class prototype as our linear classier $g$ for its easy computation, which can be performed by computing the distances between prototype representations of each class in metric space.

Nearest class prototype first computes a prototype $\mathbf{c}_k$ of each class $\mathcal{C}_k$ in the support set $\mathcal{D}_i^{support}$, as

$$\mathbf{c}_k = \frac{1}{|\mathcal{C}_k|} \sum_{(x_i, y_i) \in \mathcal{C}_k} f_\phi(x_i) \tag{3}$$

Then, the prediction score $p_\phi$ of each query sample on each category can be obtained by the following equation:

$$p_\phi(y = k|x) = \frac{\exp\left(-d\left(f_\phi(x), \mathbf{c}_k\right)\right)}{\sum_{k'} \exp\left(-d\left(f_\phi(x), \mathbf{c}_{k'}\right)\right)} \tag{4}$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance. The learning procedure is implemented by minimizing the negative log-probability $J(\phi)$ of the ground true class $k$ via backpropagation, where the loss $J(\phi)$ can be formulated as follows:

$$J(\phi) = -\log p_\phi(y = k|x) \tag{5}$$

When the linear classifier during training and testing are both nearest class *prototype*, the proposed model is named as **RF-MML-Proto**. Note that $g$ can be replaced by other linear classifiers, for comparison purpose, we also explore the performance of *SVMs* during testing. Thus, another model named **RF-MML-SVM** is introduced, which is trained by the nearest class prototype, but tested with SVMs.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset

We evaluate the proposed model on both AID [5] and NWPU-RESISC45 [6]. The AID has a number of 10,000 images within 30 classes, the numbers of sample images in each class varies from 220 up to 420, while the NWPU-RESISC45 has 45 classes with 700 samples in each class. For AID, we first randomly select 50% samples from each class as training set, and then select 100 samples from the rest for each class to construct a balanced testing set. The remaining part is used as the validation set. As for NWPU-RESISC45, in each class, 140 samples are selected as the training set, 400 samples as the testing set and the rest is used for validation.

To verify the effectiveness of our model for class imbalanced classification, the training set is divided into two subsets according the number of samples within each class: majority classes and minority classes. Here, each minority class contains only a small number of samples (i.e., $s = 5$ and $s = 10$), while the number of training samples in majority classes stays the same. We select 10 and 15 classes to construct the minority classes for AID and NWPU-RESISC45 respectively. Note that, the number of samples for minority classes is manually setup, so that we can obtain class imbalanced training sets.

### 3.2. Implementation Details

***Network Architecture*** We use Wide ResNet-50 (WResNet50) [17] pre-trained on ImageNet as the backbone of our network architecture. For our proposed model, all the layers before the last global average are used as the architecture of $f_\phi$, while for other baseline deep learning methods, we replace the last 1000 dimensional fully connected (FC) layer of WResNet50 by a $|\mathcal{C}^{train}|$ dimensional FC layer.

***Training Setup*** The model is trained with 80,000 episodes for RET and 2000 episodes for ACF, where Adam optimizer is employed with a learning rate of $5 \times 10^{-6}$ in RET, and $1 \times 10^{-6}$ in ACF. Other baseline deep learning methods are trained by Adam optimizer with a $5 \times 10^{-6}$ learning rate by 500 epochs, the learning rate is divided by 5 for every 200 epochs. We use the validation set to select the model with the best accuracy.

### 3.3. Experimental Results

***Performance Comparison*** We first conduct experiments to compare the performance of our proposed methods, i.e., RF-MML-Proto and RF-MML-SVM, with other state-of-the-art methods. Here, the baseline model (Plain) is WResNet50 [17] without employing any strategy for class imbalanced problem. The results are given in Table 1. As can be seen from Table 1, both RF-MML-Proto and RF-MML-SVM can improve the performance on minority classes. Although the recognition accuracy on majority classes has a moderate drop, the overall performance is improved. What's more, the fewer the number of training samples in minority classes, the more improvement of our proposed model can achieve.

To further show the effectiveness of our proposed model, the confusion matrix obtained by RF-MML-SVM on AID is

**Table 1**: Performance comparison in terms of classification accuracy. Each training set is divided into majority and minority classes according to the number of training samples. $s$ denotes the number of training samples in each class.

| Methods | AID: $\lvert\mathcal{C}^{maj}\rvert = 20, \lvert\mathcal{C}^{min}\rvert = 10$ | | | | | | NWPU-RESISC45: $\lvert\mathcal{C}^{maj}\rvert = 30, \lvert\mathcal{C}^{min}\rvert = 15$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Majority $s > 100$ | Minority $s = 10$ | Overall | Majority $s > 100$ | Minority $s = 5$ | Overall | Majority $s = 140$ | Minority $s = 10$ | Overall | Majority $s = 140$ | Minority $s = 5$ | Overall |
| Plain [17] | 94.75% | 48.60% | 79.37% | 94.85% | 37.00% | 75.57% | 89.46% | 50.75% | 76.56% | 90.96% | 30.38% | 70.77% |
| ReWeight [9] | 93.35% | 56.70% | 81.13% | 91.75% | 47.80% | 77.10% | 89.07% | 55.62% | 77.92% | 86.41% | 39.10% | 70.64% |
| ReSample [8] | 93.50% | 50.60% | 79.20% | 93.10% | 40.30% | 75.50% | 90.04% | 50.73% | 76.94% | 88.67% | 32.52% | 69.95% |
| RF-MML-Proto (Ours) | 92.15% | **63.20%** | **82.50%** | 90.80% | **54.50%** | **78.67%** | 87.52% | **61.49%** | **78.73%** | 86.13% | **50.24%** | **74.11%** |
| RF-MML-SVM (Ours) | 92.15% | **62.70%** | **82.33%** | 91.80% | **53.09%** | **78.99%** | 88.31% | **63.86%** | **80.80%** | 87.87% | **50.99%** | **75.51%** |

