Large-Scale Land Cover Mapping with Fine-Grained Classes via Class-Aware Semi-Supervised Semantic Segmentation

Runmin Dong[1,4], Lichao Mou[2], Mengxuan Chen[1,4], Weijia Li[3], Xin-Yi Tong[2],
Shuai Yuan[1,4], Lixian Zhang[1,4], Juepeng Zheng[3,4], Xiao Xiang Zhu[2], Haohuan Fu[1,4,*]
[1]Tsinghua University [2]Technical University of Munich [3]Sun Yat-Sen University
[4]Tsinghua University - Xi'an Institute of Surveying and Mapping Joint Research Center
drm@mail.tsinghua.edu.cn, haohuan@tsinghua.edu.cn

Abstract

Semi-supervised learning has attracted increasing attention in the large-scale land cover mapping task. However, existing methods overlook the potential to alleviate the class imbalance problem by selecting a suitable set of unlabeled data. Besides, in class-imbalanced scenarios, existing pseudo-labeling methods mostly only pick confident samples, failing to exploit the hard samples during training. To tackle these issues, we propose a unified Class-Aware Semi-Supervised Semantic Segmentation framework. The proposed framework consists of three key components. To construct a better semi-supervised learning dataset, we propose a class-aware unlabeled data selection method that is more balanced towards the minority classes. Based on the built dataset with improved class balance, we propose a Class-Balanced Cross Entropy loss, jointly considering the annotation bias and the class bias to re-weight the loss in both sample and class levels to alleviate the class imbalance problem. Moreover, we propose the Class Center Contrast method to jointly utilize the labeled and unlabeled data. Specifically, we decompose the feature embedding space using the ground truth and pseudo-labels, and employ the embedding centers for hard and easy samples of each class per image in the contrast loss to exploit the hard samples during training. Compared with state-of-the-art class-balanced pseudo-labeling methods, the proposed method improves the mean accuracy and mIoU by 4.28% and 1.70%, respectively, on the large-scale Sentinel-2 dataset with 24 land cover classes.

1. Introduction

Land cover mapping provides pixel-level information for urban management, climate change research, ecosystem protection, and other sustainability-related applications [21, 27]. With the rapid progress of computer vision technology, it is desired to automatically gain large-scale and fine-grained land cover information from remote sensing images to support the demand of existing earth system science studies and bring new opportunities for intelligent city study [31]. However, fine-grained annotations for land cover mapping are high-cost and time-consuming, limiting labeled data collection for large-scale applications. Semi-supervised learning (SSL) is a potential solution to this problem, as it utilizes both existing labeled data and a large number of unlabeled data [37].

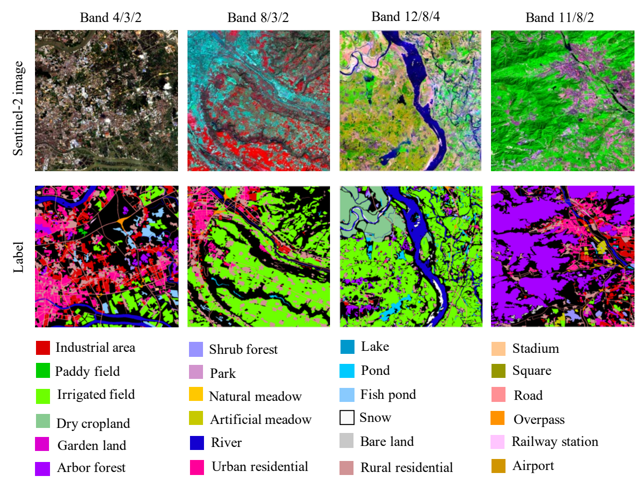

Figure 1. Examples of the labeled data. The top row shows Sentinel-2 images containing 13 bands. The bottom row shows the corresponding labels. We use a fine-grained classification system with 24 land cover classes in this work.

Large-scale land cover mapping with fine-grained classes usually suffers from class imbalance, as shown in Figure 1 and 3. The labeled data tends to be biased due to the difficulty of data collection and label annotation varying by class. The class imbalance leads to suboptimal per-

*Corresponding author

formance in the minority classes. Most SSL methods for large-scale land cover mapping are typically based on the assumption that labeled and unlabeled data have the same class distribution [41, 37]. However, this assumption does not hold in most scenarios, resulting in the bias of class estimation. As such, it is non-trivial to jointly utilize the labeled and unlabeled data to train the network for semi-supervised semantic segmentation. Although many pseudo-labeling methods aim to assign pseudo-labels more steadily and correctly for different classes on the unlabeled data, e.g., dynamically setting threshold or proportion [17, 44], they still rely on setting hyperparameters and empirical principles, resulting in a lack of flexibility and generalization ability in applications. Besides, pseudo-labeling methods mostly rely on easy samples with high confidence scores, thus failing to exploit the hard samples on the unlabeled data.

To alleviate the above issues, we propose a unified Class-Aware Semi-Supervised Semantic Segmentation framework. We exploit the potential of the unlabeled data and alleviate the class imbalance problem in large-scale land cover mapping with fine-grained classes. The proposed class-aware unlabeled data selection method constructs an SSL dataset and compensates for the class imbalance on the labeled data. Based on the built SSL dataset, we propose a class-balanced learning method to remove the annotation bias and class bias on the SSL dataset. We dynamically estimate the class prior information on the entire SSL dataset during training to re-weight both samples and classes of the loss function for the labeled data. Finally, we propose the Class Center Contrast method to jointly use the labeled and unlabeled data for training. The ground truth and pseudo-labels are leveraged as the guide to decompose the feature embedding space. We estimate the hard, easy, and overall embedding centers for each class per image and apply the contrastive loss to optimize the distances between class centers to utilize both easy and hard samples. The class imbalance issue is alleviated by utilizing the proposed method, and the performance of large-scale land cover mapping with fine-grained classes is effectively improved.

