Neurocomputing 538 (2023) 126201

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Texture based prototypical network for few-shot semantic segmentation of forest cover: Generalizing for different geographical regions

Gokul Puthumanaillam [a,b], Ujjwal Verma [b,c,*]

[a] Department of Aerospace Engineering and the Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, USA
[b] Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India
[c] Department of Electronics and Communication Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India

ARTICLE INFO

Article history:
Received 2 October 2022
Revised 16 March 2023
Accepted 28 March 2023
Available online 31 March 2023
Communicated by Zidong Wang

Keywords:
Semantic segmentation
Forest identification
Cross-geography generalization
Sustainable development goals
Few shot learning

ABSTRACT

Forest plays a vital role in reducing greenhouse gas emissions and mitigating climate change, besides maintaining the world's biodiversity. The existing satellite-based forest monitoring system utilizes supervised learning approaches limited to a particular region and depends on manually annotated data to identify forest. This work envisages forest identification as a few-shot semantic segmentation task to achieve generalization across different geographical regions. The proposed few-shot segmentation approach incorporates a texture attention module in the prototypical network to highlight the texture features of the forest. Indeed, the forest exhibits a characteristic texture different from other classes, such as road, water, etc. In this work, the proposed approach is trained for identifying tropical forests of South Asia and adapted to determine the temperate forest of Central Europe with the help of a few (one image for 1-shot) manually annotated support images of the temperate forest. An IoU of 0.62 for forest class (1-way 1-shot) was obtained using the proposed method, which is significantly higher (0.46 for PANet) than the existing few-shot semantic segmentation approach. Besides, the experimental results demonstrate that the inclusion of the texture attention module in the existing prototypical few-shot segmentation methods (PFENet and ASGNet) results in a more accurate forest identification. These results indicate that the proposed approach can generalize across geographical regions for forest identification, creating an opportunity to develop a global forest cover identification tool.

© 2023 Elsevier B.V. All rights reserved.

1. Introduction

Forest plays a vital role in regulating the ecosystem, protecting biodiversity, and supporting sustainable growth. It does also act as a significant carbon sink absorbing one-third of the CO_2 released from burning fossil fuels. Due to the increasing demand for agricultural farmland, timber, mineral, and other resources, there has been a continuous decrease in forest cover across the globe. The world has lost around 178 million *hectares* of forest since 1990, with Africa recording the largest annual rate of net forest loss [10]. Deforestation severely affects the carbon cycle, making it a significant focus area for climate change studies. Besides, deforestation also results in biodiversity loss, adversely affecting the surrounding ecosystem.

The satellite-based forest monitoring system provides a cost-effective approach for studying deforestation. This system compares the forest cover of a particular region with the earlier forest cover of the same region to locate the area of deforestation. Identifying forests in satellite images acquired over a period of time is beneficial for detecting deforestation in remote and inaccessible areas. For instance, the Brazilian government's PRODES and DETER program uses Landsat and MODIS data to calculate the deforestation region in Brazil's legal amazon. Despite the success of these approaches in identifying the areas of deforestation, these methods are limited to Brazil's legal amazon rainforest and have limited success in identifying small-scale deforestation. Besides, these approaches involve manual analysis, which can be a time-consuming task [17]. In addition, change detection techniques have also been utilized to identify regions of deforestation [25]. However, these approaches' study area is limited to a particular region such as the Brazilian Amazon [3].

Forest identification in satellite images is a crucial step in a satellite-based forest monitoring system. Majority of existing methods for automatic identification of forest utilises traditional

* Corresponding author at: Department of Electronics and Communication Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India.
E-mail address: ujjwal.verma@manipal.edu (U. Verma).

https://doi.org/10.1016/j.neucom.2023.03.062
0925-2312/© 2023 Elsevier B.V. All rights reserved.

supervised machine learning classifiers or supervised deep learning approaches [6,11] [2]. In this work, forest identification is formulated as a few-shot semantic segmentation task. In a supervised learning approach, the train and test distribution are assumed to be identical, which would limit their application to the geographical region represented in the training set. In contrast, the few-shot learning can be utilized to identify the forest of an unseen region, which is *not* present in the training set, with the help of a few labelled images of the unseen region. Unlike the supervised learning approach, the few-shot method can be utilized to identify forests, especially for regions where only a few labelled images (support set) are available [20]. Therefore, this approach can be used to create a tool for identifying forests across all regions with a limited amount of manually annotated images (one image for 1-shot). Despite the recent advances in few-shot learning, few authors have utilized this approach for forest identification. The authors in [20] studied the performance of few-shot learning (model agnostic meta-learning (MAML)) algorithm for land cover segmentation on high-resolution satellite images of South Asia. In [20], the original images were divided into non-overlapping sub-images, and a few sub-images were used as support set to perform the segmentation on other sub-images. This approach permitted the authors to study the model's performance in identifying the entire region's forest using a small number of sub-images of the *same* geographical region.

In comparison, our work focuses on evaluating the few-shot semantic segmentation method's performance in identifying the forest of a particular geographical region using the model trained on a completely *different* geographical region. This work proposes a prototypical few- shot semantic segmentation network that leverages a texture attention module. The tree canopy of the forest exhibits a characteristic texture different from other background classes, such as roads, barren lands etc. The model is trained on DeepGlobe land cover segmentation dataset [7], which contains satellite images from India, Indonesia, and Thailand. This trained model is then evaluated on the LandCover AI dataset [5], which includes images from Central Europe. This formulation of few-shot semantic segmentation allows us to develop a forest identification method, which would not be restricted by the geographical region. In this work, forest refers to the region covered with trees standing in proximity with single trees and orchards being ignored as defined in [5].

