FALSE: False Negative Samples Aware Contrastive Learning for Semantic Segmentation of High-Resolution Remote Sensing Image

Zhaoyang Zhang, Xuying Wang, Xiaoming Mei, Chao Tao, and Haifeng Li, Member, IEEE

Abstract—Self-supervised contrastive learning (SSCL) is a potential learning paradigm for learning remote sensing image (RSI)-invariant features through the label-free method. The existing SSCL of RSI is built based on constructing positive and negative sample pairs. However, due to the richness of RSI ground objects and the complexity of the RSI contextual semantics, the same RSI patches have the coexistence and imbalance of positive and negative samples, which causes the SSCL pushing negative samples far away while pushing positive samples far away, and vice versa. We call this the sample confounding issue (SCI). To solve this problem, we propose a False negative sampLes aware Contrasitive LEarning model (FALSE) for the semantic segmentation of high-resolution RSIs. Since SSCL pretraining is unsupervised, the lack of definable criteria for false negative sample (FNS) leads to theoretical undecidability, and we designed two steps to implement the FNS approximation determination: coarse determination of FNS and precise calibration of FNS. We achieve coarse determination of FNS by the FNS self-determination (FNSD) strategy and achieve calibration of FNS by the FNS confidence calibration (FNCC) loss function. Experimental results on three RSI semantic segmentation datasets demonstrated that the FALSE effectively improves the accuracy of the downstream RSI semantic segmentation task compared with the current three models, which represent three different types of SSCL models. The mean intersection over union (mIoU) on the ISPRS Potsdam dataset is improved by 0.7% on average; on the CVPR DGLC dataset, it is improved by 12.28% on average; and on the Xiangtang dataset, this is improved by 1.17% on average. This indicates that the SSCL model has the ability to self-differentiate FNS and that the FALSE effectively mitigates the SCI in SSCL.

Index Terms—False negative sample (FNS), remote sensing image (RSI), self-supervised contrastive learning (SSCL), semantic segmentation.

I. INTRODUCTION

DEEP neural networks (DNNs) trained in a supervised learning manner have made remarkable progress in remote sensing image (RSI) scene classification [1], target detection [2], and semantic segmentation [3], [4]. The dependence of this approach on massive, high-quality labeled samples has become a bottleneck for wide-scale application [5], [6], [7]. The promise of self-supervised contrastive learning (SSCL) has made it possible to learn the RSI-invariant features from massive unlabeled data [8], [9], [10], [11].

The core idea of SSCL is to cleverly obtain copies of the same image patches as positive samples and other different images as negative samples by the data augmentation methods of spatial and spectral transformations such as rotation, scaling, random color distortion, and Gaussian blur [12] and to construct self-supervised signals by pulling the positive samples closer while pushing the negative samples farther away. This is done instead of manually labeling them as supervised signals, thus forcing DNNs to obtain spatial- and spectral-invariant representations [11], [13], [14], [15].

However, due to scene complexity, ground object richness, and unbalanced distribution of samples of RSI, there is a phenomenon in which positive and negative samples coexist in the same patch and are highly unbalanced. As shown in Fig. 1, the false negative samples (FNSs) for the selected anchor sample in the last two columns contain the same ground features as the positive sample, causing the SSCL pushing negative samples far away while pushing positive samples far away, and vice versa. We call this the sample confounding issue (SCI). The performance loss of the model due to SCI is called the sample confounding effect (SCE), while the negative sample image patch containing positive samples is called the FNS because it gives the wrong feedback signal to the model.

The current methods to solve the SCE problem in SSCL are mainly considered from the perspective of samples and can be divided into two categories. One category aims to improve the quality of negative sample construction by attaching other unsupervised methods to the original SSCL and using the additional unsupervised results to guide the self-supervised
model to construct higher quality positive and negative sample pairs [16], [17], [18]. However, it is often difficult to improve the construction quality of positive and negative sample pairs in RSI processing using additional unsupervised clustering methods. Nevertheless, this may introduce defects of related unsupervised methods because the ground objects in RSI often have problems such as sample imbalance, intraclass differences, and interclass similarity, leading to the ineffectiveness of unsupervised clustering methods [19]. The second category considers abandoning the construction of negative samples [20], [21], [22], which means that the model’s performance will depend only on the construction of positive samples. Considering that the FNS is essentially positive samples in the dataset, this type of approach completely avoids the generation of FNS. Nevertheless, it also means that the model will not use the positive samples that already exist in the dataset, which may reduce the model’s ability to extract RSI-invariant features.

