CONFIDENCE GUIDED SEMI-SUPERVISED LEARNING IN LAND COVER CLASSIFICATION

Wanli Ma, Oktay Karakuş, Paul L. Rosin

School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, UK

ABSTRACT

Semi-supervised learning has been well developed to help reduce the cost of manual labelling by exploiting a large quantity of unlabelled data. Especially in the application of land cover classification, pixel-level manual labelling in large-scale imagery is labour-intensive, time consuming and expensive. However, existing semi-supervised learning methods pay limited attention to the quality of pseudo-labels during training. However, the quality of training data is one of the critical factors determining network performance. In order to fill this gap, we develop a *confidence-guided semi-supervised learning* (CGSSL) approach to make use of high-confidence pseudo labels and reduce the negative effect of low-confidence ones on training the land cover classification network. Meanwhile, the proposed semi-supervised learning approach uses multiple network architectures to increase pseudo-label diversity. The proposed semi-supervised learning approach significantly improves the performance of land cover classification compared to the classical semi-supervised learning methods, and even outperforms fully supervised learning with a complete set of labelled imagery of the benchmark Potsdam land cover dataset.

***Index Terms*—** Semi-supervised Learning, Land Cover Classification, Multi-modality, Confidence Guided Loss

1. INTRODUCTION

In recent years, with the great success of deep learning in computer vision, automated land cover classification approaches have been significantly improved by using deep learning. Nowadays, the majority of deep-learning land cover classification methods are based on supervised learning [1], which generally requires enormous annotated datasets. Although there are many well-annotated datasets in the computer vision area, it is difficult to generalize deep learning models trained by those datasets to the remote sensing domain. Meanwhile, manual annotation by experts of large-scale remote sensing data, such as satellite products and images captured by drones in complex terrain, is labour-intensive and expensive. Fortunately, a large amount of unlabelled remote sensing data is freely available. Thus, exploring semi-supervised learning approaches for remote sensing applications has become a feasible way to solve the problem of the lack of labelled data [2].

In computer vision, semi-supervised learning has achieved competitive performance in applications such as image classification [3] and semantic segmentation [4]. It has become evident that just using a portion of a labelled dataset and the rest as unlabeled data achieves competitive performance compared to fully supervised learning with a complete set of labelled data. Specifically, mainstream semi-supervised learning methods generate "*pseudo*" labels by using the current prediction of the model during training to supervise the deep learning model. To improve the performance of semi-supervised learning approaches, accuracy labels is regarded as two key points [5]. In semi-supervised land cover classification, similar emphasis is placed on enhancing the accuracy and diversity of pseudo labels. This work utilizes the confidence of pseudo labels to reweight the loss, effectively leveraging high-accuracy pseudo labels while mitigating the impact of low-accuracy ones. On the other hand, to enhance the diversity of pseudo labels, a parallel utilization of three distinct neural network models is employed.

Specifically, the contributions of this work are as follows: (1) We propose a confidence guided cross entropy loss for semi-supervised land cover classification. This is flexible and can be easily transferred to other computer vision tasks such as semantic segmentation. (2) An adaptive mechanism is designed to adjust the threshold automatically (no-interaction setting) for judging the quality of the pseudo labels based on their confidence. (3) We promote the investigation of multiple network outputs in terms of an information theory aspect – entropy – to weight confidence levels of pseudo labels from each network. Then, we use this information to optimise the unsupervised training process in semi-supervised land cover classification.

2. BACKGROUND & RELATED WORK

In general terms, semi-supervised learning is defined as an approach that lies between supervised and unsupervised learning. During the supervised learning step, various widely applied semantic segmentation methods can be used such as PSPNet [6], UNet [7], SegNet [8], DeepLabV3+[9]. In

