SemiSiROC: Semisupervised Change Detection With Optical Imagery and an Unsupervised Teacher Model

Lukas Kondmann, Sudipan Saha, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*—Change detection (CD) is an important yet challenging task in remote sensing. In this article, we underline that the combination of unsupervised and supervised methods in a semisupervised framework improves CD performance. We rely on half-sibling regression for optical change detection (SiROC) as an unsupervised teacher model to generate pseudolabels (PLs) and select only the most confident PLs for pretraining different student models. Our results are robust to three different competitive student models, two semisupervised PL baselines, two benchmark datasets, and a variety of loss functions. While the performance gains are highest with a limited number of labels, a notable effect of PL pretraining persists when more labeled data are used. Further, we outline that the confidence selection of SiROC is indeed effective and that the performance gains generalize to scenes that were not used for PL training. Through the PL pretraining, SemiSiROC allows student models to learn more refined shapes of changes and makes them less sensitive to differences in acquisition conditions.**

***Index Terms*—Change detection (CD), multitemporal, optical images, semisupervised, unsupervised.**

I. Introduction

CHANGE detection (CD) is the task of segmenting changing pixels over time in multitemporal Earth observation data. In the face of a changing planet, CD is at the core of many relevant monitoring tasks. It allows us to study the temporal evolution of forests [1], [2], [3], urban areas [4], [5], coastal and maritime regions [6], [7], and the effects of natural disasters [8], [9], [10], [11], [12]. CD methods face a number of hurdles related to the acquisition conditions between the different times the images are collected. This includes but is not limited to illumination conditions, clouds and shadows, acquisition angles, and the definition of what constitutes a change [13]. Despite these challenges, several trends have been beneficial for the methodological progress in CD in recent years. First, open data policies, for example, in the Copernicus program [14] increase accessibility and availability of multitemporal Earth observation data [15]. Second, technological progress results in increasing spatial and temporal resolution of satellite data with up to daily imagery [16]. Third, methodological progress in image recognition, particularly, deep learning [17], has also fueled a variety of improvements in artificial intelligence for Earth observation including CD [18], [19], [20].

Many recent advances are in supervised learning for binary CD from optical imagery [21], [22], [23], [24], [25], [26], [27], [28], [29]. Following the success of convolutional neural networks (CNNs) in a variety of computer vision problems [17], CNNs have been used frequently for CD problems as well. Daudt et al. [23] introduce a siamese CD architecture inspired by UNet [30]. ESCNET is a combination of superpixel enhancement and a deep CNN [31]. For CD in aerial images, Xu et al. [32] design a pseudosiamese capsule network.

More recently, the success of vision transformers [33], [34] has induced increasing attention also from the remote sensing community. For example, Bandara and Patel [21] design ChangeFormer, a siamese transformer network for building CD. In a similar spirit, Chen et al. [25] employ a self-attention-based transformer method. Further, many approaches also combine convolutional and attention-based approaches with promising results [22], [35], [36].

However, obtaining large-scale labeled data for CD remains a challenge. Unsupervised CD methods [37], [38], [39], [40], [41], therefore, learn without labeled data to circumvent this issue. Many methods also utilize the advances in deep learning for unsupervised CD. For example, Saha et al. introduce deep change vector analysis (DCVA) for high-resolution imagery, which combines ideas from classical image differencing with a deep convolutional feature extractor [37]. DCVA has also been further extended in combination with self-supervised pretraining [42] and refined further for medium-resolution images [38]. A generative approach is used in [43] to model the different image in an unsupervised fashion. Zhan et al. [44] rely on an initial classification of changing superpixels with a fully CNN. These superpixels are then categorized by uncertainty and used to train a classifier in a second step.

Still, in unsupervised CD, particularly with lower resolution, many methods reach high performance also without the use of

Manuscript received 17 October 2022; revised 26 January 2023 and 24 February 2023; accepted 29 March 2023. Date of publication 20 April 2023; date of current version 28 April 2023. This work was supported in part by the Helmholtz Association through the joint research school "Munich School for Data Science - MUDS" and Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research" under Grant W2-W3-100, and in part by 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" under Grant 01DD20001. *(Corresponding author: Xiao Xiang Zhu.)*

Lukas Kondmann is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: Lukas.Kondmann@dlr.de).

Sudipan Saha is with the Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: sudipan.saha@scai.iitd.ac.in).

Xiao Xiang Zhu is with the Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: xiaoxiang.zhu.ieee@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3268104

deep features. Sibling regression for optical change detection (SiROC) [39] is inspired by exoplanet search and compares pixels against their distant neighborhood to identify changes in optical imagery. Furthermore, image differencing also called change vector analysis [45] and its extensions [46], [47], [48] still play a role in practice.

Semisupervised approaches bridge the gap between unsupervised and supervised approaches. These methods try to combine labeled data with larger amounts of unlabeled data to support the training process. Among the first to apply semisupervised learning in CD were Bovolo et al. [49]. They use a Bayesian thresholding mechanism to set up an adequately defined binary semisupervised support vector machine (S^3VM). Modified self-organizing feature App (SOFM) uses only a limited set of initial labels to compute soft labels for unlabeled additional input [50]. Chen et al. [51] rely on probabilistic Gaussian processes (GP) as a first step with labeled and unlabeled data. The outputs of the GP classifier are then refined with a Markov random field regularizer. A Laplacian regularized metric learning mechanism is used in [52] to exploit unlabeled training data at scale for hyperspectral image CD. For very high spatial resolutions, graph convolutional networks (GCNs) are also effective for semisupervised learning by encoding multitemporal images as a graph [53].

One particularly effective direction in semisupervised learning in general image recognition is student–teacher models [54]. Typically, there is a teacher model that is trained on labeled data and predicts additional labels for images where ground truth is not available. Then, a student model uses these additional labels, referred to as pseudolabels (PL), during the training. With Earth observation data, PLs have also been shown to be effective for hyperspectral image classification [55]. PLs are also related to unsupervised CD approaches for small scenes, which rely on an initial difference image or change classification and finetune this further with another unsupervised method [43], [56], [57]. This is similar to using PLs although these approaches are purely unsupervised and are applied only to single scenes instead of large-scale training. Li et al. [58] use PLs explicitly for CD in SAR images but stay in the unsupervised domain. Similarly, Gao et al. [59] train convolutional wavelet neural networks with automatically generated labels for sea ice CD with SAR images.

In many student–teacher settings, the actual labels are used at least in some capacity in the pseudolabeling. However, this can be somewhat challenging in scenarios with limited labels as in CD. Additionally, applications of methods in regions outside their training data often require some robustness to unseen regions [60]. In this article, we therefore propose SemiSiROC where we use an unsupervised method with well-calibrated uncertainties for PL training. The uncertainty score for each prediction allows us to filter only high-quality PL for pretraining. In the second step of the semisupervised method, we finetune student models with the actual labels to improve optical CD performance. We evaluate our results on a binary version of the DynamicEarthNet benchmark [61] as well as the OSCD dataset [24] and compare the effectiveness of our strategy with five competitive CD models as students: ChangeFormer [21], BIT [25], DTCDSCN [29], FC-Siam-Diff [23], and FC-Siam-Conc [23]. Although SemiSiROC is most effective in limited label scenarios, we also find that even with a sizeable amount of 1000 labeled image pairs, SemiSiRoc boosts performance for all tested models notably. While student–teacher models themselves are not new in remote sensing, our ingenuity lies in the components specifically designed for CD on large-scale datasets and further validation on a global dataset of such scale. We have three main contributions.

1) We present SemiSiROC, a semisupervised CD method in optical remote sensing that combines advanced supervised models with unsupervised pseudolabeling.
2) Building on the confidence filtering of SiROC, we devise a mechanism to prioritize relevant scenes during PL filtering.
3) We propose a detailed experimental setup for CD subject to geographic disparity, based on the recently launched publicly available DynamicEarthNet dataset [61]. This experimental setup will be helpful for other researchers to pursue research in this direction. Our experiments on this setup and the OSCD [23] benchmark show that semisupervised learning is indeed helpful.