This work's main contributions are: 1) Re-formulation of few-shot semantic segmentation task for identifying forest across different geographical region: Training the model on images of tropical forests (South Asia), while testing its performance on temperate forests (Central Europe), 2) Development of texture based prototypical few-shot segmentation architecture for forest identification. The proposed texture attention module can be integrated with any existing prototypical network to emphasise the characteristic textures of a particular class. The rest of the paper is organized as follows: Section 2 summarizes the recent works in identification of forest cover and existing few shot semantic segmentation methods, Section 3 presents the proposed few-shot semantic segmentation approach utilized for forest identification, and Section 4 shows the result obtained.

2. Related works

Remote sensing methodologies have been widely used to study the earth and its surface [12]. Remote sensing technologies enable us to study a larger spatial area and can also be used to study the temporal evolution of the earth's surface. Remote sensing tools have been used to identify forests for various applications such as estimating deforestation. However, most existing remote sensing methods focus on a particular region (eg: Switzerland [29], Canada [30]). There are limited works that focus on estimating forest cover at a global scale [21] [24]. Most of these works [21] [24] require domain expertise in estimating the forest cover. Given the rapid change in the forest cover over the years, manually updating the forest cover map every year is time-consuming. In recent years, machine learning methodologies have been utilized to automatically identify the forest cover of a particular region. Traditional machine learning methods such as random forest, Support vector Machine, and k-Nearest Neighbour, have been used to recognize areas with forest cover [6,11]. These traditional machine learning models depend on hand-crafted features and require careful estimation of hyper-parameters. In the last decade, the deep learning methods have shown to outperform these traditional machine learning methods for various computer vision tasks such as image classification, semantic segmentation. Recently, a supervised deep learning model (UNet) was utilized to identify forest regions of Philippines [2]. However, the supervised deep learning methods requires huge amount of manually annotated images for training the model which is a time-consuming and expensive task. These supervised methods would fail to generalize in identifying forest cover of a different region. To the best of our knowledge, there are no existing methods which study the generalization of deep learning methods for estimating forest cover *across* geographical areas. The proposed work develops a deep learning model that can generalize to a different geographical region with the help of a few (one or two) labeled images.

Few Shot Learning is one of the approach utilized to learn a robust representation with few limited labelled images. In few shot learning, the representation learnt from a large dataset is adapted for a set of unseen classes with a help of few limited images. Specifically, few shot image segmentation aims to assign pixel level labels to the unseen semantic class (like bus, aeroplane) with the help of few labelled images (support set). Recently, a unified approach for segmenting object was proposed in [15] for three tasks: Video Segmentation, Image Co-Segmentation and Few Shot Semantic Segmentation. This graph neural network based approach exploits similar visual appearance to provide pixel wise segmentation. In addition, a non-parametric segmentation approach based on non-learnable prototypes are presented in [32].

The existing few shot semantic segmentation methods can be categorized into prototypical networks or parametric methods. In prototypical networks, the prototype features extracted from the support set are matched (pixel-wise) with the features of query images for each sematic class. For instance, PANet [28] assigns the semantic class represented by the learned prototype features nearest to the query features. PANet method also introduces prototype alignment regularization between support and query images. In a separate study [22] inspired with weight imprinting, multi-resolution average pooling is integrated with base embedding. The base embeddings are obtained from the support set images masked with the support set labels using VGG-16 as the base network. In another study focusing on segmentation of foreground objects [16] proposes a feature relevance term to ensure discriminative features are learnt for a given class. Besides, it also proposes a boosted inference approach using an ensemble of experts. In a separate study, superpixel-guided clustering is utilized to generate more representative features [14]. To improve the quality of prototype features, a feature enrichment module is proposed in [26]. This module extracts multi-scale information along with finer relationship between support and query features. Recently [8], regularization on support set is proposed to ensure robustness in learned prototype representation. Most of the existing prototypical networks learn the prototypical representations from support set using a backbone architecture like VGG-16. These representations might contain non-discriminative features for different classes

thereby adversely affecting the segmentation results. In contrast, the proposed approach uses a texture attention module to highlight the relevant features. Moreover, the proposed texture attention module can be integrated with an existing prototypical network to emphasize the zones with characteristics texture features.

The parametric methods based few shot semantic segmentation jointly learns the representation from support set and estimates the segmentation on query images. For instance, Fully Guided Network (FGN) is built upon Mask R-CNN wherein the support set is utilized to guide the Mask R-CNN to obtain instance level segmentation [9]. In this work, prototypical based approach is utilized due to its lightweight and compact representation.

3. Method

3.1. Problem formulation

This work focuses on identifying areas with forest cover across geographical regions with limited annotated images. The cross-geography generalization is achieved by formulating the forest identification as a few-shot segmentation task. In a few shot semantic segmentation task, images from two non overlapping sets of classes, C_{train} and C_{test} are provided such that training dataset, D_{train} consists of images of C_{train} classes and test dataset D_{test} is made from C_{test} [28]. In this work, the D_{train} contains images from South East Asia [7] while D_{test} contains images from Central Europe [5]. This evaluation paradigm focuses on cross-geography generalization by utilizing tropical forest images of South Asia for training the model while testing the model on temperate forest images of central Europe.