2. Related Work

Large-Scale Land Cover Mapping. Land cover mapping classifies each pixel of images into a unique class, which can be regarded as a semantic segmentation problem [6, 11]. Large-scale land cover mapping with fine-grained classes can provide necessary information for earth system science studies and land spatial layout optimization. However, manual or semi-manual labeling is high-cost and time-consuming [34]. Besides, the classification systems of most large-scale studies are coarse. Thus the provided land cover information is limited. For example, the Chesapeake dataset covering 160,000 km^2 costs \$1.3 million but only contains four classes [33]. DynamicEarthNet dataset covers 17,000 km^2 and includes 7 basic land cover classes [36]. To overcome the limited area of the labeled data and improve the model generalization, some works utilize openly available produced data (e.g., OpenStreetMap data and land cover products), which contains many noisy labels but can provide prior information in a large-scale area [19, 10]. These works still use coarse classification systems of land cover mapping. Five-Billion-Pixels [37] is a large-scale land cover dataset with 24 categories, covering about 60,000 km^2, which opens the door to large-scale land cover mapping with fine-grained classes. However, this dataset suffers from a severe class-imbalance problem (the common categories cover hundreds of times more pixels than the rare ones), due to the difficulty of data collection and label annotation varying by class.

Class-Imbalanced Learning. To solve the class-imbalance problem, related works include data re-sampling [3, 4], loss re-weighting [22, 32, 18], margin modification [5, 35], and decoupled learning [20, 47]. In self-supervised learning and SSL, recent works devote to re-balancing the class distribution of SSL data by assigning pseudo-labels for the unlabeled data to address the class imbalance issue [41, 15]. For example, Gui et al. [14] propose a class-aware pseudo-labeling method for SSL, dynamically adjusting the threshold for selecting pseudo-labels to obtain better performance on minority classes. Similarly, Hu et al. [17] propose a label bias removal method and dynamically determine the threshold for each class to assign more pseudo-labels for minority classes. However, the hard samples with relatively low confidence scores are ignored for optimizing the model, leading to sub-optimal performance.

Semi-Supervised Semantic Segmentation. SSL methods leverage the relationships between the labeled and unlabeled data to improve the model accuracy and generalization, alleviating the limited area of the labeled data [8]. Semi-supervised semantic segmentation methods include entropy minimization [38], consistency regularization [30], and pseudo-labeling [42]. Pseudo-labeling is widely used in semi-supervised semantic segmentation, and many works contribute to selecting pseudo-labels more steadily and correctly [29]. For example, Tong et al. [37] propose a dynamic pseudo-label assignment method, in which the number of pseudo-labels is dynamically increased with training iterations. However, these methods still rely on the manual setting of hyperparameters and suffer from the uncertainty of pseudo-labels. Alternatively, contrastive learning methods in supervised learning [39, 16] and SSL [46, 40] constrain the representations of positive samples against these negative samples in feature space, leading to promising results. Different from the manner of data augmentation-based positive sample construction [43, 23] and memory bank-based negative sample construction [1, 40], we leverage the ground-truth and pseudo-labels as a guide to decompose the feature embedding space, and estimate the embedding centers for each class per image to stably select easy and hard samples for the proposed Class Center Contrast method.

3. Methodology

3.1. The Overall Workflow of Our Approach

Our approach consists of four steps. First, a baseline model is trained on the labeled data. Second, we propose a class-aware unlabeled data selection method to build an SSL dataset that is more balanced towards minority classes. The baseline model is used to predict land cover classes on all collected data from a large unlabeled data pool and calculate the statistics of the class distribution for each image patch. The image patches containing pixels of minority classes are selected as unlabeled data. Then the SSL dataset is constructed by combining the labeled data and the selected unlabeled data. Third, we propose a Class-Balanced Cross Entropy loss and a Class Center Contrast method. To alleviate the class imbalance problem, we re-weight the cross entropy loss on labeled data in both sample and class level by jointly considering the annotation bias and the class bias on the built SSL dataset. Besides, we decompose the feature embedding space using the ground truth and pseudo-labels, and propose an embedding center selection and contrast method for class-balanced and annotation-guided learning on unlabeled data. The pipeline of the proposed class-aware SSL model is shown in Figure 2. Finally, we perform model inference for large-scale land cover mapping.

3.2. Class-Aware Unlabeled Data Selection

A suitable set selection of unlabeled data is important for SSL model development. Inspired by the observation in [41] that existing SSL algorithms can produce pseudo-labels on minority classes with high precision, we propose an unlabeled data selection strategy that tends to select unlabeled image patches containing pixels with minority classes.

First, we train a semantic segmentation model using labeled data. We apply the off-the-shelf class balance strategy [4] designed for supervised learning to guarantee the model ability of minority class estimation. The definition of minority classes depends on the class proportion and the class accuracy on the labeled data. Then, we perform model inference on all candidate unlabeled data to predict the land cover class of each pixel, thereby obtaining the class distribution of each candidate image patch. We denote the proportion of pixels of all minority classes in an image patch as P_m. Finally, we formulate a principle to automatically select unlabeled data according to the class distribution of each candidate image patch. Specifically, the proportions of the 24 classes on the labeled dataset are shown in Figure 3. The classes with a proportion lower than 1% generally achieve an accuracy of less than 30% within our dataset. We collect all the images with $P_m \geq 1\%$, which leads to about 25,000 images over the about 500,000 candidates (see Section 4.3). Then we randomly select another 25,000 images from those with $0 < P_m < 1\%$, leading to a total of 50,000 unlabeled images, which is five times the number of the labeled training data. The authors suggest selecting unlabeled patches with relatively more pixels of minority classes (i.e., larger P_m), and those encompassing rare classes characterized by fewer pixels within the minority classes. Note that although the prediction on the unlabeled data is not absolutely accurate, it can preserve image patches that are most likely to contain pixels with minority classes, which contributes to building a more balanced SSL dataset.

3.3. Class-Balanced Learning

In our built SSL dataset, the class distribution of unlabeled data differs from that of the labeled data. It can be regarded as an annotation bias problem on the entire dataset. Besides, the SSL dataset still suffers from a class imbalance issue. Consequently, using the class prior information estimated on the labeled data to the re-weight loss function is inappropriate. To solve the above issues, we propose a class-balanced learning method to jointly remove the annotation bias and class bias on the SSL dataset.

First, we re-weight samples of labeled data to remove the annotation bias. Owing to our unlabeled data selection method, the entire SSL dataset is more balanced towards minority classes than the original labeled dataset. Therefore, we can estimate the class prior on the entire SSL dataset to partially remove the annotation bias. Specifically, we denote L as a label indicator, where $L = 0$ and $L = 1$ represent the unlabeled and labeled data, respectively. The model parameters θ of labeled data in traditional SSL methods are obtained by maximizing likelihood estimation in the formula (1).