II. Method

A. SemiSiROC

Let us assume, we have two different collections of images, D and U. D is a collection of N_D bi-temporal pairs with associated pixelwise change/unchanged label. On the other hand, U is a collection of N_U unlabeled bitemporal pairs. Generally $N_U > N_D$, however this is not a strict assumption. The U and D can be acquired over different geographic areas/continents, thus they need not be representing the same geographic distribution. Our goal is to exploit both D and U to learn a CD model. Toward this, we design a semisupervised pipeline that allows exploiting U for model training even if labels for it are not available. We exploit a teacher–student model where the teacher model labels the images and selects relevant samples from U. This allows its student to exploit the label space $D \cup U$ instead of D. Therefore, we train with PLs first before we go on to real labels. This is consistent with semisupervised literature [62] and has the underlying assumption that the model can immensely benefit from PLs as a first step of training, which can be subsequently refined with actual labels.

The PLs for pretraining are based on SiROC [39], an unsupervised method for optical CD. We average the confidence on the cube level and as a default choice use the top 25%. Then, we train a student model with the preselected locations and PLs first before finetuning with the actual labels. Since the teacher model exploits SiROC in a semisupervised setting, we call our approach SemiSiROC.

Algorithm 1 outlines SemiSiROC in pseudocode in more depth. Given the unlabeled collection U, the labeled collection D, the corresponding labels L, and a supervised CD model, the desired output is a binary change segmentation. At first, we

Algorithm 1: SemiSiROC.

Input: U, D, L, model
Output: Binary Change Segmentation
C = [],P = []
for (u in U) **do**
  P_u, C_u = SiROC(u)
  C.append(C_u)
  P.append(P_u)
end for
$U_P = C_P$.top_quarter(C)
$P_P = P$.top_quarter$C$)
model.train(U_P, P_P) {PL training}
model.train(D, L) {Finetuning}

define a collection of confidence scores C and PLs P. Then, we loop over the elements of U and obtain PLs and confidence scores with SiROC for each image pair. Before semisupervised pretraining, we filter P and U to only use the scenes with the highest confidence, which is defined as U_P. These scenes are used as input for the pretraining of the CD model before training with actual labels in the final step.

While the proposed SemiSiROC approach is similar to many semisupervised learning strategies [62], note that our approach is distinct in the following three ways:

1) how we generate the PLs with an unsupervised CD method;
2) how we select the samples for student training based on a well-calibrated uncertainty;
3) how we exploit them for global CD.

B. Unsupervised Teacher Model

The goal of the teacher model is to assign PLs to some samples from U with reasonable confidence that they can be used later for training the CD (student) model. Since U and D may not necessarily be from the same distribution, the teacher model may use its learning from D and bias the distribution of PLs for U by overfitting to D. This is particularly relevant in the geo context where different locations and points in time can quickly change the data-generating distribution [63]. We argue that the teacher label should refrain from using the actual labels in any form to obtain the PLs. If the PL extraction process uses the actual labels, this would make them interdependent and hamper generalization. Thus, the teacher model should be based on unsupervised learning in this case. Additionally, semisupervised pretraining is more flexible with unsupervised PLs and our pretrained model can serve as a starting point for other CD applications without the need to retrain the teacher model on new datasets with new labels to obtain other PLs. Therefore, we propose to use an unsupervised teacher model to incentivize more robustness to spatial generalization in the PLs. This is in contrast to many other semisupervised approaches with PLs, which rely on teacher models that have seen at least some of the actual labels [62].

As unsupervised teacher model, we employ SiROC [39]. While the method is highly performant, we pick it as a PL source or so-called teacher model mainly because it comes with a built-in, well-calibrated confidence score ranging from 0 (low) to 1 (high) with its prediction for each pixel. This allows us to filter PLs based on their confidence and only train on high confidence labels. As this confidence score is closely connected to the quality of the PL, we hypothesize that algorithms should learn better with selected PLs only. Out of N_U total samples in U, N'_U are chosen after confidence filtering for pretraining. In the following, we explore SiROC in more depth.

1) Sibling Regression for Optical Change (SiROC): SiROC models a pixel as a linear combination of a set of neighboring pixels n at a certain time t in a time series. At time $t+1$, the value of the respective pixel is predicted based on the neighbors n at $t+1$. The deviation between the actual and the predicted pixel value is interpreted as a change signal. If the difference is high, this is seen as an indication of change as the pixel seems to have undergone a change compared to its neighborhood. The comparison against the neighborhood serves to eliminate local or image-wide trends as sources of false positives for changes.

More formally, given a channel of a multispectral image I at time t and $t+1$, the core of the predicted change segmentation $\hat{P}$ is based on the following equation:

$$\hat{P} = \begin{cases} 1, & \text{if } \hat{\mathbf{I}}_{\mathbf{t+1}} - \mathbf{I}_{\mathbf{t+1}} > o \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where o is the Otsu threshold [64]. $\hat{I}_{t+1}$ is the predicted image at time $t+1$ based on half-sibling regression. To extend this to multiple channels C, the absolute sum of the difference between the predicted and the actual image is taken as

$$\hat{P} = \begin{cases} 1, & \text{if } \sum_{c=1}^{C} |\hat{\mathbf{I}}_{\mathbf{t+1,c}} - \mathbf{I}_{\mathbf{t+1,c}}| > o \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

For formal details on how $\hat{I}_{t+1,c}$ is obtained given a set of neighbors, we refer to [39]. SiROC ensembles over many mutually exclusive neighborhoods and relies on majority voting between the models for its final prediction. This iterative process uses mutually exclusive sets of neighboring pixels that are increasingly more distant from the pixel of interest itself. Relevant parameters for this process are the maximum neighborhood size and the step size of the ensemble. The number of ensembles is given as the maximum neighborhood size divided by the step size. We use SiROC with its presented defaults in [39]. The respective parameter values are as follows.

1) Maximum neighborhood size: n_max=200.
2) Initial exclusion window: e_start=0.
3) Step size of ensemble: s=2.
4) Filter size of morphological operations: p=5.

One deviation is to reduce the step size of the ensemble from 8 to 2. This results in 100 models with a maximum neighborhood size of 200 and allows for more variation in the uncertainty estimates.

The number of votes, as shown in [39], can be interpreted as a well-calibrated uncertainty and is used in this work as a

confidence score. This is because the performance of SiROC is increasing in its confidence. Therefore, we use SiROC in combination with three supervised student models for CD.

C. Student Model

Once the teacher model is used to select the pseudosamples from U, ideally any machine-learning-based classifier model can be used to train the student model. The training involves the following two steps:

1) training with pseudo labeled N'_U samples from U, obtained in Section II-B;
2) fine tuning with the labeled dataset D.

To illustrate that our SemiSiROC can work with a diverse set of classifiers, we chose several competitive supervised CD architectures. They are outlined in more detail as follows.

FC-Siam-diff [23] is a fully convolutional Siamese neural network inspired by the UNet architecture [30]. Pre- and post-images are processed in two separate parallel streams with shared weights, which are only merged after the convolutional layers of the network. In contrast to a classic concatenation of features, this network takes the absolute difference of the encoding streams. This allows the model to focus on temporal differences in the image pair, which is well suited for CD tasks. These differences are infused as inputs to the upsampling steps. Allowing feature differences to be passed without further processing far into the network allows the network to treat simple decisions without unnecessary complexity.

FC-Siam-conc [23] is similar to *FC-Siam-diff* with one major distinction. Instead of taking feature differences of the encoding streams, the features are concatenated. This gives the model more flexibility but nudges it less directly toward a temporal comparison of features.

DTCDSCN [29] stands for dual task constrained deep Siamese convolutional network. It is a convolutional model, which performs semantic segmentation and CD simultaneously. This is helpful for change detection since a prior understanding of objects and their size from semantic segmentation can be utilized for the CD task.