As in the few shot semantic segmentation framework [28], the train/test set is divided into several episodes, with each episode containing a set of support image with annotation and the query image. A C-way K-shot semantic segmentation implies that support set S contains K image pairs (Image, Mask) with pixels from C different classes per episode. The goal is to segment the area of the unseen class C_{test} from each query image given K samples from the support set S.

3.1.1. Dataset and pre-processing

The dataset from the DeepGlobe challenge [7] contains high-resolution optical satellite images from South Asian countries, including India, Thailand, and Indonesia. The dataset contains 803 images of 2,448 $\times$ 2,448 resolution. A mask is provided for each image which contains seven different classes: Urban, Agriculture, Rangeland, Forest, Water, Barren and also an Unknown class. In this work, the images are tiled to create sub-images of 612 $\times$ 612 pixels. This tiling results in 12,800 sub-images that form the training set, D_{train}. Moreover, this work focuses on forest identification; therefore the following classes are merged for 1-way task: Urban, Agriculture, Water, Rangeland, Barren and Unknown.

On the other hand, the test dataset was adapted from the Land-Cover AI dataset [5] which consists of 41 RGB images acquired over central Europe (Poland). This dataset contains four classes: Building, Woodlands (trees standing in close proximity), Water, and Background (unknown). The input images in train set were resized to (128 $\times$ 128) pixels and augmented by randomly flipping the image/mask horizontally. According to the standard practice for few-shot semantic segmentation, this study fixed the number of query images as one. For a 1-way task, the classes in D_{train} are Forest and Background, while for 2-way task, the classes are Forest, Water and Background. The background class contains all other classes present in the dataset apart from forest (for 1-way) and forest, water (for 2-way). Since we are using two different datasets with different labeling policies, we have pre-processed the ground truth masks to ensure that pixel-level labels are represented uniformly in train and test split.

3.2. Method overview

The proposed method first extracts the prototypes from the support images (Section 3.2.1). Unlike the existing methods, this work emphasizes the textures features while computing the support prototypes. Indeed, the pixels corresponding to the forest exhibit characteristic textures. Subsequently, the query prototype is extracted from the query image. Finally, segmentation mask for the query image is computed using metric learning between the prototype and support features. In addition, prototype alignment regularization [28] is also implemented which learns a better representation from the support images (Fig. 1).

3.2.1. Texture based support image prototype

Given the set of support images S, VGG-16 [23] as feature extractor is utilized to compute the features from the support images. The five convolutional block of VGG-16 with maxpool4 layer set to 1 and dilated convolution (dilation = 2) is utilized in this work. Let us represent these features as F_k for the $k^{th}, k = 1, 2, \ldots, K$ image in the support set. The feature vector F_k are resized to the same size as the input image using bilinear interpolation [31]. The texture attention module then enhances these support features to highlight the characteristic textures of forests.

Texture attention module: The textures of the forest region can be an important cue in attaining the generalization. The properties of Gabor filters such as joint resolution in space and frequency along with their capability to model the response of human cortical cells have made Gabor filter a preferred choice for texture analysis [4]. Therefore, in this work the texture features are extracted as an activation of a Gabor Filter based CNN [1]. The filters for the first layer of this CNN are constrained to fit Gabor function. The filter parameters are initialized from the Gabor filter bank, and then updated by backpropagation. Let us represent T_k as the texture features extracted for a given image k using the Gabor CNN. Note that this feature vector is resized to the same size as the input image using bilinear interpolation [31].

The texture features of only the pixels corresponding to the forest are utilized as attention maps. These forest pixels are identified using GrabCut [19] based foreground segmentation approach. GrabCut uses an improved iterative Graph Cut based technique coupled with border matting. The method involved iterating over the *support* images while using the support image ground truth mask as the trimap reference keeping the interaction rectangle to be the size of the image. This is then run through a max-flow method to determine the min-cut, which once converges, separates the background from the foreground (forest). Let us represent the binary mask generated by Grab Cut as $G_k \in 0, 1$ representing the foreground (forest) regions. Note that using a foreground segmentation approach rather than a ground truth segmentation mask eliminates the label noise that might be present in the manually annotated data.

Subsequently, the texture features for only forest pixels are computed. Specifically, the final texture features from the support image is computed as

$$\widehat{T}_k = T_k(x, y)G_k(x, y) \tag{1}$$

where T_k is the texture features and G_k is the binary mask identifying foreground (forest) and background.