Unlike the above methods, our observation is as follows: the SSCL of the RSI model itself can distinguish between true and FNSs. This ability comes from the correct self-supervised signals given to the model by the true positive and true negative samples (TNSs). This ability is potentially reinforced continuously as the model is trained. We refer to this as the ability of FNS self-determination (FSD). This observation motivates us to rethink the SCE problem from the perspective of the model rather than directly from the perspective of the sample.

The fundamental difficulty of using FSD to determine the FNS is that self-supervised pretraining is essentially an unsupervised process. The lack of definable criteria for the FNS leads to theoretical undecidability, so we can only approximately determine the FNS by some strategy. Approximate determination of FNS can be divided into two steps in terms of process: coarse determination of FNS and precise calibration of FNS. The former is the initial screening of FNS to ensure completeness, and the latter is the precise selection based on the former to ensure accuracy.

We propose the False negAtive sampLes aware contraStive lEarning model (FALSE), which achieves the coarse determination of FNS through the FNS self-determination (FNSD) strategy and achieves the precise calibration of FNS by designing the FNS confidence calibration (FNCC) loss function. In the FNSD strategy, the anchor sample in the closer positive sample pair in the embedding space is used as the benchmark, and the negative sample with the highest similarity to the anchor sample is determined as the possible FNS. The FNCC loss function is designed to improve the contribution of the possible FNS to the positive sample term of the original contrastive loss function [23] and reduce its contribution to the negative sample term of the loss function to mitigate SCE in the SSCL model. The contributions in this letter are as follows.

1) We propose a FALSE for the semantic segmentation of high-resolution RSIs. FALSE determines the approximate determination of FNS in SSCL from the perspective of the model rather than samples and mitigates the SCI in SSCL of RSIs.
2) We designed the FNCC loss function quantitatively rather than qualitatively to characterize the strength of the ability of FNS self-determination (FSD) in the form of confidence weights.

3) The experimental results on three semantic segmentation datasets show that FALSE relative to SimCLR, prototypical contrastive learning (PCL), and Barlow twins improves the mean intersection over union (mIoU) on the ISPRS Potsdam dataset by 0.7% on average, improves the mIoU by 12.28% on average on the CVPR DGLC dataset, and improves the mIoU by 1.17% on average on the Xiangtan dataset.

II. METHODOLOGY

In the SSCL of RSI, the presence of true positive sample (TPS, blue dots in Fig. 2) and TNS (TNS, red dots in Fig. 2) will give the model a correct self-supervised signal about the RSI-invariant features. In contrast, FNS (FNS, pink dots in Fig. 2) in the negative sample set will give the model an incorrect signal about the RSI-invariant features, creating the SCI in SSCL.

Since SSCL pretraining is essentially an unsupervised process, the lack of definable criteria for FNS leads to theoretical undecidability, so we can only approximately determine the FNS by some strategies. The approximate determination of FNS can be divided into two steps: 1) coarse determination of FNS and 2) precise calibration of FNS. The former is the initial screening of FNS to ensure the completeness of FNS; the latter is precise on the former to ensure the accuracy of FNS selection.

A. Coarse Determination of FNS

1) Determination Benchmark: Since the goal of the SSCL model is to bring positive samples closer and push negative samples farther, if the model projects a positive sample pair to a closer location in the embedding space, then the model is currently better at learning invariant features about that positive sample pair. Based on this, the anchor sample of the closer positive sample pair in the embedding space is selected as the benchmark for determination, maximizing the use of the feature extraction information that the model has learned, thus minimizing the model’s misjudgment.

Suppose the benchmark anchor sample is denoted as \( o_{\text{key}} \), its corresponding positive sample is denoted as \( p \), and \( \text{sim}(\cdot, \cdot) \) denotes the calculation of the feature similarity between the two samples. Then, \( o_{\text{key}} \) satisfies the condition that

\[
\text{sim}(o_{\text{key}}, p) > T.
\]

In (1), \( T \) denotes the positive sample pair similarity threshold, which controls the proximity of positive sample pairs in the embedding space.

