



HAL
open science

RSMATCH: Semi-supervised Learning with Adaptive Category-Related Pseudo Labeling for Remote Sensing Scene Classification

Weiquan Lin, Jingjing Ma, Xu Tang, Xiangrong Zhang, Licheng Jiao

► **To cite this version:**

Weiquan Lin, Jingjing Ma, Xu Tang, Xiangrong Zhang, Licheng Jiao. RSMATCH: Semi-supervised Learning with Adaptive Category-Related Pseudo Labeling for Remote Sensing Scene Classification. 5th International Conference on Intelligence Science (ICIS), Oct 2022, Xi'an, China. pp.220-227, 10.1007/978-3-031-14903-0_24. hal-04666409

HAL Id: hal-04666409

<https://hal.science/hal-04666409v1>

Submitted on 1 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

RSMATCH: Semi-supervised Learning with Adaptive Category-related Pseudo Labeling for Remote Sensing Scene Classification*

WeiQuan Lin, Jingjing Ma, Xu Tang, Xiangrong Zhang, and Licheng Jiao

School of Artificial Intelligence, Xidian University, Shaanxi, 710071, China
3518076613@qq.com

Abstract. Remote sensing scene classification (RSSC) has become a hot and challenging research topic in recent years due to its wide applications. Due to the development of convolutional neural networks (CNN), the data-driven CNN-based methods have achieved expressive performance in RSSC. However, the lack of labeled remote sensing scene images in real applications make it difficult to further improve their performance of classification. To address this issue, we propose a novel adaptive category-related pseudo labeling (ACPL) strategy for semi-supervised scene classification. Specifically, ACPL flexibly adjusts thresholds for different classes at each time step to let pass informative unlabeled data and their pseudo labels according to the model's learning status. Meanwhile, our proposed ACPL does not introduce additional parameters or computation. We apply ACPL to FixMatch and construct our model RSMATCH. Experimental results on UCM data set have indicated that our proposed semi-supervised method RSMATCH is superior to its several counterparts for RSSC.

Keywords: Remote sensing scene classification (RSSC) · Adaptive category-related pseudo labeling (ACPL) · Semi-supervised method.

1 Introduction

Remote sensing image scene classification (RSSC) plays an important role in a wide range of applications, such as scene classification, ship detection and change detection [1–3]. As a basic image understanding work, it has attracted increasing attention. RSSC aims at automatically assign high-level semantic labels to local areas of remote sensing images for achieving scene-level classification. Therefore, the core of RSSC lies in obtaining the discriminative features of high-resolution remote sensing scenes.

Along with the development of deep learning [10], CNNs, trained on a large number of labeled data, show powerful feature learning ability. In RSSC community, CNNs-based methods gradually occupy the mainstream [4–6]. However, in

* This work was funded in part by the National Natural Science Foundation of China (No. 62171332) and the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University.

real applications, it is hard and time-consuming to collect and annotate enough samples of unseen categories for our learning system to retrain while limited labeled samples and abundant unlabeled samples are relatively easy to obtain. Thus, a new machine learning technology, semi-supervised learning (SSL) [11], has been a hot research topic in the last decade. SSL algorithms provide a way to explore the latent patterns from unlabeled examples and thus improve the learning performance given a small number of labeled samples.

According to loss function and model design, SSL can be divided into generative methods, consistency regularization methods, graph-based methods, pseudo-labeling methods and hybrid methods. In RS community, for generative methods, Guo et al. [7] propose SAGGAN. A gating unit, a SAG module, and an Inception V3 branch are introduced into the discriminative network to enhance the discriminant capability for facilitating semi-supervised classification. For pseudo-labeling methods, Zhang et al [8] introduced the center loss into a semi-supervised network (SSCL) by using the unlabeled samples to update centers. For hybrid methods, a pseudo labeling combined with consistency regularization is proposed [9] based on EfficientNet [12] (MSMatch) to achieve semi-supervised classification.