The support prototype for forest class is computed using masked average pooling

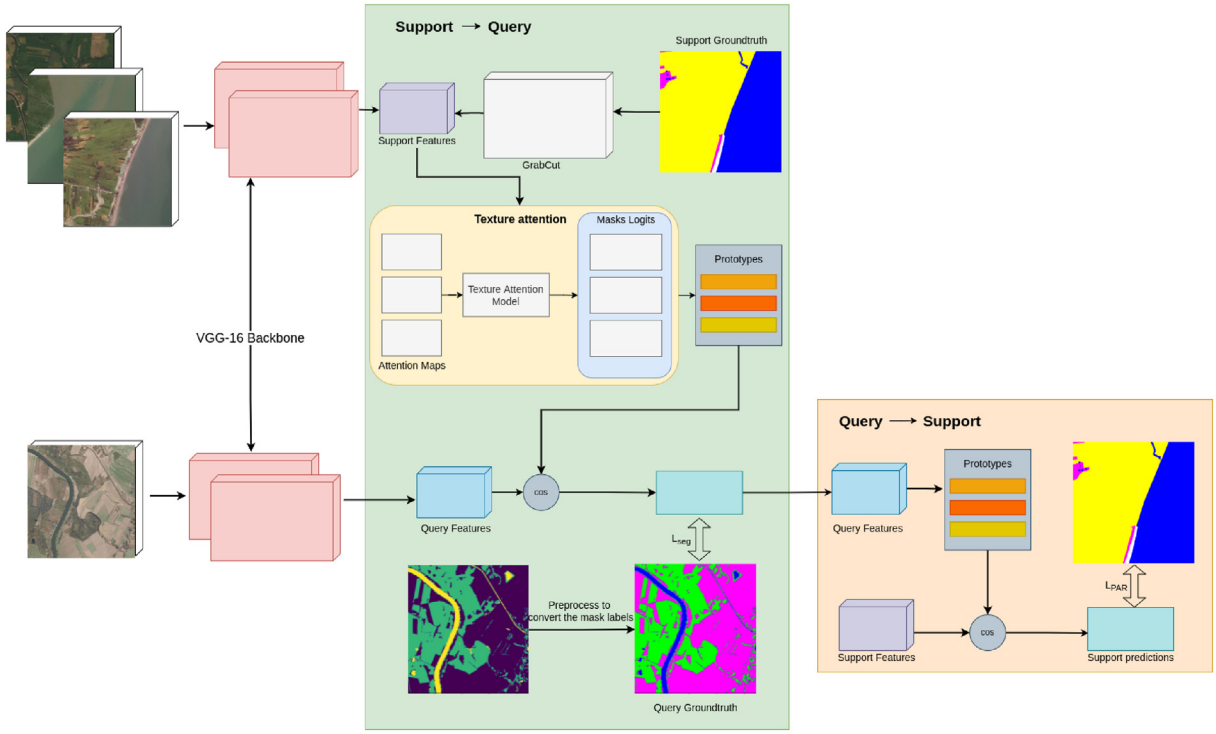

Fig. 1. Overview of the proposed texture-based prototypical network for identifying forests across geographical regions.

$$p_{forest} = \frac{1}{K}\sum_{k}\frac{\sum_{(x,y)}\widehat{T}_k(x,y)F_k(x,y)}{\sum_{(x,y)}G_k(x,y)} \tag{2}$$

The support prototype for the background class p_{bg} is computed from the masked average pooling of the VGG-16 features as follows:

$$p_{bg} = \frac{1}{K}\sum_{k}\frac{\sum_{(x,y)}G'_k(x,y)F_k(x,y)}{\sum_{(x,y)}G'_k(x,y)} \tag{3}$$

where $G'_k(x,y) = !G_k(x,y), !$ is the logical NOT operator.

3.2.2. Metric learning: support – query

For each spatial location in the query image, the optimal class is computed by comparing the distance between support prototype features and features computed from query images using the approach proposed in [28]. A brief summary of the metric learning approach is presented here. More details can be found in [28]. The features from each query image is computed using VGG-16 as a feature extractor. Let us represent query features as F_q. The distance d between query features and support prototype can be defined as

$$dist_j = d(p_j, F_q) \tag{4}$$

where $p_j$is either p_{forest} or p_{bg}. In this work, cosine distance is considered. The segmentation mask $M_q(x,y)$ can be then be computed as [28]:

$$M_q(x,y) = \text{argmax}_j \frac{\exp\left(-\alpha\, dist_j\right)}{\sum_{p_j}\exp\left(-\alpha\, dist_j\right)} \tag{5}$$

where α is a constant. The segmentation loss L_{seg} is then computed as a cross entropy loss between the segmentation mask M_q and the ground truth segmentation mask.

3.2.3. Regularization: query – support

A prototype alignment regularization proposed in [28] is also utilized in this work. In this approach, few shot learning is performed in the reverse direction, i.e., given the query image and its predicted mask, compute the segmentation mask for support image. Specifically, features from query and support images are computed using VGG-16 backbone, and masked average pooling is utilized to compute the query and support prototypes. Note that the texture features are not computed in this regularization step. The cosine similarity-based approach (Section 3.2.2) is then used to compare the query and support prototype and find predicted mask for the support image. Finally, the regularization loss (L_{PAR}) is defined as the cross-entropy loss between the predicted mask for support image and its ground truth segmentation mask. The total loss used for training the model is a weighted sum of segmentation loss (L_{seg}) and the regularization loss (L_{PAR}).

4. Results and Discussion

The proposed model was trained with a Stochastic Gradient Descent optimizer with a learning rate of 1e-3 and a momentum of 0.9 for 30,000 iterations. The model's performance was evaluated by comparing the manually annotated ground truth class labels with the predicted class labels for each pixel in the image and computing mean Intersection over Union (mIoU). The intersection over union metric (IoU) is computed as the total number of pixels common in between both predicted and ground truth (Intersection), divided by the total number of pixels in either the predicted or ground truth labels (Union).