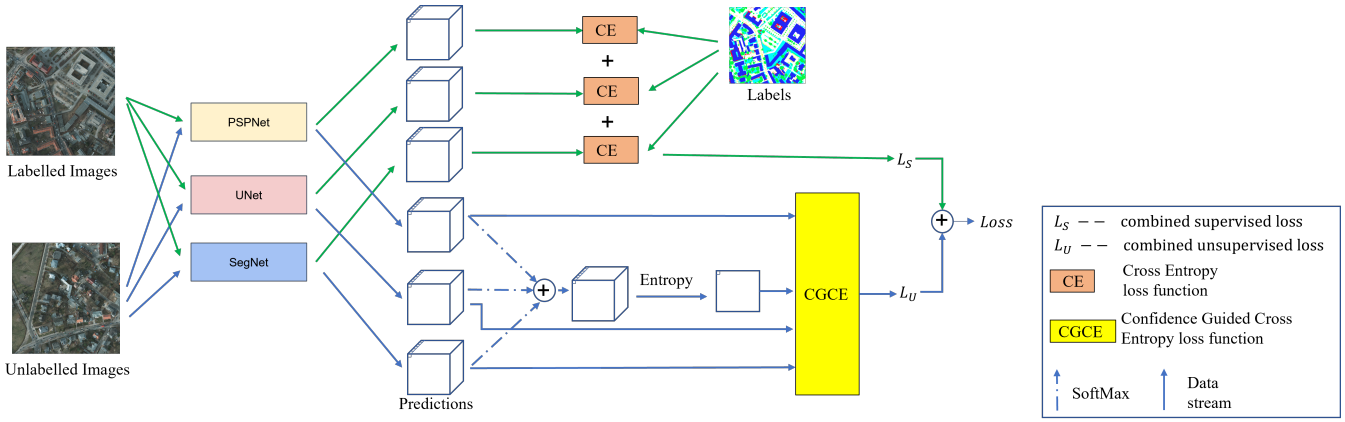

Fig. 1. Overall framework of the confidence guided semi-supervised learning (CGSSL) approach

current semi-supervised learning research [4, 10, 11] within the field of computer vision, the commonly used network is DeepLabV3+ with a pre-trained backbone e.g. ResNet 50. However, in remote sensing, network performance rankings differ and no specific architecture dominates. Especially when the number of labelled data is small, due to the exploitation of low- and high-level features via efficient skip-connections, a simpler method like U-net shows competitive (even better) results compared to DeepLabV3+ with a pre-trained backbone.

Consistency regularization [12] describes a class of unsupervised learning algorithms as a part of semi-supervised learning, that are easy to implement and widely compatible with supervised learning segmentation networks. The key idea of consistency regularization is to force perturbed models (or perturbed inputs) to have consistent outputs for unlabelled inputs. Based on this concept, cross pseudo supervision (CPS) [13] and CPS-to-n-networks (n-CPS) [14] show considerable success, which yields state-of-the-art in semantic segmentation benchmark datasets, e.g. Cityscapes. In a previous land cover mapping work [15], we have shown that consistency regularization is helpful for leveraging minimal supervision. However, CPS and n-CPS use pseudo labels to supervise the network regardless of their quality. In addition, perturbed models in those methods have the same structure which causes these networks to tend to output similar predictions after certain iterations. In order to increase the diversity of pseudo-labels in parallel using different segmentation networks stands out to be an efficient and accurate alternative [16].

3. METHOD

The proposed semi-supervised learning approach is divided into supervised learning and unsupervised learning parts as shown in Figure 1. In each training iteration, both labelled and unlabelled data are given as input into the three different networks (PSPNet [6], UNet [7], SegNet [8]). The labelled data is used in a regular supervised learning pattern to train these models by using the cross-entropy loss function. On the other hand, unlabeled data is utilized to generate pseudo labels, which are exploited to inform each network. The prediction of each network is the class probabilities for all pixels of the corresponding input image, and then the predictions from three networks are added linearly after a softmax layer to generate a comprehensive prediction. In this case, if the confidence distributions among the three networks are identical, the operation of linear addition will promote the distribution to be sharper than one of the predictions (low uncertainty). Otherwise, the combined prediction will not have a distinct strong peak (high uncertainty). Considering the fact that the information entropy is a measure of uncertainty, we calculated the entropy of the classification distribution based on the combined prediction to confirm confidence in the predictions. Furthermore, the proposed confidence-guided cross-loss function is designed to limit the negative contribution of the pseudo labels with low entropy (high uncertainty) to the network parameter optimisation. Finally, the total loss is set to a linear combination of supervised loss L_s and $\mathcal{L}_u$ unsupervised loss as

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_u, \tag{1}$$

where λ is the trade-off weight between supervised and unsupervised losses. It is worth noting that the unsupervised loss $\mathcal{L}_u$ is the linear addition of 6 losses which results from 3-model cross supervision.

As shown in Figure 2, the proposed confidence-guided cross-entropy loss module is used to calculate the unsupervised loss. The aim of the loss is to make use of the confidence of predictions to re-weight the cross entropy loss at the pixel level among the high-quality predictions based on their entropy at the class level. The mean value of entropy is regarded as a threshold to decide on the reliability of the estimated confidence. The unreliable confidence values are assumed to provide limited useful information for re-weighting the loss. Thus, the loss of these pixels is not reweighted and just uses the standard cross entropy loss function. How-