ChangeFormer [21] is also a Siamese network with a transformer-based encoder that reaches competitive performance on the LEVIR-CD [65] and DSIFN-CD [22] benchmarks. The hierarchical transformer encoder uses four transformer blocks in with shared weights in each branch. After every transformer block, a difference module is taken to compare differences at different abstraction levels. These differences are then passed to a lightweight multilayer perceptron decoder, which samples the features up and computes the final predicted change map.

Bitemporal image transformer (BIT) [25] also relies on self-attention rather than only deep convolutional features in a transformer framework. It has three main elements: a siamese semantic tokenizer, a transformer encoder, and a transformer decoder. The siamese backbone extracts convolutional features and inputs them into the semantic tokenizer. Inspired by advances in language processing, the tokenizer pools the image features into a compact set of vocabulary. The compact tokens are converted back to the pixel space and fed into a CNN prediction head. As a CNN backbone for the feature extraction, ResNet18 is used following the main paper.

III. Experimental Validation

A. Data

1) DynamicEarthNet: We base our analysis on a modified version of the DynamicEarthNet dataset [61]. This is because it allows benchmarking CD algorithms with areas of interest (AOIs) across the globe and covers a variety of different changes that are not specific to a certain use case such as buildings or urban regions only. Both of these properties make the dataset well-tailored to binary CD in an application-agnostic way. It contains monthly, manual land cover annotations for two years with Planet imagery for 75 AOIs across the globe. The locations were selected to include a wide spectrum of land cover changes across seven classes.

We pick the labels of the first and last month of each AOI and compute a binary mask of changing land cover. This maximizes change and also ensures a certain difference in the scenes. The corresponding Planet Fusion images are highly preprocessed as an analysis-ready product, which includes a variety of steps including temporal gap filling of clouds and shadow removal. Each scene is 1024×1024 pixels with 3-m resolution per pixel in size, which results in an area per scene of about $10\ \text{km}^2$. To be consistent with the image size in [21], we split each scene into 16 256×256 pixels RGB images. This results in a total of 1200 pairs of pre and post images taken 2 years apart. The class balance in the resulting dataset is about 80% no change and 20% change.

Our baseline train, validation, and test split is visible in Fig. 1. Locations are available across the globe, which is relevant to test generalizability to unseen regions where all continents except Antarctica are covered. Following the DynamicEarthNet terminology, we refer to the locations also as cubes given that the 2D images also vary in time. The cubes do not only differ by their geography but also by the type of change. The dataset covers locations from coastal areas, islands, urban regions, agricultural areas, and forests. This shows the diversity of change in practical applications, which makes this dataset challenging.

The cubes based in the continental US are used as training (blue), the validation data are taken from central America (green) and we test with the remaining cubes from across the globe. This simulates label scarcity in global CD tasks where generalizability to unseen regions is a key requirement. Particularly, annotated data in low and middle-income countries are often relatively rare. However, to validate our results against this choice, we use other splits with more training data (16, 32, 64 cubes) as an ablation study below.

2) Onera Satellite Change Detection (OSCD) [23]: As a secondary dataset, we rely on OSCD, which in total contains 24 before and after pairs of Sentinel-2 images in urban areas across the globe but we only use the ten pairs in the test set. To be consistent with our training efforts on DynamicEarthNet, we only include the RGB channels and crop 256×256 images from the original scenes. As OSCD image pairs are not square

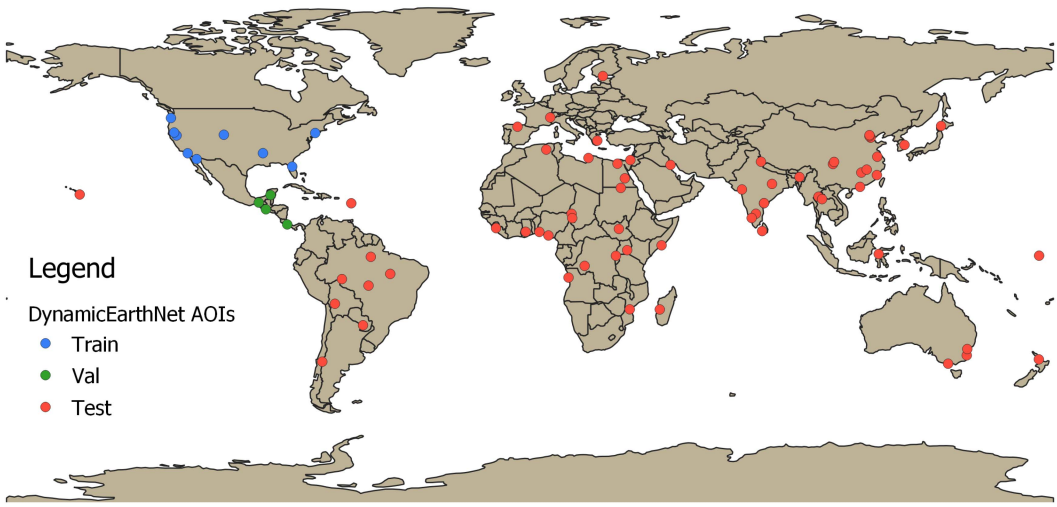

Fig. 1. Spatial train/validation/test split used as a default. The limited amount of training data simulates real-world scenarios where training data are scarce and mainly from specific regions.

and vary in size, we pad the images to the next multiple of 256 and mask the added points during evaluation of the change prediction.

B. Training and Evaluation

Our goal is to evaluate the effectiveness of a PL pretraining step. Therefore, we compare SiROC confidence pretraining for a variety of specifications including the aforementioned models but also different choices of training sets, PL sets, and training losses. We train each model until convergence with and without a pretraining step. For this study, experiments were conducted with a single NVIDIA Quadro P4000. We acknowledge that semisupervised pretraining requires an additional computational effort compared to finetuning. PL training for 50 epochs with the top quarter of scenes by confidence takes about 15 min with the P4000 for the FC-Siam-diff model. However, PL training has to be done only once and allows for all kinds of CD applications.

The following specifications are used for all experiments to ensure comparability. We train with Adam as an optimizer with a batch size of 32 and a starting learning rate of 0.0001 and linear weight decay. We evaluate our results based on three popular criteria: Accuracy, mean IOU (MIOU), and mean F1 Score. Formally, in terms of false positives (FP), true positives (TP), false negatives (FN), and true negatives (TN), these criteria have the following definitions:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FN} + \text{FP}). \quad (3)$$

Accuracy is simply asking how often is our prediction correct relative to the total number of predictions.

$$\text{MIOU} = (\text{IOU}_1 + \text{IOU}_0)/2 \quad (4)$$

with $\text{IOU} = \text{TP}/(\text{TP} + \text{FP} + \text{FN})$. In comparison to accuracy, the IOU criterion eliminates TN from the picture per class. Similarly

$$\text{MF1} = (F1_1 + F1_0)/2 \quad (5)$$

with F1 balancing precision and recall. $F1 = (2 * \text{precision} * \text{recall})/(\text{precision} + \text{recall})$. Precision is defined as $\text{TP}/(\text{TP} + \text{FP})$ and recall as $\text{TP}/(\text{TP} + \text{FN})$. Every model is run for five different seeds and reported scores are, therefore, a mean with the respective standard deviation in brackets.

C. DynamicEarthNet Results

Table I outlines the main results of this article. Overall, we test PL pretraining with SiROC with four different competitive models. Each pair of rows for one model compares the scores with and without pretraining on the confident PL. All specifications are run five times with different seeds to increase the robustness of the result against an unrepresentative seed. PL training is done with a focal loss (FL) and training with the real labels with the split of Fig. 1 and a MIOU loss with only the top 25% of cubes based on average SiROC confidence per cube.