We first evaluated the performance of a supervised deep learning semantic segmentation method to identify forest *across* regions. UNet [18] is a popular encoder-decoder based deep learning architecture for semantic segmentation. We trained UNet on images of DeepGlobe dataset to segment the tropical forests. The performance of this trained UNet was evaluated on identifying temperate forests from LandCover AI dataset. An mIoU of 0.09

was obtained using this trained UNet, which is significantly lower than that obtained using few-shot learning approach (mIoU of 0.47 using proposed approach). Indeed, the flora and canopy cover of tropical forest is significantly different from that of temperate forests. Therefore, a supervised learning approach trained to identify tropical forests would fail to segment temperate forests. In a supervised learning approach, the train and test distribution are assumed to be identical, which would limit their application to the geographical region represented in the training set.

Subsequently, the contribution of each module (viz. Support to Query (S2Q), Query to Support (Q2S) and Texture Attention (TA)) in the proposed few shot semantic segmentation was evaluated. Tables 1 and 2 shows the overall mIoU and forest class IoU for the following scenarios: Only Support to Query (S2Q), Support to Query with Texture Attention (S2Q-TA), Support to Query and Query to Support with Texture Attention (S2Q + Q2S - TA), and Support to Query with Texture Attention and Query to Support with Texture Attention (S2Q - TA + Q2S -TA). It can be observed that the inclusion of the texture attention module in support to query module results in a more accurate segmentation than support to query module without textures. Interestingly, the support to query module with texture performs better than PANet (S2Q + Q2S). Moreover, the inclusion of texture information in both support to query and query to support performs competitively when texture information is included only in support to query module. Therefore, the proposed method includes texture information only in the support to query module.

The proposed method is compared with the existing few-shot semantic segmentation methods (PANet [28], PFENet [27] and ASGNet [14]). As discussed earlier, training dataset contains images from South Asian countries, while the test dataset contains images from central Europe (Section 3). Table 1 shows the mIoU obtained for 1-way and 2-way tasks. It can be observed that the proposed approach outperforms the existing few shot segmentation methods. Notably, the proposed method achieves a higher mIoU (0.47 for 2-way 5-shot) as compared to PANet [28] (0.37 for 2-way 5-shot). Besides, the proposed method achieves an IoU of 0.69 for the forest class on 2-way 5-shot task, which is significantly higher than the existing methods (Table 2). Note that PANet uses VGG-16 features for computing prototypes and does not include texture features. The higher mIoU demonstrates the effectiveness of the proposed texture attention module. Indeed, forest exhibits characteristics texture feature which is distinct from other background classes, such as building, water, roads etc. The higher IoU demonstrates that the proposed approach can transfer the representation learned of South Asian forests to identify Central Europe's forests with the help of only limited support images of Central Europe. It can be seen that the model can identify temperate forests with the help of only *one* labeled image, even though it was trained for identifying tropical forests.

Fig. 2 compares the result obtained using the proposed method with PANet [28] on few images for 2-way 5-shot task. It can be observed that region representing the forest are identified more accurately as compared to PANet. Notably, the proposed method can distinguish between cropland and forest regions as shown in Fig. 2 (First Row, red circles). Note that PANet [28] identifies the cropland as a forest region apparently due to similar color information. Besides, the proposed method can differentiate between other classes (water, open lands) and forest as compared to PANet (Figs. 2, second and third row). Interestingly, PANet identifies the river as a forest region due to the similar color information (Fig. 2, fourth row). However, the proposed method identifies only the forest region. These results demonstrate the effectiveness of the texture attention module in discriminating between the forest and other classes.

To further highlight the importance of texture features for forest identification, the proposed texture attention module was integrated in two existing few-shot segmentation methods (PFENet [27] and ASGNet [14]). For ASGNet, the proposed texture attention map was multiplied with the features from the "expand block" for single prototype learning. For PFENet, the texture attention map was multiplied with the features obtained from the masked global pooling module. The mIoU obtained using these modified models (PFENet + TA, ASGNet + TA) is shown in Table 1. The forest class IoU for these two models are shown in Table 2. It can be seen that the inclusion of the texture attention module in the existing models improves the accuracy of the segmentation algorithm. This improved performance further substantiates the importance of texture information in identifying forests across regions.

The generation of discriminative prototype features is a crucial step in prototypical few-shot segmentation-based approach. In this work for identification of forest cover, we propose to use a foreground-background segmentation (GrabCut) to first identify the pixels representing forest and then computing the features on these foreground pixels. Limiting the region ensures that features from only forest class is learnt. In this work, foreground-background segmentation was utilized rather than using ground truth mask to eliminate the effects of label noise (if present) in the dataset. Indeed, the ground truth mask could also be utilized to identify the pixels corresponding to forest class if the label noise is minimum. We also studied the effect of foreground-background segmentation on the proposed approach. The proposed model was executed using the ground truth mask to identify the pixels corresponding to forest class. Table 3 shows the results obtained in terms of forest class IoU for the proposed method using ground

Table 1
Comparing the performance of the proposed method with existing few shot semantic segmentation methods (PANet [28], ASGNet [14], PFENet [27]) in terms of overall mIoU. Also, shown are mIoU obtained for different ablation studies (i.e. evaluating performance of Support to Query (S2Q), Query to Support (Q2S) and Texture Attention (TA) modules).