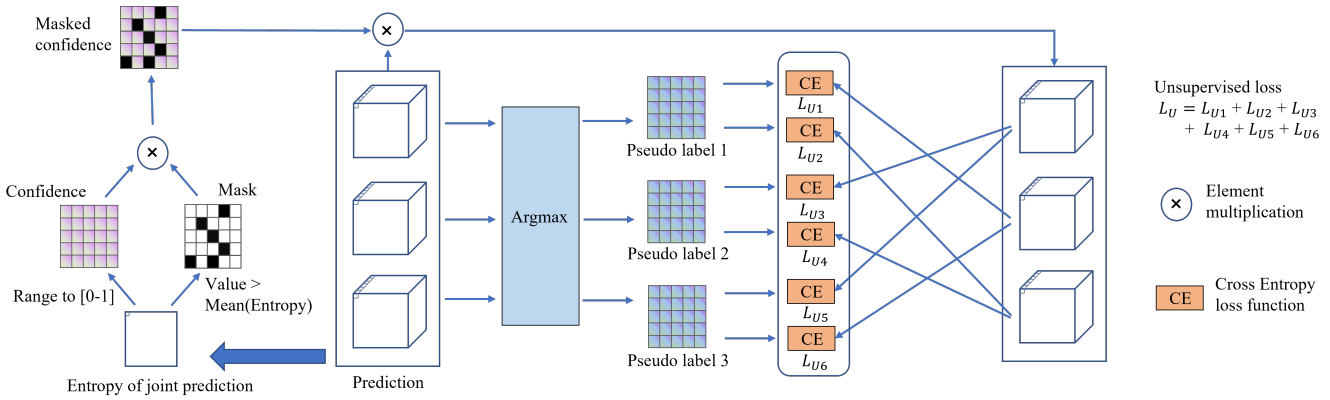

Fig. 2. Illustration of Confidence Guided Cross Supervision.

Table 1. Performance comparison of different methods for Potsdam dataset.

Model	Type	Accuracy	Precision	Recall	mIoU	F_1-score
U-Net1† [7]	Supervised	85.36%	76.75%	81.23%	67.59%	78.92%
U-Net2* [7]	Supervised	84.26%	76.45%	79.32%	66.49%	77.86%
Mean Teacher [17]	Semi-Supervised	84.58%	78.52%	80.88%	68.24%	79.68%
CPS [13]	Semi-Supervised	85.30%	77.94%	80.75%	68.38%	79.32%
CGSSL (ours)	Semi-Supervised	**86.59%**	**79.06%**	**83.54%**	**70.17%**	**81.24%**

†*U-Net1 was trained with the whole 3456 labelled samples.* **U-Net2 was trained with 1728 labelled samples.*

ever, the confidence of predictions above the mean value is regarded as reliable, which is used for entropy calculation to re-weight the loss. The weight w is defined as follows $w = \frac{\max(\mathcal{I}) - \mathcal{I}}{\max(\mathcal{I}) - \min(\mathcal{I})} + 1$, where $\mathcal{I}$ refers to the entropy of a series of class predictions for each pixel. Thus, since $w \geq 1$ the effect of these pixels is increased during training compared to the pixels with unreliable confidence values. Instead of directly using the probability of predictions to weight the loss (focal loss [18]), entropy is used as a measurement to represent the confidence of comprehensive pseudo labels from multiple distinct networks. The weight, w, is added as a factor to standard cross entropy loss $\ell(x, y)$ to favour the high-quality pseudo labels.

$$\ell(x, y) = \frac{\sum_{n=1}^{N} -w \log \frac{\exp(x_{n,y_n})}{\sum_{c=1}^{C} \exp(x_{n,c})}}{N}, \qquad (2)$$

where x represents the input, y denotes the target class, w signifies the weight, C indicates the number of classes, and N is the batch size. Finally, inspired by [13] and [16], the unsupervised loss is acquired by cross-supervision between predictions from different networks.

4. EXPERIMENTS AND RESULTS

We evaluated our method using the ISPRS Potsdam dataset [19], which consists of 38 multi-source 6000×6000 patches, including infrared, red, green, and blue orthorectified optical images, and corresponding digital surface models (DSM). We divided these data tiles into 512×512 patches, resulting in 3456 training samples and 2016 test samples. Both true orthophoto and DSM modalities have a 5 cm ground sampling distance. The dataset contains six manually classified land cover classes: *impervious surfaces, buildings, low vegetation, trees, cars,* and *clutter/background*.

In order to compare the proposed method – CGSSL – we utilised two classic semi-supervised models of Mean Teacher [17] and CPS [13]. The number of labelled data used in the aforementioned semi-supervised learning approaches is only half (1728 samples) of the whole training split of the Potsdam dataset. We remove the labels of the remaining half and just used the images in the unsupervised part. We also provide the performance of UNet [7] only in supervised learning patterns for the whole and half-labelled data which are named U-Net1 and U-Net2 in the sequel, respectively. The same test set is used to evaluate all models. Thus, if applying the proposed method in real-world scenarios, annotate only a subset of images and leave the remaining unlabelled to minimize manual labour.