2) Determination Condition: Based on the determination benchmark satisfying 1, we calculate the similarity between all the negative samples and the benchmark anchor sample \( o_{\text{key}} \) in the embedding space and determine the negative sample with the highest similarity to the benchmark anchor sample \( o_{\text{key}} \) as the possible FNS. Suppose \( n \) is used to denote the negative samples to be judged, and \( n_{pf} \) denotes the possible FNS, the above determination condition can be simply described as

\[
|\text{sim}(o_{\text{key}}, n_{pf}) - \text{sim}(o_{\text{key}}, p)| \\
\rightarrow \min |\text{sim}(o_{\text{key}}, n) - \text{sim}(o_{\text{key}}, p)|.
\]
3) Possible FNS Analysis: Influenced by the SCI, the possible FNS obtained by the FSD is not all FNS but also contains TNS. Nevertheless, since \( o_{\text{key}} \) represents the best level of the current model’s ability to extract image features, it is difficult to use the model’s FSD ability to eliminate this part of the TNS from the set of possible FNS. Noting that the indistinguishability of such negative samples is for the model, we refer to this part of the TNS among the possible FNS as the model-dependent hard negative sample (MD-HNS, brown dots in Fig. 2) in the SSCL. Suppose \( n_f \) is used to denote the FNS and \( n_h \) to denote the MD-HNS, the composition of the possible FNS obtained can be simply described as

\[
n_{pf} = n_f + n_h.
\]  

B. Precise Calibration of FNS

1) FNCC Loss Function: To calibrate the obtained possible FNS and mitigate the impact of MD-HNS on the model performance, we design the FNCC loss function by introducing confidence weights \( \alpha \) to calibrate the possible FNS as positive samples.

The original loss function of the SSCL of the RSI model mainly consists of two parts [23]: positive sample term \( e_{\text{sim}}(o, p) \) and negative sample term \( \sum_{i} e_{\text{sim}}(o, n_i) \).

The FNCC multiplies the similarity of possible FNS by the confidence weight \( \alpha \) to get \( S_{EP} \) [see (4)], and adds it to the original positive sample term of SSCL loss, increases the influence of \( n_{pf} \) on the positive sample term of the loss to enhance the correct signal, multiplies the similarity of possible FNS by \( 1 - \alpha \) to get \( S_{WN} \) [see (5)], replaces the similarity corresponding to the possible FNS in the original negative sample term with \( S_{WN} \), and reduces the influence of \( n_{pf} \) on the negative sample term of the loss to weaken the incorrect signal.

When the number of negative samples corresponding to an anchor sample is \( N \), the number of possible FNS determined to be obtained is \( N_{pf} (N_{pf} < N) \), and the FNCC loss function is defined by (6)

\[
L_{\text{FNCC}} = -\log \frac{e_{\text{sim}}(o, p) + S_{EP}}{e_{\text{sim}}(o, p) + S_{EP} + \sum_{i=N_{pf}+1}^{N} e_{\text{sim}}(o, n_i) + S_{WN}}.
\]  

In particular, when there is no possible FNS, \( N_{pf} = 0 \), and (5) degenerates to the original loss function and becomes the original SSCL model.

2) Meaning of Confidence Weights: The confidence weight \( \alpha \) represents the degree of confidence in the model’s FSD ability. When \( \alpha = 0 \), the positive sample signal enhancement term and the FNS signal weakening term of the FNCC loss are both 0. The model is the original SSCL model, and the FNSD strategy is not used. When \( \alpha = 1 \), it means that FALSE fully trusts the possible FNS obtained from the model FSD and adjusts all possible FNSs to the positive sample, eliminating the contribution of these possible FNS to the negative sample term of the FNCC. When \( \alpha \) takes any value less than 1 and greater than 0, it means that the model increases the contribution of possible FNS to the positive sample term of the FNCC from 0 to \( \alpha \) times the original contribution to the negative sample term and weakens its contribution to the negative sample term to \( 1 - \alpha \) times the original contribution.

### III. Experimental

#### A. Datasets

The experiments were selected from the public RSI semantic segmentation dataset ISPRS Potsdam [24], competition dataset CVPR DGLC [25], and Xiangtan dataset [10] from the Gaofen-2 satellite covering Xiangtan, China. The spatial resolution and number of ground object types of the three datasets are shown in Table I.

#### B. Experimental Setup

The experiments follow the general paradigm of the SSL models [9], [11], [13] and are divided into two main steps:
self-supervised pretraining and supervised fine-tuning. Self-supervised pretraining uses all the unlabeled training set data, pretrained for 200 epochs with the batch size set to 256. Next, the entire model encoder is frozen, and a small amount of labeled data is used to fine-tune and train the decoder, with the labeled data selected as 1% of the pretrained unlabeled data. The decoder is finally used to obtain the RSI semantic segmentation results.

To quantitatively explore the FSD of the FALSE, we conducted several experiments using the introduced confidence weights $\alpha$, with the positive sample pair similarity threshold $T$ set to 0.9. The results are presented in Experiment I.