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} \log P(Y|X, L=1;\theta) \\ &= \arg\max_{\theta} \sum_{(x,y)\in D_{L=1}} \log P(y|x;\theta),\end{aligned} \tag{1}$$

where $\hat{\theta}$ is the estimated model parameter, X represents samples, Y represents classes, and $D_{L=1}$ denotes the set of labeled data. Due to the annotation bias in the SSL dataset, we have $P(y|x; L = 1) \neq P(y|x)$. According to causal inference [13, 17], in the entire SSL dataset, $X \rightarrow Y \rightarrow L$ can be regarded as a one-way chain, where the class is dependent on its pixel, and whether the pixel has a label or not depends on its class. According to the

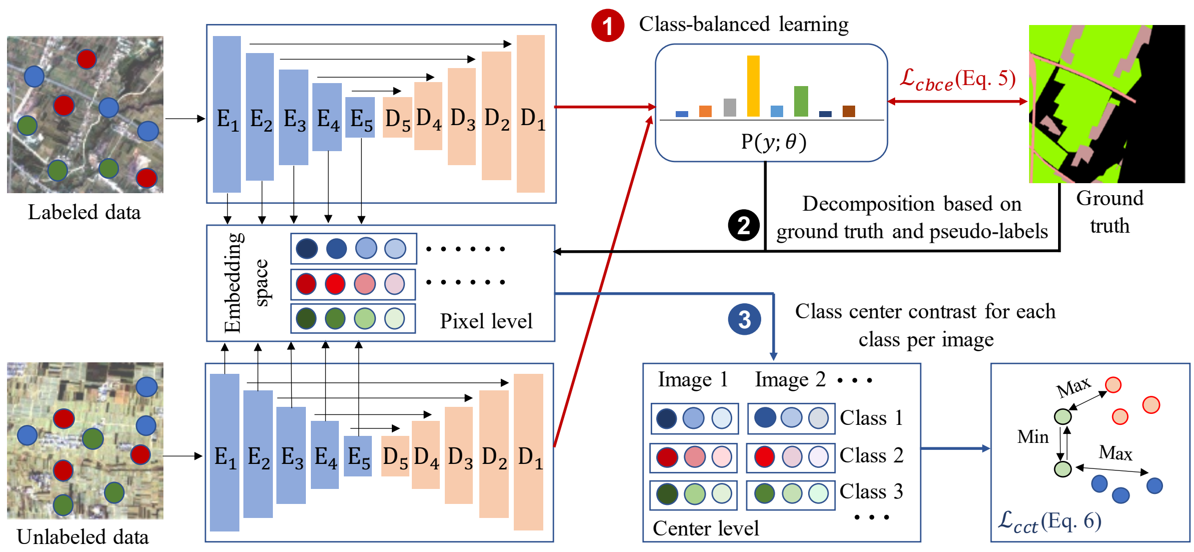

Figure 2. The pipeline of the proposed Class-Aware Semi-Supervised Semantic Segmentation framework with class-balanced learning, embedding center estimation, and class center contrast. The class-aware unlabeled data selection process is not shown in the figure.

conditional independence rule, we obtain the expression, $P(X|Y, L = 1) = P(X|Y)$ when conditioning on Y. Therefore, estimating the model parameters of the entire SSL dataset can be formalized as in the formula (2) [17], where $s(x, y) = \frac{\log P(y|x;\theta) - \log P(y;\theta)}{\log P(y|x;\theta)}$ is the loss weight of each labeled sample (x, y). Thus, it can be interpreted as modifying the class prior $P(y; \theta)$ to remove the annotation bias of the labeled data.

$$\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} \log P(X|Y;\theta) \\
&= \arg\max_{\theta} \sum_{(x,y)\in D_{L=1}} \log P(x|y;\theta) \\
&= \arg\max_{\theta} \sum_{(x,y)\in D_{L=1}} \log \frac{P(y|x;\theta)P(x;\theta)}{P(y;\theta)} \\
&= \arg\max_{\theta} \sum_{(x,y)\in D_{L=1}} \log \frac{P(y|x;\theta)}{P(y;\theta)} \\
&= \arg\max_{\theta} \sum_{(x,y)\in D_{L=1}} s(x,y) \cdot logP(y|x;\theta),
\end{aligned} \tag{2}$$

Second, although the SSL dataset is more balanced than the labeled data, it still suffers from the class imbalance problem. Therefore, we further estimate the class bias on the SSL dataset. A widely used class re-weighting loss function on supervised learning can be formalized as follows:

$$\mathcal{L}_{wce} = \sum_{k=1}^{K} W_k \sum_{i=1}^{n_k} y_{ik} \log p_{ik}, \tag{3}$$

where $W_k = \frac{1}{\log(1+r_k)}$ and r_k is the ratio of the class k on the labeled data. K is the number of classes, n_k is the number of samples of class k, y_{ik} and p_{ik} is the ground-truth and predicted probability of the i-th sample for class k, respectively. In SSL, we approximate r_k with $P(y; \theta)$, the class distribution on the entire dataset.

Finally, we perform an online estimation strategy for $P(y; \theta)$ to simplify the training process. We compute the average of class probability distributions over all pixels in a mini-batch, which is denoted as $P(Y; B_t, \theta_t)$. B_t and θ_t are the samples and model parameters of the current batch, respectively. Then the approximate class distribution, denoted as $\tilde{P}(Y)$, is updated by the weighted sum of the obtained class distribution of this mini-batch and the approximate class distribution in the previous iteration, which is formalized as follows:

$$\tilde{P}(Y) \leftarrow \mu\tilde{P}(Y) + (1-\mu)P(Y; B_t, \theta_t), \tag{4}$$

where $\mu \in [0, 1]$ is a weighting coefficient. We set $\mu = \frac{B}{N}$, in which B is the batch size and N is the total number of samples in the dataset. We initialize $\tilde{P}(Y)$ with the class distribution calculated on the labeled data. As such, the proposed Class-Balanced Cross Entropy loss on the labeled data is formulated as follows:

$$\mathcal{L}_{cbce} = \sum_{k=1}^{K} W_k \sum_{i=1}^{n_k} s(x_{ik}, y_{ik}) y_{ik} \log p_{ik}. \tag{5}$$