In this paper, based on the recent work, FixMatch [13], we upgrade it by replacing the fix threshold with an adaptive category-related strategy to obtain pseudo labels for unlabeled samples for semi-supervised RSSC. This operation is called ACPL, which takes the learning difficulties of different classes into account and thus achieves promising classification performance.

The rest of this paper is organized as follows, Section 2 introduces our proposed network in detail, Section 3 exhibits the experimental configurations and results, and conclusion are summarized in Section 4.

2 Proposed Method

2.1 Preliminary knowledge

Consistency regularization is a simple but effective technique in SSL [11]. With the introduction of pseudo labeling techniques, the consistency regularization is converted from ℓ -2 loss [14] to an entropy minimization process [15], which is more suitable for the classification task. The improved consistency loss with pseudo labeling can be formulated as:

$$\frac{1}{\mu B} \sum_{i=1}^{\mu B} \mathbb{1}(\max(p(y|\alpha(u_i))) > \tau) H(\hat{p}(y|\alpha(u_i)), p(y|\alpha(u_1))), \quad (1)$$

where $\mathbb{1}(\cdot)$ is the indicator function, B is the batch size of labeled data, μ is the ratio of unlabeled data to labeled data, α is a random data augmentation, u_i denotes a batch of unlabeled data in an iteration, p represents the output probability of the model, H is cross-entropy, $\hat{p}(y|\alpha(u_i))$ is the pseudo label and τ is the pre-defined threshold, which is used to mask out noisy unlabeled data that have low prediction confidence.

In FixMatch[13], weak augmentation operates on unlabeled data to generate artificial labels, which are then used as the target of strongly-augmented data. The unsupervised loss term based on consistency regularization between weakly-augmented and strongly-augmented data can be formulated as:

$$\frac{1}{\mu B} \sum_{i=1}^{\mu B} \mathbb{1}(\max(p(y|\alpha(u_i))) > \tau) H(\hat{p}(y|\alpha(u_i)), p(y|v(u_i))), \quad (2)$$

where v is a strong augmentation function compared with weak augmentation α .

2.2 Adaptive Class-related Pseudo Labeling

The framework of our proposed RSMATCH is exhibited in Fig. 1. The main idea of our method is to use ACPL algorithm to upgrade the fixed threshold in FixMatch to adaptive category-related one and thus learns the scenes in RS data set that are difficult to be recognized better.

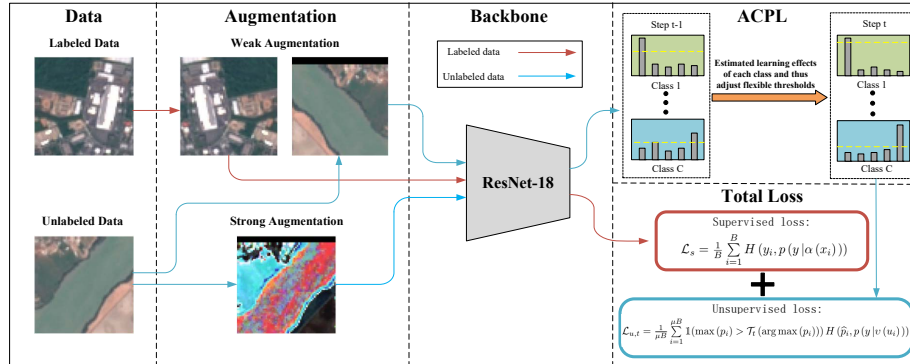


Fig. 1. Framework of our proposed RSMATCH.

The threshold τ in Eq. 2 is a pre-defined scale, which implicitly forces FixMatch treat the learning difficulty of each category in data set as same. However, different scenes always contain various volume of semantic information for model to learn. For example, in UCM data set¹, “dense residential”, “medium residential”, and “sparse residential” are difficult for model to recognize them because of the latent content in these scene images are similar.