At first, FC-Siam-diff with PL pretraining reaches an overall accuracy of 0.7812 with a MIOU score of 0.4854 and a Mean F1 Score of 0.6029. This makes it the best model in Table I overall according to all three criteria and notably better than its counterpart without pretraining. FC-Siam-diff without SiROC pretraining is about 15 percentage points (p.p.) lower in accuracy, 7 p.p. lower in MIOU, and about 3 p.p. lower in terms of mean F1 score. Further, standard deviations of performance are visibly lower with confidence-filtered PL pretraining for FC-Siam-diff. FC-Siam-Conc does not seem competitive here in comparison with a fairly low accuracy of around 62% with PLs and 56% without them. It seems that without the explicit feature difference, the model is not incentivized to pay enough attention to temporal differences for the final change segmentation. Therefore, it has trouble to distinguish changes from nonchanges. This is improved by the use of PLs but the issue remains large in comparison to FC-Siam-diff.

Similarly, the scores of ChangeFormer improve and stabilize notably by an even larger margin although the baseline performance is comparably bad. The general effectiveness is also

TABLE I
QUANTITATIVE RESULTS DYNAMICEARTHNET GROUPED BY PL USE

Model	PL	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	✓	FL	MIOU	**0.7812** (+-0.0104)	**0.4854** (+-0.0037)	**0.6029** (+-0.0018)
FC-Siam-diff [23]			MIOU	0.6359 (+-0.0405)	0.419 (+-0.0288)	0.5706 (+-0.0244)
FC-Siam-conc [23]	✓	FL	MIOU	0.6174 (+-0.0306)	0.3862 (+-0.0146)	0.5288 (+-0.0109)
FC-Siam-conc [23]			MIOU	0.5592 (+-0.0892)	0.3481 (+-0.0366)	0.4942 (+-0.0156)
ChangeFormer [21]	✓	FL	MIOU	0.736 (+-0.0448)	0.4586 (+-0.0185)	0.584 (+-0.0101)
ChangeFormer [21]			MIOU	0.4848 (+-0.0923)	0.305 (+-0.0627)	0.4545 (+-0.0644)
DTCDSCN [29]	✓	FL	MIOU	0.7208 (+-0.0286)	0.4602 (+-0.0099)	0.5935 (+-0.002)
DTCDSCN [29]			MIOU	0.6844 (+-0.0393)	0.441 (+-0.0236)	0.5815 (+-0.0206)
BIT [25]	✓	FL	MIOU	0.7303 (+-0.0158)	0.4598 (+-0.0086)	0.5887 (+-0.0058)
BIT [25]			MIOU	0.6242 (+-0.0418)	0.4074 (+-0.0227)	0.5587 (+-0.0151)
SiROC [39]				0.6946	0.4408	0.5769

The bold entries correspond to high scores in the column. This is common practice in machine learning.

confirmed when looking at BIT and DTCDSCN although the margins seem slightly lower. Given that DTCDSCN, and particularly, FC-Siam-conc seem weaker convolutional baselines than FC-Siam-diff, we focus on the latter, ChangeFormer and BIT, for the remainder of this paper for the sake of brevity. As an additional baseline, the performance of SiROC on the test set is given as a reference point.

Generally, SiROC places decently on the dataset given that it is an unsupervised method and often even outscores the supervised baselines with few labels. The information contained in the PLs and the capacity of the methods combine effectively in our semisupervised strategy. The respective scores are consistently substantially higher than in the SiROC baseline with the PLs.

Fig. 2 visualizes model predictions for eight image pairs of the models in Table I. On top are the preimage [see Fig. 2(a)] and postimage [see Fig. 2(b)] samples together with the ground truth [see Fig. 2(c)] from left to right. Large forest changes are, for example, visible in the image on the left or in middle. Notably, the illumination conditions between the pre- and postimages differ slightly, which is often a challenge in CD problems [37]. The first comparison is for FC-Siam-Diff with training on PLs in Fig. 2(d) and the corresponding version without it in Fig. 2(e). Fig. 2(d) was the best performing model quantitatively in Table I, which is confirmed by the visual inspection of the predictions.

The location and the shape of large changes are segmented well with limited mistakes. While the model does miss some smaller changes on the right, regions in the middle are segmented well. In comparison to Fig. 2(e) without PLs, the results are visibly better in Fig. 2(d). The plain FC-Siam-Diff is thrown off by different shades of green, which results in false positives in the middle and on the right. The PL version helps to reduce these false positives due to acquisition conditions and further seems to improve not only the location but also the shapes of segmented changes.

As also visible in Table I, the segmentation performance of ChangeFormer and BIT is generally worse in comparison to FC-Siam-Diff. SiROC PLs brought the biggest improvement for ChangeFormer in Table I, which is also visible in Fig. 2(f) and (g). The no PL version predicts change for virtually all grassland regions since it interprets the change in illumination as change. It is, therefore, too sensitive to the change class and struggles to extract meaningful change. This improves visibly with the PL training. For example, the shapes in the middle are fit notably better.

Similarly, the PLs bring improvement with BIT as shapes get more refined and there are fewer false positives on the right.

The impressions of Fig. 2 are generally confirmed when inspecting predictions for a more complex urban scene in Fig. 3. Again, the upper panels for each method show pre- and postimages as well as the ground truth. For all three models, the upper prediction with PL pretraining shows more refined shapes. This becomes particularly visible for ChangeFormer [see Fig. 3(f) and (g)] and BIT [see Fig. 3(h) and (i)], where the predictions without PLs are visibly more blurry and overall worse. The difference is smaller for FC-Siam-diff but the no PL version [see Fig. 3(i)] predicts a number of false positives that are predicted correctly with PLs [see Fig. 3(d)] particularly on the left and center right. On the other hand, both models miss key changes in this complex scene where the no PL variant seems keener on classifying something as a change. Overall, the qualitative inspection of scenes confirms our finding that confidence-filtered PLs help increase CD performance.

Table I shows that PL training is effective in addition to supervised use of labels. Table II outlines what happens when other PLs based on CVA or DCVA are used as semisupervised baselines. The training setup is identical to Table I and the scores for SiROC PL are the same. What varies is the source of the PLs in the pretraining step listed in the second column. FC-Siam-Diff with SiROC PLs reaches high scores in accuracy and MIOU. Accuracy is 2–3 p.p. higher compared to other PLs, which is significant but the MIOU edge is rather small. For MF1, it seems that CVA and DCVA PLs, although lacking behind in accuracy, reach a slightly more balanced classification with 61.21% MF1 each. For ChangeFormer and BIT, the scores are again lower on average. Compared to CVA, the Change Former SiROC combination scores visibly better across all three categories (+ 8 p.p. accuracy,+ 3 p.p. MIOU, + 2 p.p. MF1). ChangeFormer with SiROC PLs notably exceeds accuracy and MIOU compared to its DCVA baseline and obtains a similar MF1 score. The picture for BIT is similar with higher accuracy and MIOU and slightly better (CVA) or marginally worse (DCVA) F1 scores. Overall, SiROC PLs perform visibly better in accuracy and MIOU where the edge is particularly apparent for ChangeFormer and BIT.

Fig. 2. Qualitative results of eight sample image pairs with ground truth and respective model predictions with and without PLs. In general, the PLs seem to help the models reduce false positives based on illumination differences. Examples of this are deforestation in the middle and on the right. (a) Preimages. (b) Postimages. (c) Ground truth. (d) FC-siam-diff PL. (e) FC-siam-diff no PL. (f) ChangeFormer PL. (g) ChangeFormer no PL. (h) BIT PL. (i) BIT no PL.