3.4. Class Center Contrast

Training via pseudo-labeling is commonly used to utilize unlabeled data in SSL. In fine-grained classes, assigning reliable labels for minority classes is challenging because the segmentation model tends to be biased towards the majority classes [26]. Besides, existing label assignment methods usually use only high-confidence predictions, which may offer limited information to model learning. In this work, our method selects reliable and hard samples. Moreover, instead of directly using the pseudo-labels for supervised learning on unlabeled data, we leverage the pseudo-labels as the guide to decompose the feature embedding space. In general, the distance of feature embeddings from the same class should be minimized, while the distance of those from different classes should be maximized. Therefore, we propose an online embedding center selection method that divides each class into hard and easy sample regions on each image. Specifically, we use the prediction probabilities of ground truth or pseudo-labels to separate easy and hard samples. The predicted probability of each pixel is associated with its feature embedding, and we then calculate the average prediction probability of each class for each image. Feature embeddings with above-average probabilities for the corresponding class are considered easy samples and vice versa. Thus we can obtain three feature embedding centers, including easy, hard, and all feature embedding centers, as contrastive samples. With this method, on the one hand, the obtained feature embeddings for each class are with relatively high confidence, as the samples are close to feature local centers for each class in the embedding space. On the other hand, these feature embeddings include hard samples of each class.

Inspired by the existing contrastive learning method [39], we perform cross-image contrastive learning in each mini-batch. Since both labeled and unlabeled data are trained in a mini-batch, the embeddings of the labeled data could help the feature embedding learning of the unlabeled data. Formally, denote P_m and N_m as the feature embedding collections of the positive and negative samples in a mini-batch for the m-th feature embedding center $\boldsymbol{c_m}$. The proposed Class Center Contrastive loss for the m-th feature embedding center is formulated as follows:

$$\mathcal{L}_{cct}^{m} = -\frac{1}{|P_m|} \sum_{\boldsymbol{c_m^+} \in P_m} \log \frac{e^{(\boldsymbol{c_m} \cdot \boldsymbol{c_m^+}/\tau)}}{e^{(\boldsymbol{c_m} \cdot \boldsymbol{c_m^+}/\tau)} + \sum_{\boldsymbol{c_m^-} \in N_m} e^{(\boldsymbol{c_m} \cdot \boldsymbol{c_m^-}/\tau)}}, \tag{6}$$

in which τ is the temperature parameter, $\boldsymbol{c_m^+}$ and $\boldsymbol{c_m^-}$ are the positive and negative samples, respectively.

The overall loss function can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_{cbce} + \lambda \mathcal{L}_{cct}, \tag{7}$$

where λ is the loss weight.

4. Data

4.1. Image Collection and Preparation

Considering free access, multispectral bands, and relatively high spatial resolution, all satellite images used in this work are acquired from Sentinel-2 satellites. Sentinel-2 imagery has 13 spectral bands at 10 m (R, G, B, and near-infrared bands), 20 m (six red edge and shortwave infrared bands), and 60 m (three atmospheric correction bands) resolutions [12]. The abundant multispectral information and up to 10 m spatial resolution benefit land cover mapping with a fine-grained classification system.

We collect Sentinel-2 images with radiometric correction, geometric corrections, and atmospheric correction, covering more than 60 cities in China. We apply an off-the-shelf Sentinel-2 sharpening method, DSen2Net [24], to reconstruct low-resolution bands and uniform the resolution of all bands of Sentinel-2 images to 10 m.

4.2. Label Collection and Classification System

The land cover labels and classification system are from the Five-Billion-Pixels dataset [37], labeled with 4 m resolution. It includes 24 land cover classes. We resize the 4 m labels to 10 m to match Sentinel-2 images. The labeled area covers about 60,000 km^2 in China. The classification system includes the industrial area (C1), paddy field (C2), irrigated field (C3), dry cropland (C4), garden land (C5), arbor forest (C6), shrub forest (C7), park (C8), natural meadow (C9), artificial meadow (C10), river (C11), urban residential (C12), lake (C13), pond (C14), fish pond (C15), snow (C16), bare land (C17), rural residential (C18), stadium (C19), square (C20), road (C21), overpass (C22), railway station (C23), and airport (C24).

4.3. Dataset

Examples of the labeled data are shown in Figure 1. The training, validation, and test datasets include 9,995, 2,000, and 2,503 images with a size of 256 × 256, respectively, according to the geographical division. The unlabeled data is from more than 60 dispersed administrative districts in China and has no overlap with the labeled data. The number of candidate unlabeled images is about 500,000, each with a size of 256 × 256. According to our minority class-biased unlabeled data selection strategy, we filter 50,000 images to build the unlabeled set, which covers more scenes with minority classes. Figure 3 shows the proportion of each land cover class on labeled training data, unlabeled training data, and all training data. The majority classes cover hundreds of times more pixels than the minority classes on the labeled training data.