To solve the limitation of FixMatch, we propose an adaptive class-related pseudo labeling strategy (ACPL) to adjust the thresholds according to the model’s learning status of each class. This operation allows the model to assign the pseudo labels to different classes at different time steps. Nevertheless, it does not introduce additional parameters.

¹ <http://vision.ucmerced.edu/datasets/landuse.html>

Specifically, based on the assumption [13] that when the threshold is high, the learning effect of a class can be reflected by the number of samples whose predictions fall into this class and above the threshold, we can dynamically adjust the threshold according to the learning status estimated on learning effect, which can be formulated as:

$$\sigma_t(c) = \sum_{i=1}^N \mathbb{1}(\max(p_t(y|u_i)) > \tau) \cdot \mathbb{1}(\arg \max(p_t(y|u_i) = c)), \quad (3)$$

where $\sigma_t(c)$ reflects the learning effect of class c at time step t . $p_t(y|u_i)$ is the model's prediction for unlabeled data u_i at time step t , and N is the total number of unlabeled data. After normalizing $\sigma_t(c)$ to make its range between 0 to 1, it can be used to scale the fixed threshold τ :

$$\beta_t(c) = \frac{\sigma_t(c)}{\max_c \sigma_t}, \quad (4)$$

$$\mathcal{T}_t(c) = \beta_t(c) \cdot \tau, \quad (5)$$

where $\mathcal{T}_t(c)$ is the flexible threshold for class c at time step t .

Note that because of the confirmation bias caused by the parameter initialization, the model may predict most unlabeled samples into a certain class mistakenly at this stage. Therefore, a warm-up operation is introduced to Eq. 5 as:

$$\beta_t(c) = \frac{\sigma_t(c)}{\max\left\{\max_c \sigma_t, N - \sum_{c=1}^C \sigma_t(c)\right\}}, \quad (6)$$

where C is the number of categories per iteration. $N - \sum_{c=1}^C \sigma_t(c)$ can be regarded as the number of unlabeled data that have not been used. This ensures that at the beginning of the training, all estimated learning effects gradually rise from 0 until most unlabeled data participate in model training.

In addition to the issue we discussed above that will affect the performance of the model, another problem may contribute to an unstable model. In Eq. 5, the flexible threshold depends on the normalized estimated learning effects via a linear mapping. However, in the real training process, this linear mapping will be sensitive to the change of $\beta_t(c)$ in the early stage when the predictions of the model are still unstable, and only make small fluctuations after the class is well-learned in mid and late training stage. Therefore, it is preferable if the flexible thresholds can be more sensitive when $\beta_t(c)$ is large and vice versa.

By rewriting the Eq. 5, a non-linear mapping function is used to enable the threshold to have a non-linear increasing curve when $\beta_t(c)$ ranges uniformly from 0 to 1 as:

$$\mathcal{T}_t(c) = \frac{\beta_t(c)}{2 - \beta_t(c)} \cdot \tau \quad (7)$$

This non-linear mapping strategy ensures the threshold grows slowly when $\beta_t(c)$ is small, and becomes more sensitive as $\beta_t(c)$ gets larger.

After obtaining the category-related threshold, APCL can be achieved by calculating the unsupervised loss, which can be formulated as:

$$\mathcal{L}_{u,t} = \frac{1}{\mu B} \sum_{i=1}^{\mu B} \mathbb{1}(\max(p_i) > \mathcal{T}_t(\arg \max(p_i))) H(\hat{p}_i, p(y|v(u_i))), \quad (8)$$

where $p_i = p(y|\alpha(u_i))$. The category-related flexible thresholds are updated at each iteration. Finally, the total loss \mathcal{L}_t of APCL as the weight combination of supervised and unsupervised loss by a pre-defined scale λ can be formulated as:

$$\mathcal{L}_t = \mathcal{L}_s + \lambda \mathcal{L}_{u,t}, \quad (9)$$

where \mathcal{L}_s is the supervised loss on labeled data:

$$\mathcal{L}_s = \frac{1}{B} \sum_{i=1}^B H(y_i, p(y|\alpha(x_i))). \quad (10)$$

From the above discussion, ACPL is a cost-free method. Practically, once the prediction confidence of unlabeled data u_i is above the fixed threshold τ , the data and its predicted label are marked and will be used for calculating $\beta_t(c)$ at the next step. Such a simple marking action forces the model to pay more attention to those scenes that are difficult to be recognized. Moreover, ACPL can easily cooperate with other pseudo-labeling semi-supervised methods to improve their performance and does not introduce new parameters.