D. DynamicEarthNet Ablation Studies

1) Amount of Training Data: One may be concerned that the edge of our approach is limited by the small number of training cubes with real labels. Therefore, we iteratively add more training cubes to explore differences in the edge depending on this parameter. Table III presents these scores on a harmonized test set for this table. As we use up to 64 cubes for training and aim to keep the scores comparable, we use the respective test set for all specifications in this table. All PL specifications are again pretrained with the top 25% of cubes in confidence. We use all available training cubes with FC-Siam-diff and Change Former in the upper panel. Despite the increasing amount of training data, FC-Siam-diff remains better than ChangeFormer by a

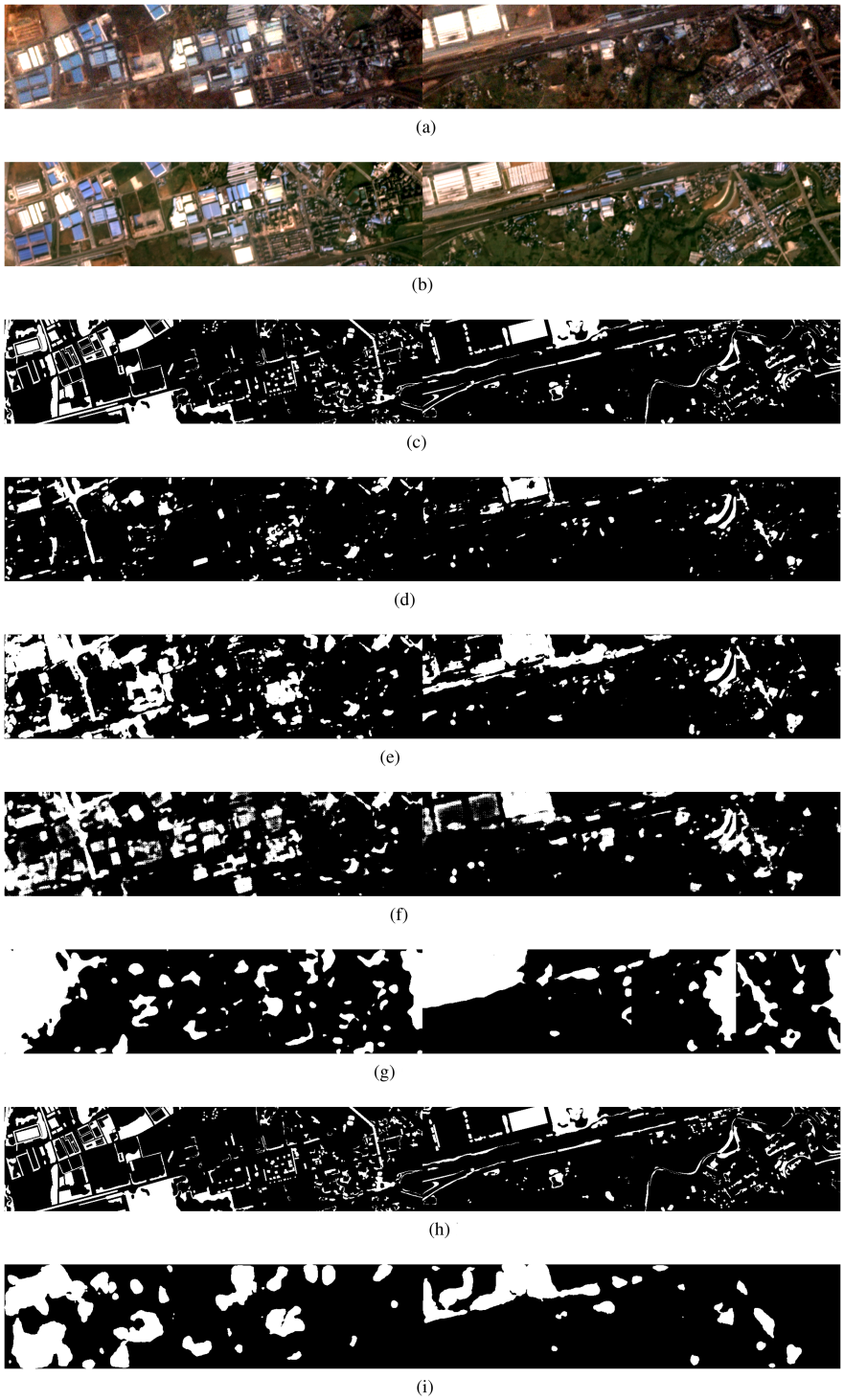

Fig. 3. Qualitative results of eight sample image pairs with ground truth and respective model predictions with and without PLs here for a complex urban scene. (a) Preimages. (b) Postimages. (c) Ground truth. (d) FC-siam-diff PL. (e) FC-siam-diff no PL. (f) ChangeFormer PL. (g) ChangeFormer no PL. (h) BIT PL. (i) BIT no PL.

significant margin. In both specifications, SemiSiroc exceeds the no PL baseline again visibly.

In the lower panel, we compare FC-Siam-diff against versions with fewer training data (25% and 50% of the aforementioned training set). Interestingly, the performance of SemiSiROC increases only marginally with additional real training data. This may indicate that a large part of potential gains through additional training data could already have been exploited by the PLs. Conversely, the gap between PL and no PL gets smaller with 16 training cubes. Then, performance from 16 to 32 cubes drops slightly, which is unexpected. One reason could be that the additional training cubes are somewhat more unrepresentative of the remaining cubes on the other side of the globe compared to the previous cubes. The highest scores with and without PLs

TABLE II
QUANTITATIVE RESULTS DYNAMICEARTHNET WITH DIFFERENT PLS

Model	PL	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	SiROC	FL	MIOU	**0.7812** (+-0.0104)	**0.4854** (+-0.0037)	0.6029 (+-0.0018)
FC-Siam-diff [23]	CVA	FL	MIOU	0.7599 (+-0.0124)	0.4853 (+-0.0072)	**0.6121** (+-0.0048)
FC-Siam-diff [23]	DCVA	FL	MIOU	0.7553 (+-0.0077)	0.4838 (+-0.0015)	**0.6121** (+-0.002)
ChangeFormer [21]	SiROC	FL	MIOU	0.736 (+-0.0448)	0.4586 (+-0.0185)	0.584 (+-0.0101)
ChangeFormer [21]	CVA	FL	MIOU	0.6589 (+-0.0414)	0.4232 (+-0.0254)	0.5666 (+-0.0196)
ChangeFormer [21]	DCVA	FL	MIOU	0.678 (+-0.0264)	0.4423 (+-0.0193)	0.5864 (+-0.0166)
BIT [25]	SiROC	FL	MIOU	0.7303 (+-0.0158)	0.4598 (+-0.0086)	0.5887 (+-0.0058)
BIT [25]	CVA	FL	MIOU	0.6886 (+-0.0091)	0.4437 (+-0.006)	0.5839 (+-0.0049)
BIT [25]	DCVA	FL	MIOU	0.7004 (+-0.0117)	0.4543 (+-0.006)	0.594 (+-0.0038)

The bold entries correspond to high scores in the column. This is common practice in machine learning.

TABLE III
ABLATION STUDY: VARYING THE TRAINING SET SIZE

Model	PL	# Training Cubes	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	✓	64	MIOU	**0.9227** (+-0.0038)	**0.5376** (+-0.0012)	**0.6127** (+-0.0028)
FC-Siam-diff [23]		64	MIOU	0.8538 (+-0.0076)	0.4865 (+-0.0055)	0.5685 (+-0.0048)
ChangeFormer [21]	✓	64	MIOU	0.813 (+-0.0113)	0.4613 (+-0.0098)	0.5494 (+-0.0102)
ChangeFormer [21]		64	MIOU	0.7792 (+-0.0277)	0.4528 (+-0.0239)	0.5516 (+-0.0449)
FC-Siam-diff [23]	✓	32	MIOU	0.9159 (+-0.0115)	0.5324 (+-0.0094)	0.6082 (+-0.0088)
FC-Siam-diff [23]		32	MIOU	0.8215 (+-0.0715)	0.4764 (+-0.031)	0.5681 (+-0.0328)
FC-Siam-diff [23]	✓	16	MIOU	0.9162 (+-0.0127)	0.5338 (+-0.007)	0.6101 (+-0.0047)
FC-Siam-diff [23]		16	MIOU	0.8488 (+-0.0205)	0.4851 (+-0.0107)	0.569 (+-0.0074)

The bold entries correspond to high scores in the column. This is common practice in machine learning.