Method/mIoU	1-way		2-way	
	1 shot	5 shot	1 shot	5 shot
S2Q	0.121	0.145	0.137	0.145
S2Q-TA	0.348	0.349	0.364	0.407
S2Q + Q2S - TA	0.330	0.371	0.334	0.388
S2Q - TA + Q2S - TA	0.448	0.496	0.451	0.500
PANet (S2Q + Q2S)	0.317	0.364	0.320	0.370
ASGNet	0.181	0.238	0.211	0.239
PFENet	0.260	0.280	0.248	0.321
ASGNet-TA	0.221	0.252	0.261	0.251
PFENet-TA	0.303	0.333	0.296	0.405
S2Q -TA + Q2S (Ours)	0.420	0.471	0.431	0.473

Table 2
Comparing the performance of the proposed method with existing few show semantic segmentation methods (PANet [28], ASGNet [14], PFENet [27]) in terms of forest class IoU. Also, shown is the forest class IoU for the ablation studies (i.e. evaluating performance of Support to Query (S2Q), Query to Support (Q2S) and Texture Attention (TA) modules).

Method/IoU (forest)	1-way		2-way	
	1 shot	5 shot	1 shot	5 shot
S2Q	0.473	0.539	0.496	0.591
S2Q-TA	0.517	0.556	0.534	0.618
S2Q + Q2S - TA	0.510	0.609	0.526	0.641
S2Q - TA + Q2S - TA	0.633	0.691	0.621	0.705
PANet (S2Q + Q2S)	0.465	0.521	0.501	0.580
ASGNet	0.181	0.238	0.211	0.239
PFENet	0.260	0.280	0.248	0.321
ASGNet-TA	0.221	0.252	0.261	0.251
PFENet-TA	0.303	0.333	0.296	0.405
S2Q -TA + Q2S (Ours)	0.626	0.678	0.624	0.699

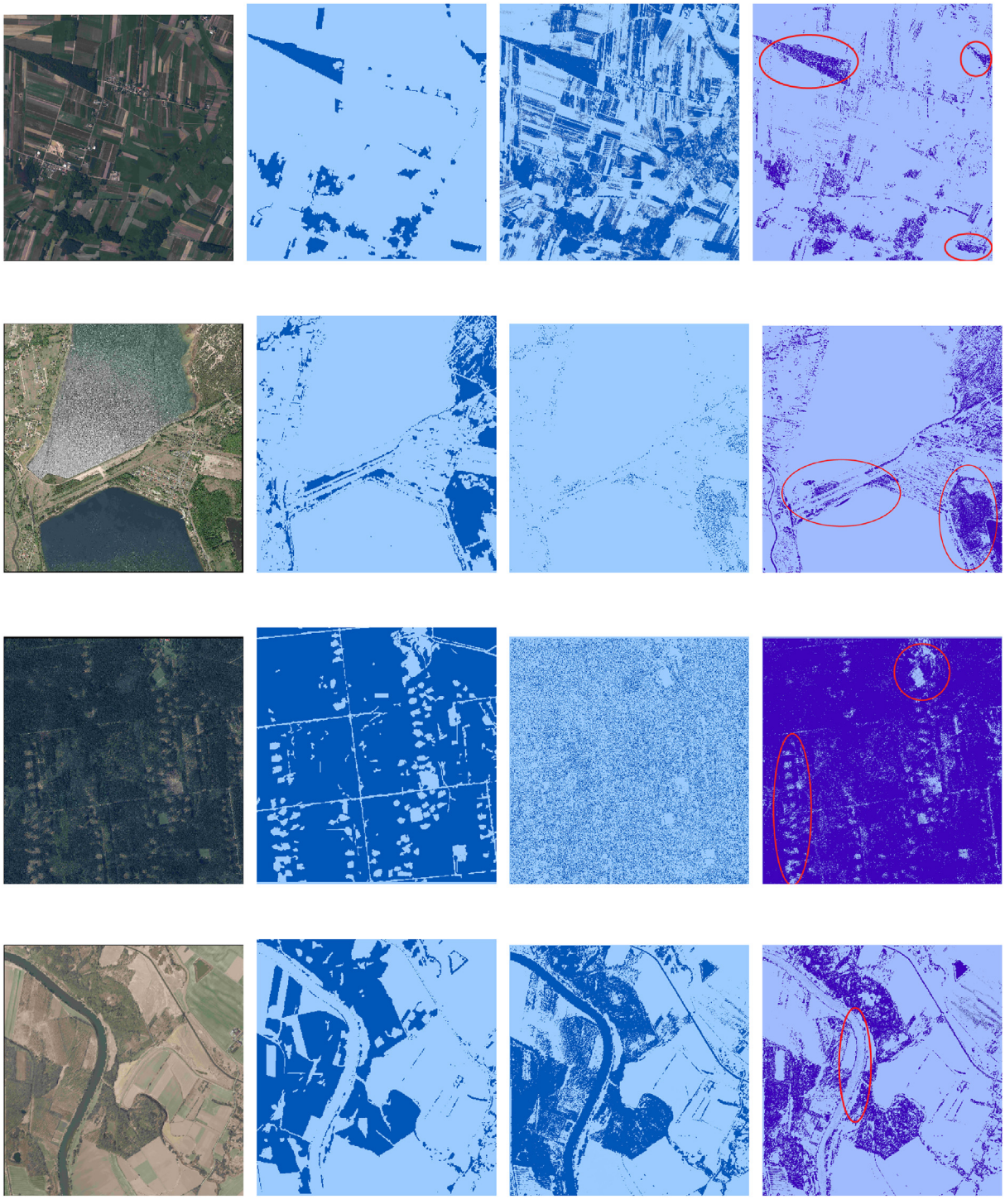

Fig. 2. Comparing the performance of the proposed method with PANet [28] in identifying temperate forest of Central Europe. The first and second column represent the original image and its corresponding ground truth mask, while the third and fourth column shows the results obtained using PANet and the proposed method respectively for 2-way 5-shot task. Dark blue represents forest class while light blue represents other classes.

truth mask and foreground-background segmentation. It can be seen that both approaches perform competitively with a slight minor improvement in *overall* IoU when the foreground-background segmentation (GrabCut) is used to identify forest cover. An alternate approach would be to utilize the ground truth mask to locate the pixels corresponding to forest. However, a few images in LandCover AI dataset suffers from label noise. Therefore, we propose to explicitly identify the pixels corresponding to forests in the image using GrabCut algorithm instead of relying on ground truth mask.