3 Experiments

3.1 Experimental settings

To verify the effectiveness of our RSMATCH, we select the UC Merced Land Use (UCM) data set. There are 2100 high-resolution aerial images with a pixel resolution of 1 foot. UCM contains 21 semantic categories, each of which has 100 images. The size of images is 256×256 .

All of the experiments are implemented by the PyTorch platform using a workstation with two Amax GeForce RTX 3090 GPUs. For a fair comparison, the optimizer for all experiments is stochastic gradient descent (SGD) with a momentum of 0.9 [16]. We use an initial learning rate of 0.03 with a cosine learning rate schedule [17] as $\eta = \eta_0 \cos\left(\frac{7\pi k}{16K}\right)$, where η_0 is the initial learning rate, k is the current training step and K is the total training step that is set to 2^{17} . An exponential moving average with the momentum of 0.999 is also utilized. The batch size of labeled data is 32. μ is set to 7, which means the batch size of unlabeled data is 32×7 for RSMATCH and its compared methods. τ is set to 0.95 for FixMATCH and our RSMATCH. The strong augmentation function used in our experiments is RandAugment [18] while the weak augmentation is a standard flip-and-shift augmentation strategy. We use ResNet-18 [19] as our backbone.

3.2 Classification results

To evaluate the effectiveness of our model, we conduct the classification experiments and compare our method with several semi-supervised algorithms proposed in recent years, including FixMatch [13], MSMatch [9], and SSCL [8]. UCM data set is split into 20% for validation, 20% for test, and 60% for training. Among them, all training samples are further divided into 10% labeled training data and 50% unlabeled training data. We train all the methods with 5 label samples, 8 label samples and 10 label samples per class, respectively. Following the common training definition in semi-supervised methods [13, 14], we use the label amount to define these three classification tasks as 105 labels, 168 labels and 210 labels, respectively. The classification results are exhibited in Table 1. What's more, we pose the confusion matrix of RSMatch of three semi-supervised classification scenarios in Fig. 2. From the observation of the results, we can find RSMatch achieves the best performance than others. The promising results illustrate our method is effective for semi-supervised RSSC task.

Table 1. Classification accuracy (%) on UCM data set.

Method	UCM		
	105 labels	168 labels	210 labels
MSMatch	78.36±0.77	86.48±1.01	88.43±0.56
SSCL	78.43±0.67	87.10±0.44	90.26±0.68
FixMatch	77.08±0.27	86.33±0.56	88.42±0.18
RSMatch (ours)	78.64± 0.32	87.56± 0.17	90.75±0.24

4 Conclusion

In this paper, RSMatch, which utilizes adaptive category-related pseudo labeling (ACPL) strategy, is adopted for remote sensing scene classification based on limited labeled samples and enough unlabeled samples to learn discriminative feature representation according to the learning effects of each class in data set and the learning status of the model. The positive experimental results have demonstrated our proposed RSMatch outperforms the several comparison semi-supervised methods.