TABLE IV
ABLATION STUDY: ROBUSTNESS TO FINETUNING LOSS

Model	PL	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	✓	FL	FL	0.787 (+-0.0088)	0.4858 (+-0.0021)	0.6008 (+-0.0051)
FC-Siam-diff [23]			FL	0.693 (+-0.0657)	0.4426 (+-0.0246)	0.5798 (+-0.0163)
FC-Siam-diff [23]	✓	FL	MIOU	0.7812 (+-0.0104)	0.4854 (+-0.0037)	**0.6029** (+-0.0018)
FC-Siam-diff [23]			MIOU	0.6359 (+-0.0405)	0.419 (+-0.0288)	0.5706 (+-0.0244)
FC-Siam-diff [23]	✓	FL	CE	0.7945 (+-0.0088)	**0.4868** (+-0.0043)	0.5987 (+-0.0096)
FC-Siam-diff [23]			CE	0.7988 (+-0.0233)	0.4466 (+-0.0219)	0.5304 (+-0.0483)
ChangeFormer [21]	✓	FL	FL	0.6762 (+-0.0538)	0.4355 (+-0.034)	0.5769 (+-0.027)
ChangeFormer [21]			FL	0.5644 (+-0.0164)	0.3548 (+-0.0085)	0.5036 (+-0.0088)
ChangeFormer [21]	✓	FL	MIOU	0.736 (+-0.0448)	0.4586 (+-0.0185)	0.584 (+-0.0101)
ChangeFormer [21]			MIOU	0.4848 (+-0.0923)	0.305 (+-0.0627)	0.4545 (+-0.0644)
ChangeFormer [21]	✓	FL	CE	**0.8068** (+-0.0122)	0.4399 (+-0.0158)	0.5155 (+-0.0321)
ChangeFormer [21]			CE	0.7735 (+-0.0471)	0.4237 (+-0.0088)	0.5067 (+-0.0178)
BIT [25]	✓	FL	FL	0.7133 (+-0.0203)	0.4531 (+-0.0088)	0.5864 (+-0.0066)
BIT [25]			FL	0.6673 (+-0.0774)	0.412 (+-0.0318)	0.5447 (+-0.0222)
BIT [25]	✓	FL	MIOU	0.7303 (+-0.0158)	0.4598 (+-0.0086)	0.5887 (+-0.0058)
BIT [25]			MIOU	0.6242 (+-0.0418)	0.4074 (+-0.0227)	0.5587 (+-0.0151)
BIT [25]	✓	FL	CE	0.7593 (+-0.0145)	0.4639 (+-0.0027)	0.581 (+-0.0098)
BIT [25]			CE	0.7876 (+-0.0256)	0.4236 (+-0.0055)	0.4984 (+-0.0139)
SiROC				0.6946	0.4408	0.5769

The bold entries correspond to high scores in the column. This is common practice in machine learning.

are achieved with the maximum number of training cubes of 64, which is about 85% of our dataset with over 1000 image pairs, where the rest is used for testing and validation. Still, the PL specification remains better than its baseline with a sizeable gap. Overall, the main takeaway remains unaffected. With both a few and a larger amount of labels, SemiSiroc is an effective strategy for CD on this dataset.

2) Varying the Finetuning Loss: However, the edge of our strategy may be specific to the loss combination used. Therefore, we test the robustness of our results with other losses at the finetuning step in Table IV for ChangeFormer, BIT, and FC-Siam-diff. We do not vary the PL loss here as this would leave the baselines without SiROC pretraining unaffected. In total, there are six specifications per model given three loss combinations each. The MIOU scores are identical to Table I.

The choice of the finetuning loss leaves SemiSiROC largely unaffected with minor differences in scores. It is marginally better in accuracy and MIOU compared to the MIOU loss and slightly lower in terms of Mean F1. The focal loss baseline with FC-Siam-diff is slightly stronger than with MIOU but still lacks

TABLE V
ABLATION STUDY: PL TRAINING NOT ON TEST IMAGES WITH SIAMUNET

Model	PL	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	✓	FL	MIOU	**0.7541** (+-0.0115)	**0.4621** (+-0.004)	**0.581** (+-0.0017)
FC-Siam-diff [23]			MIOU	0.5965 (+-0.0419)	0.3902 (+-0.0286)	0.5448 (+-0.0246)

The bold entries correspond to high scores in the column. This is common practice in machine learning.

TABLE VI
ABLATION STUDY: DIFFERENT CONFIDENCE SPLITS

Model	PL Split	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	Top 25%	FL	MIOU	0.7812 (+-0.0104)	**0.4854** (+-0.0037)	0.6029 (+-0.0018)
FC-Siam-diff [23]	Random 25%	FL	MIOU	0.7541 (+-0.0113)	0.4801 (+-0.002)	**0.6074 (+-0.003)**
FC-Siam-diff [23]	Bottom 25%	FL	MIOU	0.7576 (+-0.0093)	0.4767 (+-0.005)	0.6009 (+-0.0075)
FC-Siam-diff [23]	Top 50%	FL	MIOU	0.7794 (+-0.0067)	0.4839 (+-0.004)	0.6016 (+-0.0031)
FC-Siam-diff [23]	All	FL	MIOU	**0.787** (+-0.0055)	0.4824 (+-0.0043)	0.5958 (+-0.0054)

The bold entries correspond to high scores in the column. This is common practice in machine learning.

behind the comparable SemiSiROC specification by about 9 p.p. in accuracy, 4 p.p. in MIOU, and 2 p.p. in Mean F1.

Expectedly, training with a cross-entropy (CE) loss pushes the FC-Siam-diff baseline to almost exclusively predict the majority no change class. This results in an accuracy high score of almost 0.80, which even marginally surpasses the respective SemiSiROC score although with a higher standard deviation. However, the corresponding Mean F1 score, which is comparably sensitive to large discrepancies in predictive performance across the classes falls behind by almost 7 p.p. to the SemiSiROC CE score.

For the ChangeFormer model, the observations of the MIOU finetuning seem to be confirmed. Similar to FC-Siam-diff, CE training leads to the prediction of mostly no change. The FL results are somewhat better than the MIOU results but still comparably bad. Overall, Table IV confirms the impression of the effectiveness of our semisupervised strategy.

At last, the results for the BIT model mirror the aforementioned results. Pseudolabeling is highly effective across all categories with an FL or MIOU loss. With CE, the model again tends to overfit largely to the no-change class, which is why the accuracies are higher. Even though the no PL version with CE loss reaches the highest accuracy among BIT models, the results are visibly unbalanced. While the PL version lacks behind 3 p.p. in accuracy, it makes more balanced choices with more than 8 p.p. more MF1.

3) Results on Unseen Geographic Areas: Note that for the two previous tables, we did not restrict the PLs to be outside of the test set. While during training, no model sees any actual labels from the test set, one could argue that the images of the test set may be advantageous for our strategy.

To ensure that our strategy is effective also on cubes that were also not part of the PL training, we split the former test set in two where we use the western half from the perspective of Fig. 1 for PL training and the eastern half for testing with the FC-Siam-diff as the most effective model overall. The respective scores are reported in Table V and cannot be directly compared to the scores of previous tables anymore because of the difference in the test cubes. Still, the PL step remains better in comparison by a wide margin that seems even bigger than in previous comparisons. The gap is substantial at 15 p.p. in accuracy and 7 p.p. in MIOU.

4) PL Filtering: Another ablation study concerns the effectiveness of the PL filtering. Since labels are limited, the preselection discards additional information, which may be useful in training. Therefore, we mix up the cube selection with a random selection and the lowest 25% in confidence. The respective results are reported in Table VI . The top 25% cubes score best in terms of accuracy and MIOU and fall just short of the random selection in terms of MF1. Still, with a difference of almost 3 p.p. with similar MIOU and F1 values, it seems that the confidence prefiltering indeed extracts meaningful PLs, which result in more effective learning. Additionally, we notice decreasing marginal returns of adding a higher fraction of PLs in our case. Using the top half or even all cubes with their respective PLs results in a similar performance than only using the top quarter. Therefore, we choose the threshold of 25% for more efficient training. Even though SiROC PLs improve performance already without filtering, the confidence selection further pushes the CD performance.