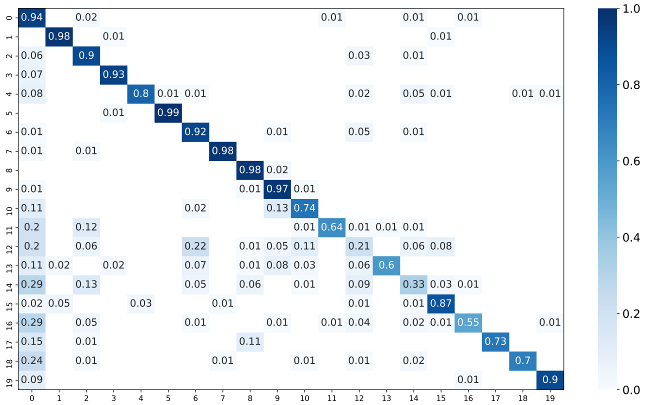

**Fig. 2**: Confusion matrix obtained by RF-MML-SVM on AID testing set. The range of minority class is between 10 and 19, which contain only 10 training samples in each class. For a clearer presentation, we have integrated the first 10 majority categories into class 0.

illustrated in Fig. 2. The recognition accuracy comparison for each class of RF-MML-SVM and ReWeight [9] is shown in Fig. 3, which details the classification accuracy improvement of RF-MML-SVM relative to ReWeight in each class. As shown in Fig. 2 and Fig. 3, although the improvement of the overall recognition accuracy is limited, the proposed model can achieve a higher recognition accuracy on the minority categories.

***Ablation Study*** To show the effectiveness of both RET and ACF, an ablation study is conducted on NWPU-RESISC45, where 15 minority classes with 10 samples in each class are manually created. The results are given in Table 2, here the baseline method denotes the model that is trained without RET and ACF, which is exactly same as ProtoNet in [14]. RF-MML-Proto/w/R denotes the proposed model without using RET, and RF-MML-Proto/w/F denotes the proposed model without using ACF. As can be seen from Table 2, both training strategies are effective, and the combination can further improve the model's performance and achieve the best classification result for class imbalanced data.

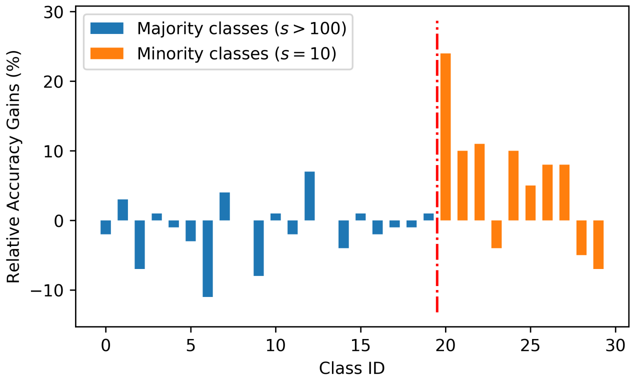

**Fig. 3**: Precision comparison for each class of the RF-MML-SVM and the ReWeight method. Where the Y-axis denotes per class classification accuracy improvement of RF-MML-SVM relative to the ReWeight, and the X-axis denotes the class index of each category.

**Table 2**: Ablation Study on NWPU-RESISC45

| Methods | Majority | Minority | Overall |
|---|---|---|---|
| Baseline(ProtoNet) [14] | 79.23% | 60.67% | 72.85% |
| RF-MML-Proto/w/R | 83.40% | 61.39% | 76.11% |
| RF-MML-Proto/w/F | 79.29% | 64.14% | 74.18% |
| RF-MML-Proto | 87.52% | 61.49% | 78.73% |

## 4. CONCLUSION

We have presented a metric based meta-learning model and a new training procedure to deal with the class imbalance in aerial scene classification. With a slight performance drop on the majority classes, our proposed model can greatly improve the recognition accuracy on the minority classes and finally improve the overall performance. Such improvement is approximately proportional to the degree of class imbalance. Both RET and ACF can improve the performance of meta-learning paradigm in class imbalanced classification.

## 5. REFERENCES

[1] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework," *GIScience & Remote Sensing*, vol. 54, no. 5, pp. 741–758, 2017.

[2] X. Bian, C. Chen, L. Tian, and Q. Du, "Fusing local and global features for high-resolution scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2889–2901, 2017.

[3] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[4] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14 680–14 707, 2015.

[5] G. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[6] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *arXiv preprint arXiv:1703.00121*, 2017.

[7] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 21, pp. 1263–1284, 2009.

[8] L. Shen, Z. Lin, and Q. Huang, "Relay backpropagation for effective learning of deep convolutional neural networks," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 467–482.

[9] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5375–5384.

[10] T. Shi, J. Wang, P. Wang, Q. Cai, and Y. Han, "The impact of imbalanced training datasets on cnn performance in typical remote scenes classification," *DEStech Transactions on Computer Science and Engineering*, no. pcmm, 2018.

[11] S. Sukhanov, C. Debes, and A. Zoubir, "Dynamic selection of classifiers for fusing imbalanced heterogeneous data," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 5361–5365.

[12] Y. Wang, D. Ramanan, and M. Hebert, "Learning to model the tail," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 7032–7042.

[13] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 3637–3645.

[14] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 4080–4090.

[15] S. Gidaris and N. Komodakis, "Dynamic few-shot visual learning without forgetting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4367–4375.

[16] K. Lee, S. Maji, A. Ravichandran, and S. Soatto, "Meta-learning with differentiable convex optimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 657–10 665.

[17] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.