References

1. X. Tang, Q. Ma, X. Zhang, F. Liu, J. Ma and L. Jiao, "Attention Consistent Network for Remote Sensing Scene Classification," in IEEE Journal of Selected Topics in

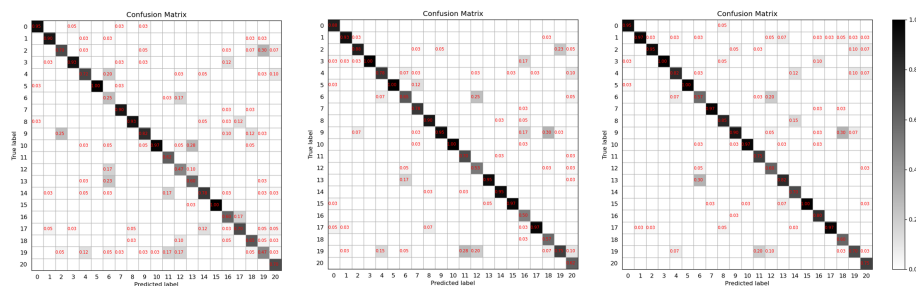


Fig. 2. Confusion matrix of UCM data set. The scene numbers and names are summarized as follows. 1-Agricultural, 2-Airplane, 3-Baseball Diamond, 4-Beach, 5-Buildings, 6-Chaparral, 7-Dense Residential, 8-Forest, 9-Freeway, 10-Golf Course, 11-Harbor, 12-Intersection, 13-Medium Density Residential, 14-Mobile Home Park, 15-Overpass, 16-Parking Lot, 17-River, 18-Runway, 19-Sparse Residential, 20-Storage Tanks, and 21-Tennis Courts,

- Applied Earth Observations and Remote Sensing, vol. 14, pp. 2030-2045, 2021, doi: 10.1109/JSTARS.2021.3051569.
2. Y. Yang et al., "AR2Det: An Accurate and Real-Time Rotational One-Stage Ship Detector in Remote Sensing Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, 2022, Art no. 5605414, doi: 10.1109/TGRS.2021.3092433.
 3. X. Tang et al., "An Unsupervised Remote Sensing Change Detection Method Based on Multiscale Graph Convolutional Network and Metric Learning," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-15, 2022, Art no. 5609715, doi: 10.1109/TGRS.2021.3106381.
 4. 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," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811-2821, May 2018, doi: 10.1109/TGRS.2017.2783902.
 5. Q. Wang, S. Liu, J. Chanussot and X. Li, "Scene Classification With Recurrent Attention of VHR Remote Sensing Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1155-1167, Feb. 2019, doi: 10.1109/TGRS.2018.2864987.
 6. X. Liu, Y. Zhou, J. Zhao, R. Yao, B. Liu and Y. Zheng, "Siamese Convolutional Neural Networks for Remote Sensing Scene Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1200-1204, Aug. 2019, doi: 10.1109/LGRS.2019.2894399.
 7. D. Guo, Y. Xia and X. Luo, "GAN-Based Semisupervised Scene Classification of Remote Sensing Image," in *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 12, pp. 2067-2071, Dec. 2021, doi: 10.1109/LGRS.2020.3014108.
 8. J. Zhang, M. Zhang, B. Pan and Z. Shi, "Semisupervised Center Loss for Remote Sensing Image Scene Classification," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1362-1373, 2020, doi: 10.1109/JSTARS.2020.2978864.
 9. P. Gomez and G. Meoni, "MSMatch: Semisupervised Multispectral Scene Classification With Few Labels," in *IEEE Journal of Selected Topics in Applied Earth*

- Observations and Remote Sensing, vol. 14, pp. 11643-11654, 2021, doi: 10.1109/JS-TARS.2021.3126082.
10. Li Y, Zhang H, Xue X, et al. Deep learning for remote sensing image classification: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(6): e1264.
 11. Yang X, Song Z, King I, et al. A survey on deep semi-supervised learning[J]. arXiv preprint arXiv:2103.00550, 2021.
 12. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.
 13. Sohn K, Berthelot D, Carlini N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[J]. Advances in Neural Information Processing Systems, 2020, 33: 596-608.
 14. Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.
 15. Grandvalet Y, Bengio Y. Semi-supervised learning by entropy minimization[J]. Advances in neural information processing systems, 2004, 17.
 16. Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning[C]//International conference on machine learning. PMLR, 2013: 1139-1147.
 17. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.
 18. Cubuk E D, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 702-703.
 19. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.