One of the major contributions of this work is the inclusion of texture information in the prototypical few-shot learning approach. The

Table 3
Evaluating the performance of the proposed model with foreground-background segmentation (GrabCut) and using ground truth information. The metric shown below is IoU for forest cover.

	Proposed Method (Using Ground Truth)	Proposed Method (Using GrabCut)
1-way 1-shot	0.624	0.626
1-way 5-shot	0.677	0.678
2-way 1-shot	0.623	0.624
2-way 2-shot	0.698	0.699

texture information ensures that the forest can be distinguished from roads, water bodies, etc. However, in certain regions, the texture features of forests and farmlands/agricultural fields might be similar. This similar texture appearance might adversely affect the performance of the proposed algorithm. Besides, the proposed few-shot segmentation method is based on a prototypical network. The proposed method utilizes episodic learning to mimic the scenario encountered while evaluating few-shot learning algorithms. In episodic learning, the training and the test set are divided into smaller batches of limited data. However, recent studies [13] have indicated that episodic learning might be detrimental to the model's performance. Instead, a non-episodic learning approach could be preferred for prototypical networks, which is much simpler with fewer hyper-parameters and improves the model's performance compared to the episodic learning counterpart. In the future, a non-episodic learning approach will be studied to further enhance the proposed model's performance.

5. Conclusion and future work

The satellite-based forest monitoring system provides a cost-effective tool to identify the change in forest coverage of a region. The existing forest identification methods are based on supervised learning approaches that depend on manually annotated data and are limited to a particular geographical region. This work proposes a texture-based prototype alignment few-shot semantic segmentation network for forest identification. In this work, support prototype features are enhanced by incorporating texture features to highlight the difference in the texture of forest and other classes. In addition, the proposed method focuses on cross-geography generalization wherein the model is trained on the images of South Asia and evaluated on images of Central Europe. The proposed method is compared with the existing few shot segmentation methods and an IoU of 0.62 (1-way 1-shot) is observed for forest class as compared to existing methods (0.46 for PANet [28]). Interestingly, the proposed method can accurately differentiate between forest and other classes (cropland, water, etc) compared to PANet. The results demonstrate that features learned for identifying tropical forests can be transferred to identify temperate forests with the help of very few labeled images (one image for 1-shot) of temperate forests.