E. OSCD Results

To further investigate the transferability and generizability of the proposed approach, we evaluate SemiSiROC also on OSCD [23], which is a widely used binary CD benchmark based on Sentinel-2 with a focus on urban regions. The results of our experiments are presented in Table VII. The models used are identical to the ones in Table I. We merely apply them to the OSCD test set instead of the DynamicEarthNet test set directly to analyze the transferability of models. Similar to Table I, we test a variation with additional PL pretraining and without it for each model. At first, FC-Siam-diff [23] remains a strong model and achieves an average accuracy of above 95% with a MIOU of 55.47% and an MF1 score of 62.06% across the five runs. There is a notable difference across all three scoring criteria between the PL and the no PL version. Most significantly, accuracy drops about 15 p.p. without DynamicEarthNet-based PL pretraining. This is the case even though both models were trained with

TABLE VII
QUANTITATIVE RESULTS OSCD TEST SET TRAINED ON DYNAMICEARTHNET AND GROUPED BY PL USE

Model	PL	Loss PL	Loss	Accuracy	MIOU	MF1
FC-Siam-diff [23]	✓	FL	MIOU	**0.9575** (+-0.0096)	0.5547 (+-0.0185)	0.6206 (+-0.0252)
FC-Siam-diff [23]			MIOU	0.8083 (+-0.1035)	0.4927 (+-0.0892)	0.5966 (+-0.082)
ChangeFormer [21]	✓	FL	MIOU	0.8592 (+-0.0692)	0.5145 (+-0.0457)	0.6085 (+-0.0356)
ChangeFormer [21]			MIOU	0.384 (+-0.2976)	0.2139 (+-0.1703)	0.2984 (+-0.1892)
BIT [25]	✓	FL	MIOU	0.9248 (+-0.0154)	**0.5585** (+-0.0115)	**0.6422** (+-0.012)
BIT [25]			MIOU	0.7273 (+-0.059)	0.4082 (+-0.0321)	0.5066 (+-0.0238)

These are the models of Table I applied to the OSCD test set without retraining.
The bold entries correspond to high scores in the column. This is common practice in machine learning.

real DynamicEarthNet labels. Interestingly, the accuracies are in the range (94–96%) of FC-Siam models in [23] based on supervised training on OSCD, whereas our approach does not use OSCD labels at all. The contrast to no PLs gets even larger for ChangeFormer although some of the ChangerFormer models seem to tilt toward predicting mostly change on this dataset, which results in unstable average performance. Even when excluding these runs, however, the maximum performance of ChangeFormer on the OSCD test set is 74.13% accuracy, 41.86% MIOU, and 51.71% which is substantially below the average with PLs. Third, BIT model PLs is arguably the best model here since it is only slightly inferior to FC-Siam-diff in accuracy but achieves high scores in MIOU and MF1 with 55.85% and 64.22%, respectively. Again, the difference to no PLs is large across all categories. Overall, the OSCD results confirm the previous impression that PL pretraining with SemiSiROC can be highly effective in optical CD applications.

IV. DISCUSSION

A. Comparing Teacher and Students

The previous section outlines the effectiveness of SiROC as an unsupervised teacher model for CD with limited labels. This is because it is an effective method and can prioritize PLs based on a well-calibrated confidence. The mechanism for these improvements seems to be higher robustness to false positives because of acquisition conditions and more refined shapes of changes.

Since SiROC models analyze how much a pixel changes in comparison to its neighborhood, it seems intuitive that it would guide a student model toward higher robustness to false positives. Consider the example of Fig. 2. Grassland seems much greener in the post images but since this affects virtually all pixels in the grassland neighborhood of a pixel, SiROC would not necessarily view this as change. This is something the student models seem to pick up on without modeling this explicitly. Another property of SemiSiROC seems to be more refined change shapes, which is also a strength of the initial SiROC model [39]. This may incentivize the student model to learn more about likely shapes and spatial dependencies of changes.

B. Relative Weakness of Transformer Models

Second, we notice that throughout our results, the two transformer models seem to perform worse compared to the siamese UNet. This results in large gains through PL pretraining and underlines the effectiveness of our strategy. There are several possible explanations for this relative weakness. A likely candidate is model size and label availability. ChangeFormer, in particular, is a large model, which makes it data hungry and its success on other datasets such as Levir-CD in [21] may be related to the fact that more labels are available there. This seems plausible for Levir-CD, which was about 10x more labeled pixels than the binary DynamicEarthNet we use here.

However, DSIFN only has 25% more labeled pixels than our dataset. Therefore, another reason could be that both of these methods have been tested in the context of urban CD only with a focus on buildings. Maybe the different kinds of change applications across the globe within DynamicEarthNet pose a challenge to these models and the smaller siamese model adjusts to this more quickly. Nevertheless, the SemiSiROC framework shows effectiveness for all the methods we tested here and shows promise for CD applications with optical data in practice. Our model pretrained with PLs converges faster during fine tuning (i.e., training with actual labels). Thus, our proposed method reduces the time requirement of the training phase with actual samples.

V. CONCLUSION

Monitoring changes of the Earth's surface over time with satellite imagery is an integral part of remote sensing. In this article, we combine unsupervised and supervised techniques in a semisupervised framework. This framework, called SemiSiROC, relies on pretraining a student model with PLs that we filter by confidence. This enables the student model to learn from additional, meaningful high-confidence examples in a pretraining step before finetuning with actual labels. We evaluate SemiSiROC with three different supervised backbones: FC-Siam-Diff, ChangeFormer, and BIT. We evaluate the models with and without filtered PL pretraining on a binary version of the DynamicEarthNet benchmark that is based on Planet Fusion imagery with 3-m resolution. We pick only the cubes with the 25% highest confidence scores during pretraining. For all three models, we find a notable boost in performance for our baseline specification in Table I with eight cubes, which corresponds to 124 training scene pairs with real labels. Additionally, we outline that SemiSiROC remains competitive in the eye of semisupervised student–teacher baselines based on DCVA and CVA PLs.

Further, we evaluate the SemiSiROC models on scenes not seen during PL training, which results in similar performance gains. This ensures that the learned features are not specific to scenes close to the PLs. Even with 64 training cubes with

over 1000 labeled pairs, SemiSiROC is effective compared to its non-PL baseline, where gains are still large. Additional evaluations on the OSCD benchmark confirm the effectiveness of our SemiSiROC strategy also on an urban CD dataset based on Sentinel-2. Qualitative inspections of the predictions shed light on what the teacher model seems to teach its students: Compared to its no PL counterparts, the SemiSiROC models predict more refined shapes and seem to be less sensitive to false positives.

Our results point toward several potentially promising future research directions. At first, our work could be applied to related tasks such as multiclass CD or different input sensors. Second, more experiments are necessary to understand the role of teacher models in spatial generalization generally and particularly in CD.

References

[1] J. A. Cardille, E. Perez, M. A. Crowley, M. A. Wulder, J. C. White, and T. Hermosilla, "Multi-sensor change detection for within-year capture and labelling of forest disturbance," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112741.

[2] G. Chen and G. J. Hay, "An airborne lidar sampling strategy to model forest canopy height from quickbird imagery and geobia," *Remote Sens. Environ.*, vol. 115, no. 6, pp. 1532–1542, 2011.

[3] C. Senf and R. Seidl, "Mapping the forest disturbance regimes of Europe," *Nature Sustainability*, vol. 4, no. 1, pp. 63–70, 2021.

[4] D. Lu, E. Moran, and S. Hetrick, "Detection of impervious surface change with multitemporal landsat images in an urban–rural frontier," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 298–306, 2011.

[5] S. Ji, Y. Shen, M. Lu, and Y. Zhang, "Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1343.

[6] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for sea ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.

[7] K. Rokni, A. Ahmad, K. Solaimani, and S. Hazini, "A new approach for surface water change detection: Integration of pixel level image fusion and image classification techniques," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 226–234, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0303243414001780

[8] R. Gupta et al., "Creating xBD: A dataset for assessing building damage from satellite imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 10–17.