On this basis, we selected SimCLR, which represents the original SSCL model, PCL, which represents the SSCL model that joins clusters, and the Barlow twins, which represents the SSCL model without constructing negative samples. These are used as the baseline tested on the ISPRS Potsdam, CVPR DGLC, and Xiangtan datasets, which are then compared with the FALSE model having a confidence weight of 1. The results are presented in Experiment II.

### C. Experimental Results and Analysis

1) **Experiment I, Analysis of Confidence Weight:** We selected five values that were uniformly distributed in the interval from 0 to 1: 0, 0.3, 0.5, 0.7, and 1. The experiments were repeated five times for each confidence weight while keeping the other parameters consistent.

Fig. 3 shows the semantic segmentation mIoU results of the FALSE model on the ISPRS Potsdam, CVPR DGLC, and Xiangtan datasets with five different confidence weights. The model performed best when fully trusting the possible FNS obtained by the model’s FSD ($\alpha = 1$). Compared with the original SSCL model, FALSE improves the mIoU by 0.72% on the ISPRS Potsdam dataset, 0.8% on the CVPR DGLC dataset, and 0.56% on the Xiangtan dataset. Moreover, when the model partially trusted the possible FNS obtained from the model’s FSD ($0 < \alpha < 1$), the semantic segmentation performance became less stable. Considering that a confidence weight between 0 and 1 reduces the contribution of possible FNS to both positive and negative sample terms of the FNCC loss function, this phenomenon implies that the possible FNS (including FNS and MD-HNS) affects the stability of the semantic segmentation performance of the FALSE model.

Tables II–IV show the IoU of the FALSE model for various types of ground objects on the three datasets with five different confidence weights.
TABLE V

| Method          | ISPRS Potsdam | CVPR DGLC | Xiangtan |
|-----------------|---------------|-----------|----------|
|                 | OA  | mIoU  | mAcc  | OA  | mIoU  | mAcc  | OA  | mIoU  | mAcc  |
| SimCLR          | 59.63 | 42.03 | 0.54   | 26.59 | 60.49 | 70.41 | 59.20 | 35.41 | 41.25 |
| PCL             | 60.22 | 42.25 | 54.20  | 77.36 | 60.54 | 72.56 | 77.85 | 35.49 | 41.55 |
| Barlow twins    | 60.03 | 41.87 | 54.06  | 66.71 | 32.45 | 39.28 | 77.57 | 33.49 | 39.33 |
| FALSE (ours)    | 60.46 | 42.75 | 54.77  | 81.47 | 61.29 | 71.43 | 79.64 | 35.97 | 41.60 |

For the classification, compared with five different confidence weight models, the FALSE model with confidence weight 1 achieved the best mIoU for five of all six types of ground features on the Potsdam dataset, the best mIoU for three and the second-ranked mIoU for three of all seven types of ground features on the DGLC dataset, and the best mIoU for four and the second-ranked mIoU for three of all eight types of ground features on the Xiangtan dataset.

2) Experiment II, Comparison of Four Types of SSCL Models for Semantic Segmentation: Table V shows the semantic segmentation results of the four different types of positive and negative sample construction strategy SSL models on the ISPRS Potsdam, CVPR DGLC, and Xiangtan datasets. The FALSE model achieves the best semantic segmentation results on these three datasets. Its overall accuracy (OA), mIoU, and mean class accuracy (mAcc) outperform SimCLR, which represent the original SSCL model, PCL, which represents the SSL model that joins clusters, and the Barlow twins, which represents the SSL model without constructing negative samples.

IV. CONCLUSION

In this letter, we proposed the FALSE for the semantic segmentation of high-resolution RSIs. Under the restriction that self-supervised pretrained FNSs are theoretically undecidable, the FALSE model achieves approximate determination of the FNS by coarse determination and precise calibration of FNS and quantitatively characterizes the ability of FNS self-determination (FSD) using confidence weights. The experiments on three RSI semantic segmentation datasets showed that FALSE effectively alleviates the SCE caused by SCI in the original SSCL of RSI. Compared with SimCLR, which represents the original SSCL model, PCL, which represents the SSL model that joins clusters, and the Barlow twins, which represents the SSL model without constructing negative samples, FALSE improves the mIoU by 0.7% on average on ISPRS Potsdam, improves the mIoU by 12.28% on average on DGLC CVPR2018, and improves the mIoU by 1.17% on average on Xiangtan.

The current method is only a simple implementation of the model’s FSD, introducing manually set confidence weights. Through the experiment, we found that the confidence weight corresponding to the best segmentation accuracy of different ground objects is not the same, so how to adjust the confidence weight adaptively for different ground objects in the dataset and give full play to the model’s ability of FSD is a further issue to be considered in the future for the FALSE model.