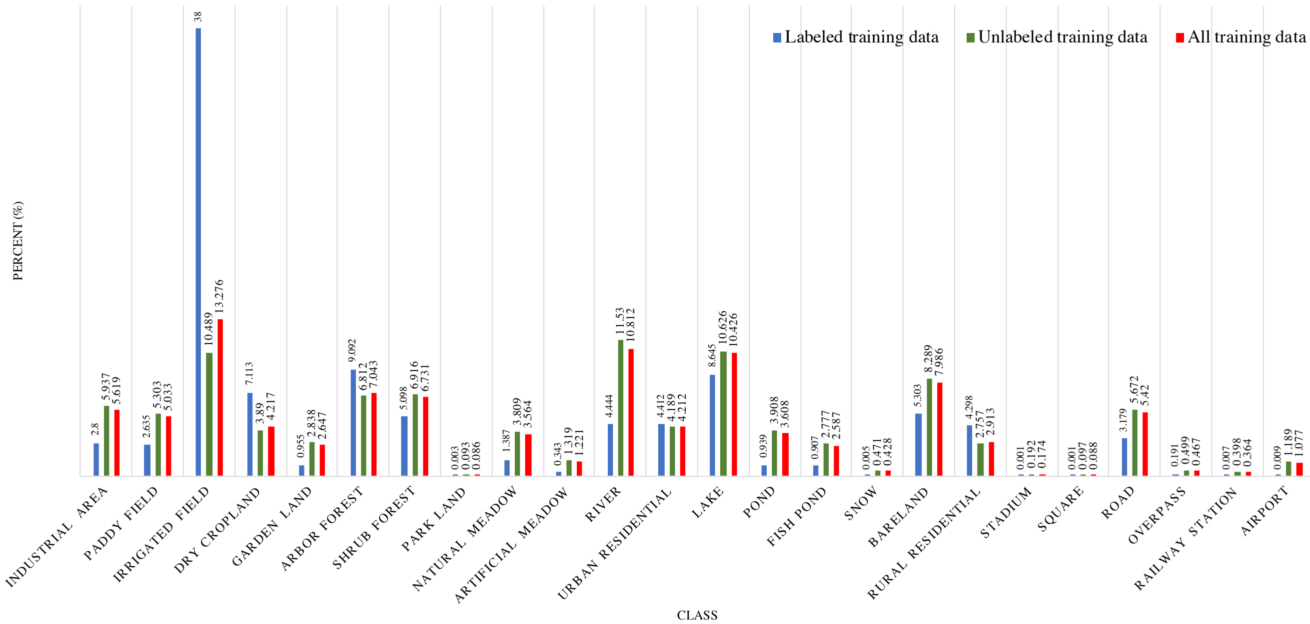

Figure 3. Percentage of each land cover class on the labeled training data, unlabeled training data, and all training data.

Method	OA	mAcc	mIoU	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Baseline	73.24	45.77	35.61	71.24	66.25	88.66	17.90	16.93	62.26	27.01	17.26	44.65	11.33
SimCLR [7] +finetune	74.07	45.81	35.17	74.49	63.33	90.88	14.10	17.80	62.10	23.57	18.39	48.42	21.02
Pseudo -labeling [25]	74.61	48.83	38.03	76.85	69.60	91.55	13.47	20.52	59.19	31.63	37.53	44.90	10.68
Advent [38]	74.81	49.25	38.98	75.88	68.38	91.98	12.20	15.01	62.46	25.91	44.77	54.75	8.10
CADR [17]	75.49	47.83	36.63	76.32	52.43	91.48	21.54	24.35	64.74	31.85	22.59	59.88	17.25
DPA [37]	75.36	48.57	37.77	76.05	69.40	92.33	15.51	19.79	66.30	27.05	23.25	52.41	12.48
Ours	**75.97**	**53.53**	**40.68**	81.71	80.07	88.89	22.27	37.32	67.06	34.87	26.44	56.73	17.16

C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
70.86	82.76	78.93	24.49	57.71	3.04	48.67	77.23	10.80	7.90	71.42	45.83	34.95	60.34
71.47	83.27	79.41	30.54	64.73	0.75	48.77	68.82	2.00	0	73.59	44.53	33.87	63.72
72.40	82.90	83.56	16.94	52.44	13.66	55.06	72.58	21.05	13.03	70.27	60.60	38.78	62.83
69.98	83.19	83.38	19.72	49.34	15.50	46.56	75.16	22.40	17.55	70.99	58.04	41.62	69.23
65.90	86.27	84.53	26.09	72.57	0	51.78	74.96	0.95	0	65.23	48.51	33.64	75.02
71.97	83.33	81.80	16.77	50.66	0.88	52.37	72.61	21.42	6.64	72.15	59.10	49.15	72.37
70.61	82.86	81.69	17.13	78.53	26.28	55.80	80.69	31.83	26.49	66.23	64.65	46.39	42.97

Table 1. Quantitative comparison with different methods. We also show the precision of each class. Bold indicates the best results (%).

5. Experiments

5.1. Implementation Details and Metrics

We choose ResUNet [9] in our experiments, as it is stable and lightweight for land cover mapping. We train for 100 epochs for all experiments using stochastic gradient descent (SGD) with a momentum of 0.9 and a weight decay of 10^{-4}. The initial learning rate is set to 0.05. We adopt the batch size as 12, including 6 labeled images and 6 unlabeled images. The loss weight λ is set to 0.005. The temperature τ of contrastive loss is set to 0.07.

The performance of the proposed approach and other competing methods are assessed with overall accuracy (OA), mean accuracy (mAcc), mean intersection over union (mIoU), and precision of each class.

5.2. Comparison Results

Table 1 shows the comparison results with typical and state-of-the-art class-balanced pseudo-labeling methods. For a fair comparison, all competing methods and our

Method	OA	mAcc	mIoU
Random selection	75.50	51.35	39.09
Majority class -biased selection	75.75	49.98	39.33
Minority class -biased selection	**75.97**	**53.53**	**40.68**

Table 2. Ablation study on the proposed class-aware unlabeled data selection method. Bold indicates the best results (%).

Method	OA	mAcc	mIoU
Without class balance	75.41	48.16	37.27
Class-balanced method in [38]	73.61	51.33	38.61
Class-balanced method in [37]	75.51	49.66	35.51
Ours	**75.97**	**53.53**	**40.68**

Table 3. Comparison results on different class-balanced learning methods. Bold indicates the best results (%).

method use the same model architecture. All experiments are measured on the same computing platform. The baseline applies the supervised learning method to the labeled data. The SimCLR+finetune method uses a self-supervised learning method [7] on the unlabeled data to obtain a pretrain model and performs fine-tuning on the labeled data. Pseudo-labeling method [25] first generates pseudo-labels of the unlabeled data by using the baseline model, then retrains the model with ground truth and pseudo-labels. Advent [38], DPA [37], and CADR [17] are three state-of-the-art SSL methods.

The results demonstrate that our approach significantly improved over 4.28% in mAcc, 1.70% in mIoU, and 0.58% in OA. Compared to the baseline for each class, our method can significantly improve the accuracy of the minority classes and maintain the accuracy of the majority classes simultaneously. In contrast, the competing methods often tend to predict the pixels to majority classes and suffer from low precision on minority classes. Regarding the computational cost, our method incurs 53% additional training time than the basic SSL algorithm (i.e., Advent) due to Class-Balanced Learning and Class Center Contrast. The inference times are the same for all SSL methods, as we use the same backbone. Please find the visual comparison in the supplementary.