Our experiments are implemented by Pytorch. we use a mini-batch SGD optimizer adopted a poly learning rate policy. All the experiments were performed on NVIDIA A100-sxm in the GW4 Isambard. We thoroughly evaluated all models using class-related performance metrics, including accuracy, precision, recall, mean intersection over union (mIoU), and F_1-score. As shown in Table 1, CGSSL shows the best performance in terms of all performance metrics. Especially, CGSSL improves recall significantly due to the great reduction of false negatives in prediction. Even though CGSSL

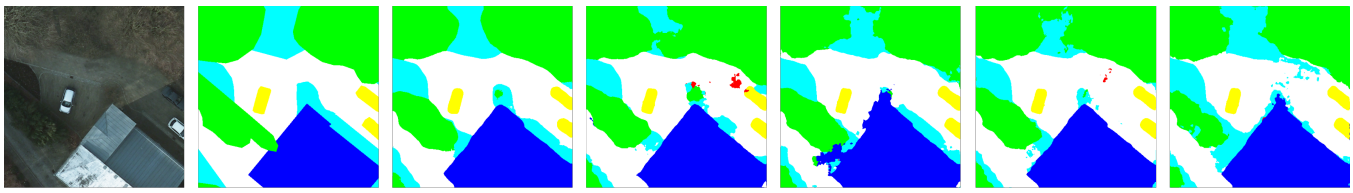

Fig. 3. Visual Results of each method on Potsdam Dataset. Values between parentheses refer to accuracy in percentages. #U-Net1 was trained with the whole 3456 labelled samples. *U-Net2 was trained with 1728 labelled samples.

only uses half of the labelled data, its performance is even better than UNet1 which is trained with the whole dataset. Figure 3 shows a case of predictions for all methods where CGSSL is mostly close to the ground truth and no other classes are predicted.

5. CONCLUSION

In this paper, we introduced an innovative semi-supervised learning approach for land cover classification that utilizes a confidence-guided cross-entropy loss. Especially, an adaptive loss was provided for the semi-supervised learning to exploit pseudo labels with an information theory perspective. This is also flexible to be transferred to various other semi-supervised learning tasks. The proposed method shows considerable performance and benefits from unlabeled data for land cover classification. Meanwhile, since three networks are required to increase the diversity of pseudo labels in training processing, one of the drawbacks of this method is the extensive computational requirement and might not be efficient to implement in edge computing devices for practical applications. This is already listed as our future work to further develop lighter segmentation architectures for semi-supervised learning.

6. REFERENCES

[1] A. Vali et al., "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sensing*, vol. 12, no. 15, p. 2495, 2020.

[2] J.-X. Wang et al., "Semi-supervised semantic segmentation of remote sensing images with iterative contrastive network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[3] B. Zhang et al., "FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling," *in NeurIPS*, vol. 34, pp. 18 408–18 419, 2021.

[4] H. Hu et al., "Semi-supervised semantic segmentation via adaptive equalization learning," *in NeurIPS*, vol. 34, pp. 22 106–22 118, 2021.

[5] S. Zhang et al., "Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition," *Knowledge-Based Systems*, vol. 229, p. 107340, 2021.

[6] H. Zhao et al., "Pyramid scene parsing network," in *Proceedings of CVPR*, 2017, pp. 2881–2890.

[7] O. Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[8] V. Badrinarayanan et al., "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *TPAMI*, vol. 39, no. 12, pp. 2481–2495, 2017.

[9] L.-C. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the ECCV*, 2018, pp. 801–818.

[10] Y. Wang et al., "Semi-supervised semantic segmentation using unreliable pseudo-labels," in *Proceedings of CVPR*, 2022, pp. 4248–4257.

[11] H. Xu et al., "Semi-supervised semantic segmentation with prototype-based consistency regularization," *in NeurIPS*, vol. 35, pp. 26 007–26 020, 2022.

[12] G. French et al., "Semi-supervised semantic segmentation needs strong, varied perturbations," *arXiv preprint arXiv:1906.01916*, 2019.

[13] X. Chen et al., "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of CVPR*, 2021, pp. 2613–2622.

[14] D. Filipiak et al., "n-CPS: Generalising cross pseudo supervision to n networks for semi-supervised semantic segmentation," *arXiv preprint arXiv:2112.07528*, 2021.

[15] W. Ma, O. Karakuş, and P. L. Rosin, "AMM-FuseNet: attention-based multi-modal image fusion network for land cover mapping," *Remote Sensing*, vol. 14, no. 18, p. 4458, 2022.

[16] X. Luo et al., "Semi-supervised medical image segmentation via cross teaching between CNN and transformer," in *MIDL*. PMLR, 2022, pp. 820–833.

[17] A. Tarvainen et al., "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *in NeurIPS*, vol. 30, 2017.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE ICCV*, 2017, pp. 2980–2988.

[19] F. Rottensteiner et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Annals*, vol. 1, no. 1, pp. 293–298, 2012.