CRediT authorship contribution statement

Gokul Puthumanaillam: Formal Analysis, Data Curation, Methodology, Software, Validation, Writing – Orignal Draft, Writing – review and editing. **Ujjwal Verma:** Conceptualization, Formal Analysis, Methodology, Validation, Writing -Orignal Draft, Writing – review and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Alekseev, A. Bobe, Gabornet: Gabor filters with learnable parameters in deep convolutional neural networks, 2019. arXiv preprint arXiv:1904.13204.
[2] R.D.D. Altarez, A. Apan, T. Maraseni, Deep learning u-net classification of sentinel-1 and 2 fusions effectively demarcates tropical montane forest's deforestation, Remote Sens. Appl.: Soc. Environ. 29 (2023), https://doi.org/10.1016/j.rsase.2022.100887.
[3] P.P. de Bem, O.A. de Carvalho Junior, R. Fontes Guimarães, R.A. Trancoso Gomes, Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks, Remote Sens. 12 (2020) 901.
[4] F. Bianconi, A. Fernández, Evaluation of the effects of gabor filter parameters on texture classification, Pattern Recogn. 40 (2007) 3325–3335, https://doi.org/10.1016/j.patcog.2007.04.023.
[5] A. Boguszewski, D. Batorski, N. Ziemba-Jankowska, T. Dziedzic, A. Zambrzycka, Landcover.ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, USA, 2021, pp. 1102–1110, https://doi.org/10.1109/CVPRW53098.2021.00121.
[6] G.D. Caffaratti, M.G. Marchetta, L.D. Euillades, P.A. Euillades, R.Q. Forradellas, Improving forest detection with machine learning in remote sensing data, Remote Sens. Appl.: Soc. Environ. 24 (2021), https://doi.org/10.1016/j.rsase.2021.100654.
[7] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, R. Raskar, Deepglobe 2018: A challenge to parse the earth through satellite images, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018.
[8] H. Ding, H. Zhang, X. Jiang, Self-regularized prototypical network for few-shot semantic segmentation, Pattern Recogn. 133 (2023).
[9] Z. Fan, J.G. Yu, Z. Liang, J. Ou, C. Gao, G.S. Xia, Y. Li, Fgn: Fully guided network for few-shot instance segmentation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 9172–9181.
[10] FAO, Global forest resources assessment 2020 – key findings, 2020. doi: 10.4060/ca8753en.
[11] R. Gupta, L.K. Sharma, Mixed tropical forests canopy height mapping from spaceborne lidar gedi and multisensor imagery using machine learning models, Remote Sens. Appl.: Soc. Environ. 27 (2022), https://doi.org/10.1016/j.rsase.2022.100817.
[12] H.J. Kramer, Earth observation remote sensing: survey of missions and sensors, Springer Science & Business Media, 2012.
[13] S. Laenen, L. Bertinetto, On episodes, prototypical networks, and few-shot learning, Advances in Neural Information Processing Systems 34 (2021) 24581–24592.
[14] G. Li, V. Jampani, L. Sevilla-Lara, D. Sun, J. Kim, J. Kim, Adaptive prototype learning and allocation for few-shot segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8334–8343.
[15] X. Lu, W. Wang, J. Shen, D.J. Crandall, L. Van Gool, Segmenting objects from relational visual data, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2021) 7885–7897.
[16] K. Nguyen, S. Todorovic, Feature weighting and boosting for few-shot segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 622–631.
[17] M.X. Ortega, J.D. Bermudez, P.N. Happ, A. Gomes, R.Q. Feitosa, Evaluation of deep learning techniques for deforestation detection in the amazon forest. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W7, 121–128, 2019.
[18] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention, Springer. (2015) 234–241.
[19] C. Rother, V. Kolmogorov, A. Blake, grabcut: Interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (2004) 309–314.
[20] M. Rußwurm, S. Wang, M. Korner, D. Lobell, Meta-learning for few-shot land cover classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 200–201.
[21] M. Shimada, T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, R. Lucas, New global forest/non-forest maps from alos palsar data (2007–2010), Remote Sens. Environ. 155 (2014) 13–31, https://doi.org/10.1016/j.rse.2014.04.014.
[22] M. Siam, B. Oreshkin, M. Jagersand, Amp: Adaptive masked proxies for few-shot segmentation, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 5248–5257. DOI: 10.1109/ICCV.2019.00535.
[23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv:1409.1556.
[24] H. Tang, J. Armston, S. Hancock, S. Marselis, S. Goetz, R. Dubayah, Characterizing global forest canopy cover distribution using spaceborne lidar, Remote Sens. Environ. 231 (2019).
[25] A.P. Tewkesbury, A.J. Comber, N.J. Tate, A. Lamb, P.F. Fisher, A critical synthesis of remotely sensed optical image change detection techniques, Remote Sens. Environ. 160 (2015) 1–14.
[26] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, J. Jia, Prior guided feature enrichment network for few-shot segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2020) 1, https://doi.org/10.1109/TPAMI.2020.3013717.
[27] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, J. Jia, Prior guided feature enrichment network for few-shot segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 1050–1065, https://doi.org/10.1109/TPAMI.2020.3013717.

[28] K. Wang, J.H. Liew, Y. Zou, D. Zhou, J. Feng, Panet: Few-shot image semantic segmentation with prototype alignment, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 9197–9206.
[29] L.T. Waser, C. Fischer, Z. Wang, C. Ginzler, Wall-to-wall forest mapping based on digital surface models from image-based point clouds and a nfi forest definition, Forests 6 (2015) 4510–4528, https://doi.org/10.3390/f6124386, URL: https://www.mdpi.com/1999-4907/6/12/4386.
[30] M.A. Wulder, T. Hermosilla, G. Stinson, F.A. Gougeon, J.C. White, D.A. Hill, B.P. Smiley, Satellite-based time series land cover and change information to map forest area consistent with national and international reporting requirements, Forestry: Int. J. Forest Res. 93 (2020) 331–343.
[31] X. Zhang, Y. Wei, Y. Yang, T.S. Huang, Sg-one: Similarity guidance network for one-shot semantic segmentation, IEEE Trans. Cybern. 50 (2020) 3855–3865.
[32] T. Zhou, W. Wang, E. Konukoglu, L. Van Gool, Rethinking semantic segmentation: A prototype view, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2582–2593.

Gokul Puthumanaillam completed his BTech from Manipal Institute of Technology, Manipal in Electronics and Communication Engineering in 2022. He is currently working towards his Ph.D. at University of Illinois Urbana-Champaign. His areas of research are decision making in highly adversarial areas where data is unavailable or intrinsically rare.

Dr. Verma received his Ph.D. from Télecom ParisTech, University of Paris-Saclay, Paris, France, in Image Analysis and his M.S. (Research) from IMT Atlantique (France) in Signal and Image Processing. Dr. Verma is currently an Associate Professor and Head of the Department of Electronics and Communication Engineering at Manipal Institute of Technology, Bengaluru, India. His research interests include Computer Vision and Machine Learning; focusing on variational methods in image segmentation, deep learning methods for scene understanding, and semantic segmentation of aerial images. He is a recipient of the "ISCA Young Scientist Award 2017–18" by the Indian Science Congress Association (ISCA), a professional body under the Department of Science and Technology, Government of India. Dr. Verma is the Co-Lead for the Working Group on Machine/Deep Learning for Image Analysis (WG-MIA) of the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society. He is Guest Editor for Special Stream in IEEE Geoscience and Remote Sensing Letters and a reviewer for several journals (IEEE Transactions on Image Processing, IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters). He is also a Sectional Recorder for the ICT Section of the Indian Science Congress Association for 2020–22. Dr. Verma is a Life Member of the Indian Science Congress Association.