5.3. Ablation Study

Class-Aware Unlabeled Data Selection. We compare our method with the opposite strategy (i.e., majority class-biased data selection) and widely opted random selection method. Table 2 shows that our method achieves the best results. On the contrary, the mAcc value of the majority class-biased data selection method is unsatisfactory because the SSL model is biased toward predicting the majority classes.

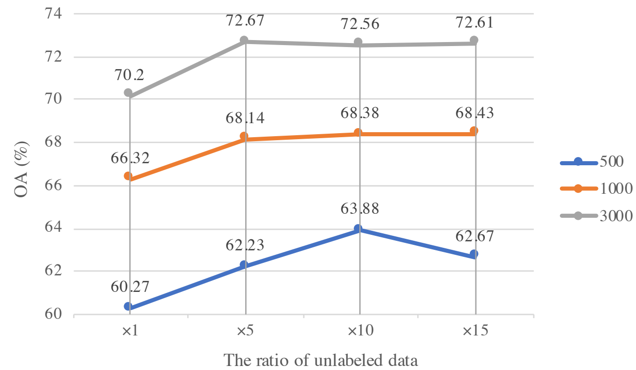

Figure 4. Results of using the different number of labeled images (i.e., 500, 1,000, and 3,000) and the different ratios of unlabeled to labeled images in SSL.

Besides, to verify the effectiveness of unlabeled data selection, we show the statistics of class distribution in Figure 3. Estimating the class distribution of unlabeled data is based on the pseudo-labels produced by the baseline model. Owing to the proposed class-aware unlabeled data selection method, the class distribution of the unlabeled data is more balanced than that of the labeled data. As a result, the entire training dataset is more balanced towards minority classes than the original labeled dataset. Therefore, we can estimate the class prior information on the entire dataset to partially remove the annotation bias.

Class-Balanced Learning. In general, class re-weighting sacrifices the accuracy of majority classes to improve the accuracy of minority classes, which decreases OA and mIoU. From Table 3, two comparison methods suffer from the above problem, while our method improves the mAcc, OA, and mIoU compared to the baseline (i.e., without loss weights). The reason is that the two comparison methods use class prior information of labeled data and ignore the discrepancy of class distributions between labeled and unlabeled data in SSL. Besides, Table 4 shows the ablation study on Class-Balanced Cross Entropy loss. Utilizing sample re-weighting improves performance by removing the annotation bias on the semi-supervised learning dataset. Using class re-weighting for cross-entropy loss removes class bias and improves the performance of mAcc and mIoU by a large margin. Furthermore, jointly utilizing the sample and class weights achieves the best results in Table 4.

Class Center Contrast Method. Table 5 shows the ablation study on the proposed class center contrast method. Contrastive loss utilizes inter-class and intra-class information and achieves better results than center loss used in [28]. Compared with only training contrastive loss with pseudo-labels on unlabeled data, using pseudo-labels on both labeled and unlabeled data achieves better results. Further-

Cross entropy loss	Sample re-weighting	Class re-weighting	OA	mAcc	mIoU
✓			75.41	48.16	37.27
✓	✓		75.74	48.96	38.40
✓		✓	75.43	52.06	40.07
✓	✓	✓	**75.97**	**53.53**	**40.68**

Table 4. Ablation study on class-balanced cross entropy loss. Bold indicates the best results (%).

Center loss	Contrastive loss	Pseudo label	Ground truth	On labeled data	On unlabeled data	OA	mAcc	mIoU
✓		✓			✓	74.07	48.13	37.34
	✓	✓			✓	75.23	50.06	39.29
	✓	✓		✓	✓	75.26	51.42	39.43
	✓	✓	✓	✓	✓	**75.97**	**53.53**	**40.68**

Table 5. Ablation study on the class center contrast method. Bold indicates the best results (%).

Method	OA	mAcc	mIoU
Random sample selection	74.68	49.31	38.80
Simple sample selection	74.74	51.57	40.42
Hard sample selection	74.57	51.14	40.31
Feature embedding centers	**75.97**	**53.53**	**40.68**

Table 6. Results of different contrastive sample selection manners. Bold indicates the best results (%).

Method	OA	mAcc	mIoU
Within-image method	73.36	53.48	40.03
Cross-image method	**75.97**	**53.53**	**40.68**

Table 7. Results of within-image and cross-image contrastive learning. Bold indicates the best results (%).

more, introducing the ground truth of labeled data to the training of contrastive loss can significantly improve performance. We also compare different sample selection strategies for contrastive learning. We select three contrastive samples for each class per image for a fair comparison. As shown in Table 6, leveraging high-confidence predicted samples from the unlabeled data for semi-supervised learning yields better results than employing uncertain samples. Nonetheless, samples with high-confidence predictions provide limited information for model learning, leading to only marginal enhancements. The proposed Class Center Contrast method involves both high-confidence and hard samples, achieving the best results. Besides, we compare within-image and cross-image contrastive learning results in Table 7. Cross-image method uses labeled and unlabeled images in a mini-batch and achieves better results, while the within-image method only uses local information of labeled or unlabeled images.

Hyperparameter Tuning of Loss Weight. λ is to balance two loss terms in the formula (7). Table 8 shows the

Loss weight	OA	mAcc	mIoU
$\lambda = 0.05$	75.07	51.62	40.15
$\lambda = 0.01$	75.26	51.42	39.43
$\lambda = 0.005$	**75.97**	53.53	**40.68**
$\lambda = 0.001$	75.27	**53.84**	40.26

Table 8. Hyperparameter tuning of loss weight. Bold indicates the best results (%).

results using different loss weights λ. We conclude that a large loss weight for $\mathcal{L}_{cct}$ reduces the model performance. In this work, we set $\lambda = 0.005$.

The Number and Ratio of Labeled Images in SSL. We conduct a comparative analysis of the enhancement in accuracy resulting from the increase of unlabeled data. As shown in Figure 4, the utilization of a larger quantity of unlabeled data, specifically ten times the amount of labeled training data containing only 500 images, has been observed to result in a substantial improvement in accuracy. Meanwhile, the performance becomes unstable as the ratio between unlabeled and labeled images increases further because of the limited number of labeled images. As the amount of labeled data increased, using the unlabeled data, which is five times the amount of labeled data, can yield superior results compared to using just one time the amount of labeled data. Subsequently, the performance is relatively stable as the number of unlabeled data increases. Therefore, to balance the computational cost and improvement in accuracy, we use a ratio of 5:1 for unlabeled and labeled data.