[9] Z. Lv, H. Huang, L. Gao, J. A. Benediktsson, M. Zhao, and C. Shi, "Simple multiscale UNet for change detection with heterogeneous remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, May 22, 2022, Art. no. 2504905.

[10] Z. Lv, F. Wang, G. Cui, J. A. Benediktsson, T. Lei, and W. Sun, "Spatial–spectral attention network guided with change magnitude image for land cover change detection using remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 22, 2022, Art. no. 4412712.

[11] L. Moya et al., "Detecting urban changes using phase correlation and l1-based sparse model for early disaster response: A case study of the 2018 Sulawesi Indonesia earthquake-tsunami," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111743.

[12] M. Zanetti et al., "A system for burned area detection on multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 22, 2022, Art. no. 5404315.

[13] Z. Lv, T. Liu, J. A. Benediktsson, and N. Falco, "Land cover change detection techniques: Very-high-resolution optical images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 44–63, Mar. 2022.

[14] J. Aschbacher, "ESA's earth observation strategy and copernicus," in *Satellite Earth Observations and Their Impact on Society and Policy*. Singapore: Springer, 2017, pp. 81–86.

[15] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[16] C. Kwan et al., "Assessment of spatiotemporal fusion algorithms for planet and worldview images," *Sensors*, vol. 18, no. 4, 2018, Art. no. 1051.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[19] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.

[20] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[21] W. G. C. Bandara and V. M. Patel, "A transformer-based siamese network for change detection," *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 207–210.

[22] C. Zhang et al., "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 183–200, 2020.

[23] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[24] R. C. Daudt, B. L. Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2115–2118.

[25] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5607514.

[26] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

[27] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 506.

[28] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[29] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, "Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 811–815, May 2021.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv.*, 2015, pp. 234–241.

[31] H. Zhang, M. Lin, G. Yang, and L. Zhang, "ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 28–42, Jan. 2023.

[32] Q. Xu, K. Chen, X. Sun, Y. Zhang, H. Li, and G. Xu, "Pseudo-siamese capsule network for aerial remote sensing images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6000405.

[33] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[34] A. Vaswani et al., "Attention is all you need," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[35] J. Chen et al., "DASNet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1194–1206, 2021.

[36] F. I. Diakogiannis, F. Waldner, and P. Caccetta, "Looking for change? Roll the dice and demand attention," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3707.

[37] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[38] S. Saha, Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "Unsupervised deep transfer learning-based change detection for HR multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 856–860, May 2021.

[39] L. Kondmann, A. Toker, S. Saha, B. Schölkopf, L. Leal-Taixé, and X. X. Zhu, "Spatial context awareness for unsupervised change detection in optical satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 22, 2022, Art. no. 5614615.

[40] C. Ren, X. Wang, J. Gao, and H. Chen, "Unsupervised change detection in satellite images with generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59 no. 12, pp. 10047–10061, 2020.

[41] N. Falco, G. Cavallaro, P. R. Marpu, and J. A. Benediktsson, "Unsupervised change detection analysis to multi-channel scenario based on morphological contextual analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 3374–3377.
[42] S. Saha, L. Mou, C. Qiu, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Unsupervised deep joint segmentation of multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8780–8792, Dec. 2020.
[43] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative adversarial networks for change detection in multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2310–2314, Dec. 2017.
[44] T. Zhan, M. Gong, X. Jiang, and M. Zhang, "Unsupervised scale-driven change detection with deep spatial-spectral features for VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5653–5665, Aug. 2020.
[45] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with landsat," in *Proc. LARS Symposia*, 1980, pp. 33–37.
[46] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.
[47] F. Thonfeld, H. Feilhauer, M. Braun, and G. Menz, "Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 50, pp. 131–140, 2016.
[48] L. Li, X. Li, Y. Zhang, L. Wang, and G. Ying, "Change detection for high-resolution remote sensing imagery using object-oriented change vector analysis method," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 2873–2876.
[49] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
[50] S. Ghosh, M. Roy, and A. Ghosh, "Semi-supervised change detection using modified self-organizing feature map neural network," *Appl. Soft Comput.*, vol. 15, pp. 1–20, 2014.
[51] K. Chen, Z. Zhou, C. Huo, X. Sun, and K. Fu, "A semisupervised context-sensitive change detection technique via gaussian process," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 236–240, Mar. 2013.
[52] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, 2015.
[53] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 607–611, Apr. 2021.
[54] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10687–10698.
[55] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.
[56] M. Gong, H. Yang, and P. Zhang, "Feature learning and change feature classification based on deep learning for ternary change detection in SAR images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 129, pp. 212–225, 2017.
[57] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 321–333, Jan. 2019.
[58] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, Aug. 2019.
[59] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.
[60] S. Saha, B. Banerjee, and X. X. Zhu, "Trusting small training dataset for supervised change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2031–2034.
[61] A. Toker et al., "DynamicEarthNet: Daily multi-spectral satellite dataset for semantic change segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 21158–21167.
[62] I. Z. Yalniz, H. Jégou, K. Chen, M. Paluri, and D. Mahajan, "Billion-scale semi-supervised learning for image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13975–13985.
[63] L. Kondmann et al., "DENETHOR: The dynamicearthnet dataset for harmonized, inter-operable, analysis-ready, daily crop monitoring from space," in *Proc. 35th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2021, pp. 1–13.
[64] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[65] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662.

Lukas Kondmann received the bachelor degree in economics from the Ludwig-Maximilians University of Munich, Munich, Germany, in 2016, the honors degree in technology management from the Center for Digital Technology Management, Munich, in 2017, and the master degree in social data science from the University of Oxford, Oxford, U.K., in 2019. He is currently working toward the Ph.D. degree in engineering with the Technical University of Munich and the German Aerospace Center, Munich.

He was a Visiting Researcher working on Big Data for social good with the UC Berkeley School of Information in spring 2017. His research interests include time-series analysis of multispectral remote sensing imagery with a focus on monitoring the sustainable development goals.

Sudipan Saha received the M.Tech. degree in electrical engineering from the Indian Institute of Technology Bombay Mumbai, India, in 2014, and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento, Italy, in 2020.

He was a Postdoctoral Researcher with the Technical University of Munich (TUM), Munich, Germany, and has worked as an Engineer with TSMC Limited, Hsinchu, Taiwan, from 2015 to 2016. In 2019, he was a Guest Researcher with the Technical University of Munich (TUM), Munich, Germany. He is currently an Assistant Professor with the Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi, India. His research interests include multitemporal remote sensing image analysis, domain adaptation, time-series analysis, image segmentation, deep learning, image processing, and pattern recognition.

Dr. Saha was the recipient of Fondazione Bruno Kessler Best Student Award 2020. He is a Reviewer for several international journals and served as a Guest Editor at Remote Sensing (MDPI) special issue on "Advanced Artificial Intelligence for Remote Sensing: Methodology and Application."

Xiao Xiang Zhu (Fellow, IEEE) received the master (M.Sc.), doctor of engineering (Dr.-Ing.), and "Habilitation" degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She was a Guest Scientist or Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan, and the University of California, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is the Chair Professor with Data Science in Earth Observation, TUM and was the Founding Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR). Since 2019, she has been a Co-coordinator with the Munich Data Science Research School (www.mu-ds.de). Since 2019, she has also been the Head of the Helmholtz Artificial Intelligence—Research Field "Aeronautics, Space and Transport." Since May 2020, she is the PI and Director of the international future AI lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has also been serving as a Co-Director with the Munich Data Science Institute (MDSI), TUM. She is currently a Visiting AI Professor with Phi-lab, European Space Agency. Her main research interests include remote sensing and Earth observation, signal processing, machine learning, and data science, with their applications in tackling societal grand challenges, e.g., Global Urbanization, UN's SDGs, and Climate Change.

Dr. Zhu is a Member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the Scientific Advisory Board in several research organizations, among others the German Research Center for Geosciences (GFZ) and Potsdam Institute for Climate Impact Research (PIK). She is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and serves as the Area Editor responsible for special issues of IEEE SIGNAL PROCESSING MAGAZINE.