5.4. Examples of Land Cover Mapping in China

As an application of the proposed method, we produce large-scale land cover maps with 24 classes over 1 million km^2. As shown in Figure 5, we compare our land cover mapping results with two existing public products produced by ESA [45] and Google [2]. The example in the top row

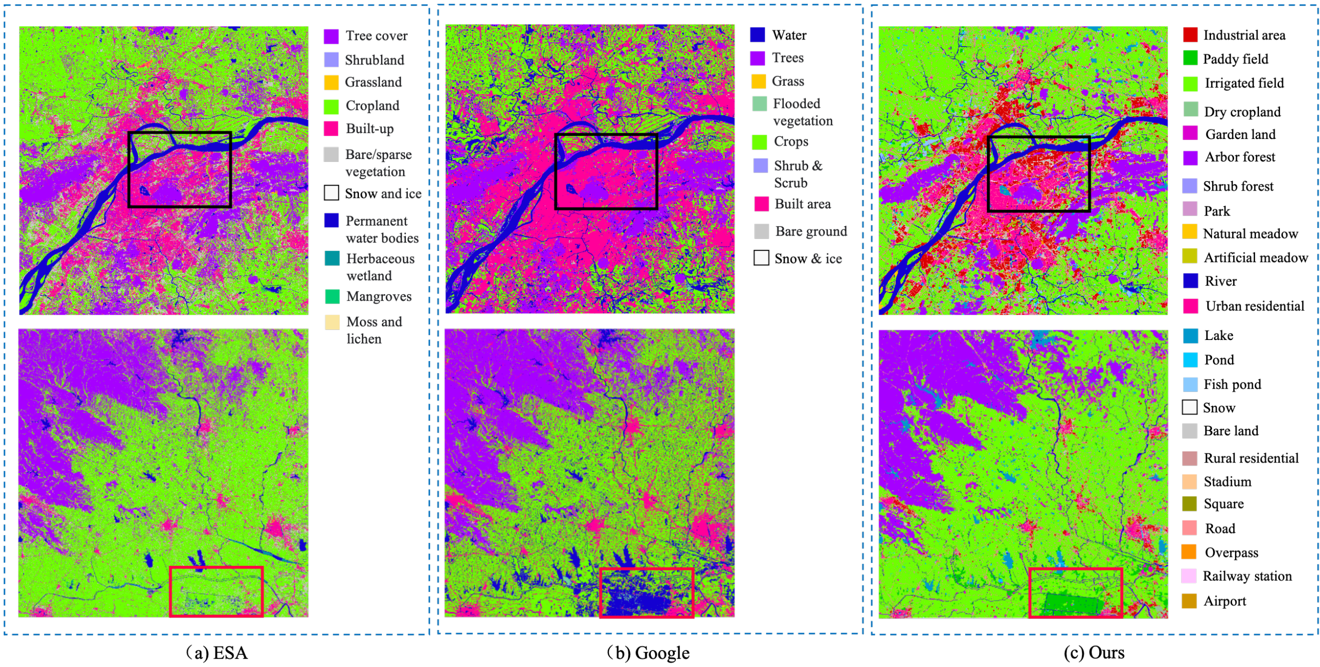

Figure 5. Examples of Land Cover Mapping in China. The land cover products of ESA and Google are also generated from Sentinel-2 images but use different classification systems. In the black rectangles, our land cover results with fine-grained classes can provide more land cover information. In the red rectangles, our method recognizes the paddy field correctly, while the other two products classify it as cropland and water, respectively, due to the coarse classification system and the misleading spectral information of the paddy field.

shows the results of an urban area located in the east of China, and that in the bottom row gives the results of a rural area located in the center of China. The land cover products of ESA and Google contain 11 and 9 classes, while our classification system includes 24 classes. From the black rectangles, the two comparison products mainly distinguish built areas, water, trees, and crops. Owing to the fine-grained classification system and the proposed class-aware SSL methods, our results can recognize more land cover categories, including industrial area, urban residential, rural residential, road, overpass, railway station, arbor forest, artificial meadow, irrigated field, bare land, lake, river, pond, and fish pond. Our land cover maps with fine-grained classes can provide more land cover information. Besides, from the red rectangles, our method recognizes the paddy field correctly, while the other two products classify it as cropland and water, respectively, due to the coarse classification system and the misleading spectral information of the paddy field. From the above qualitative comparison, our method provides land cover maps with fine-grained classes, significantly improving the land cover information.

6. Conclusion

In this work, we propose a unified Class-Aware Semi-Supervised Semantic Segmentation approach for large-scale land cover mapping with fine-grained classes. To address the class-imbalance issue, we propose a class-aware unlabeled data selection method to build an SSL dataset more balanced towards minority classes. Then, we propose a Class-Balanced Cross Entropy loss, considering the annotation and class biases on the built SSL dataset. Moreover, we propose the Class Center Contrast method to jointly use the labeled and unlabeled data for training. Experimental results validate the effectiveness of our method in large-scale land cover mapping with fine-grained classes.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China (Grant No. T2125006), Jiangsu Innovation Capacity Building Program (Project No. BM2022028), China Postdoctoral Science Foundation (Grant No. 2023M731871), the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (Grant No. 01DD20001), and Shuimu Tsinghua Scholar Project.

References

[1] Inigo Alonso, Alberto Sabater, David Ferstl, Luis Montesano, and Ana C Murillo. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8219–8228, 2021. 2

[2] Christopher F Brown, Steven P Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, Valerie J Pasquarella, Robert Haertel, Simon Ilyushchenko, et al. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1):251, 2022. 8

[3] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106:249–259, 2018. 2

[4] Jonathon Byrd and Zachary Lipton. What is the effect of importance weighting in deep learning? In *International Conference on Machine Learning*, pages 872–881, 2019. 2, 3

[5] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019. 2

[6] Bingyu Chen, Min Xia, and Junqing Huang. Mfanet: A multi-level feature aggregation network for semantic segmentation of land cover. *Remote Sensing*, 13(4):731, 2021. 2

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607, 2020. 6, 7

[8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 2

[9] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020. 6

[10] Runmin Dong, Weizhen Fang, Haohuan Fu, Lin Gan, Jie Wang, and Peng Gong. High-resolution land cover mapping through learning with noise correction. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2021. 2

[11] Runmin Dong, Cong Li, Haohuan Fu, Jie Wang, Weijia Li, Yi Yao, Lin Gan, Le Yu, and Peng Gong. Improving 3-m resolution land cover mapping through efficient learning from an imperfect 10-m resolution map. *Remote Sensing*, 12(9):1418, 2020. 2

[12] Runmin Dong, Lixian Zhang, Weijia Li, Shuai Yuan, Lin Gan, Juepeng Zheng, Haohuan Fu, Lichao Mou, and Xiao Xiang Zhu. An adaptive image fusion method for sentinel-2 images and high-resolution images with long-time intervals. *International Journal of Applied Earth Observation and Geoinformation*, 121:103381, 2023. 5

[13] Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016. 3

[14] Qian Gui, Xinting Wu, and Baoning Niu. Class-aware pseudo labeling for non-random missing labels in semi-supervised learning. In *2022 IEEE Eighth International Conference on Multimedia Big Data*, pages 138–143. IEEE, 2022. 2

[15] Lanzhe Guo and Yufeng Li. Class-imbalanced semi-supervised learning with adaptive thresholding. In *International Conference on Machine Learning*, pages 8082–8094, 2022. 2

[16] Hanzhe Hu, Jinshi Cui, and Liwei Wang. Region-aware contrastive learning for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16291–16301, 2021. 2

[17] Xinting Hu, Yulei Niu, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. On non-random missing labels in semi-supervised learning. In *International Conference on Learning Representations*, 2022. 2, 3, 4, 6, 7

[18] Muhammad Abdullah Jamal, Matthew Brown, Ming-Hsuan Yang, Liqiang Wang, and Boqing Gong. Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7610–7619, 2020. 2

[19] Noah Johnson, Wayne Treible, and Daniel Crispell. Opensentinelmap: A large-scale land use dataset using openstreetmap and sentinel-2 imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1333–1341, 2022. 2

[20] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. 2019. 2

[21] Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C Mazzariello, Mark Mathis, and Steven P Brumby. Global land use/land cover with sentinel 2 and deep learning. In *2021 IEEE international Geoscience and Remote Sensing Symposium IGARSS*, pages 4704–4707, 2021. 1

[22] Salman H Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3573–3587, 2017. 2

[23] Donghyeon Kwon and Suha Kwak. Semi-supervised semantic segmentation with error localization network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9957–9967, 2022. 2

[24] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319, 2018. 5

[25] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks.

In *Workshop on challenges in representation learning, International Conference on Machine Learning*, volume 3, page 896, 2013. 6, 7

[26] Ruihuang Li, Shuai Li, Chenhang He, Yabin Zhang, Xu Jia, and Lei Zhang. Class-balanced pixel-level self-labeling for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11593–11603, 2022. 5

[27] Weijia Li, Runmin Dong, Haohuan Fu, Jie Wang, Le Yu, and Peng Gong. Integrating google earth imagery with landsat data to improve 30-m resolution land cover mapping. *Remote Sensing of Environment*, 237:111563, 2020. 1

[28] Zhuohong Li, Hongyan Zhang, Fangxiao Lu, Ruoyao Xue, Guangyi Yang, and Liangpei Zhang. Breaking the resolution barrier: A low-to-high network for large-scale high-resolution land-cover mapping using low-resolution labels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:244–267, 2022. 7

[29] Xiaoqiang Lu, Licheng Jiao, Fang Liu, Shuyuan Yang, Xu Liu, Zhixi Feng, Lingling Li, and Puhua Chen. Simple and efficient: A semisupervised learning framework for remote sensing image semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. 2

[30] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 2

[31] Prem Chandra Pandey, Nikos Koutsias, George P Petropoulos, Prashant K Srivastava, and Eyal Ben Dor. Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto International*, 36(9):957–988, 2021. 1

[32] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *International Conference on Machine Learning*, pages 4334–4343, 2018. 2

[33] Caleb Robinson, Le Hou, Kolya Malkin, Rachel Soobitsky, Jacob Czawlytko, Bistra Dilkina, and Nebojsa Jojic. Large scale high-resolution land cover mapping with multi-resolution data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12726–12735, 2019. 2

[34] Vladan Stojnic and Vladimir Risojevic. Self-supervised learning of remote sensing scene representations using contrastive multiview coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1182–1191, 2021. 2

[35] Jingru Tan, Changbao Wang, Buyu Li, Quanquan Li, Wanli Ouyang, Changqing Yin, and Junjie Yan. Equalization loss for long-tailed object recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11662–11671, 2020. 2

[36] Aysim Toker, Lukas Kondmann, Mark Weber, Marvin Eisenberger, Andrés Camero, Jingliang Hu, Ariadna Pregel Hoderlein, Çağlar Şenaras, Timothy Davis, Daniel Cremers, et al. Dynamicearthnet: Daily multi-spectral satellite dataset for semantic change segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21158–21167, 2022. 2

[37] Xinyi Tong, Guisong Xia, and Xiao Xiang Zhu. Enabling country-scale land cover mapping with meter-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:178–196, 2023. 1, 2, 5, 6, 7

[38] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 2, 6, 7

[39] Wenguan Wang, Tianfei Zhou, Fisher Yu, Jifeng Dai, Ender Konukoglu, and Luc Van Gool. Exploring cross-image pixel contrast for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7303–7313, 2021. 2, 5

[40] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022. 2

[41] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10857–10866, 2021. 2, 3

[42] Linshan Wu, Leyuan Fang, Xingxin He, Min He, Jiayi Ma, and Zhun Zhong. Querying labeled for unlabeled: Cross-image semantic consistency guided semi-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[43] Fan Yang, Kai Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14421–14430, 2022. 2

[44] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4268–4277, 2022. 2

[45] D Zanaga, R Van De Kerchove, W De Keersmaecker, N Souverijns, C Brockmann, R Quast, J Wevers, A Grosu, A Paccini, S Vergnaud, et al. Esa worldcover 10 m 2020 v100. zenodo, 2021. 8

[46] Yuanyi Zhong, Bodi Yuan, Hong Wu, Zhiqiang Yuan, Jian Peng, and Yu-Xiong Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7273–7282, 2021. 2

[47] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9719–9728, 2020. 2