Generalized Few-Shot Semantic Segmentation for Remote Sensing Images

Yuyu Jia, Jiabo Li, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Few-shot segmentation (FSS) techniques enhance pixel-level interpretation of unseen classes while reducing reliance on extensive labeled data. However, FSS still faces significant limitations in practical applications: it is restricted to segmenting novel classes and relies on manually constructed support-query pairs during inference. We are the first to introduce the Generalized Few-Shot Segmentation (GFSS) task to remote sensing analysis. It enables simultaneous segmentation of base and novel classes without manual prior interventions. The most intuitive construction is to extend a pre-trained *base* classifier with a *novel* classifier. Nevertheless, since the latter is aggregated from a limited number of supports while the former is trained on abundant data, this disparity inevitably introduces a base class bias, leading to suboptimal segmentation results. This paper proposes a Background-aware Self-mining Prototype Learning (BSPL) strategy to address the issues above. Specifically, we design a dynamic prototype update mechanism during training to enhance the model's adaptability in few-shot scenarios and thereby mitigate the base class bias. Considering the intra-class variation and complex background elements in remote sensing images, we customize segmentation guidance for each query through background-aware self-mining, achieving more precise segmentation performance. Compared to peer algorithms, extensive experiments demonstrate that BSPL achieves the best overall segmentation performance for both base and novel classes, indicating its significant practicality.**

***Index Terms*—Generalized few-shot segmentation, intra-class variation, complex background elements, remote sensing.**

I. Introduction

SEMANTIC segmentation, as a fundamental image analysis technique, is widely applied in various real-world remote sensing applications [1], [2], [3], [4], [5], [6]. Over the past 6-10 years, deep learning-based semantic segmentation algorithms [7], [8], [9] have achieved remarkable results, driven by the availability of large-scale labeled datasets. As a result, challenges related to data collection in specialized scenarios and the high cost of manual annotation have emerged. Consequently, researchers have increasingly focused on reducing the model's dependency on large datasets [10], [11], [12], [13]. Few-Shot Segmentation (FSS) [14], [15], [16], [17], following extensive training on abundant data, aims to leverage a minimal number of annotated samples to guide the segmentation of novel/unseen classes effectively.

This work was supported by the National Natural Science Foundation of China under Grant 62301385, 62471394 and U21B2041.

Yuyu Jia, Jiabo Li, and Qi Wang are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

E-mail: jyy2019@mail.nwpu.edu.cn, ligapt@foxmail.com, crabwq@gmail.com.

Qi Wang is the corresponding author.

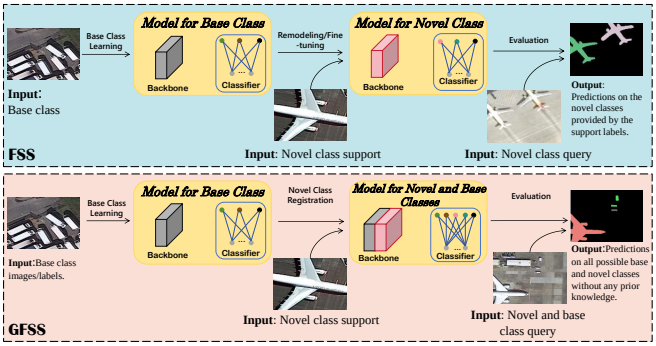

Fig. 1. Differences in task settings between FSS and GFSS.

FSS datasets are composed of a *base* set with abundant labeled samples and a *novel (unseen)* set. To simulate the few-shot scenario, FSS models typically utilize a meta-learning paradigm that divides the base set into multiple episodes during training. In each episode, a few pixel-level labeled supports are aggregated into class prototypes and then used to guide the segmentation of queries within that class. Manually selected support-query pairs are fed into the model during inference to obtain segmentation results for novel queries. Although FSS has enhanced the practicality of pixel-level image interpretation to some extent, it encounters three critical limitations when attempting to advance further in real-world applications [18], [19]: (i) FSS requires that categories in the supports must also be present in the queries, necessitating a manual selection to construct support-query pairs. This introduces an unrealistic prior knowledge assumption for unseen scenarios. (ii) Even when queries encompass multiple categories, most FSS methods are restricted to binary segmentation. (iii) Although performance on unseen classes has improved, the ability to effectively segment base classes has not been adequately preserved, as shown in Fig. 1(a).

With these considerations in mind, we introduce the task of Generalized Few-Shot Segmentation (GFSS) specifically for remote sensing imagery (Fig. 1(b)). Without manually constructed support-query pairs, GFSS enables simultaneous segmentation of both known and unknown classes in a query sample, aligning better with real-world application demands [20], [21]. The most intuitive construction strategy is: one initially trains a base classifier using ample base class data and subsequently aggregates a novel classifier from the novel supports through methods like prototypical learning. Finally, the two classifiers are unified to perform the segmentation task. However, this strategy still faces three significant challenges that severely impact the performance: (i)The base classifier

is trained on a substantial dataset, while the novel classifier is derived from a limited number of supports. This disparity inevitably results in a bias towards the base classes. (ii) The significant intra-class variances in remote sensing images prevent a limited number of supports from providing high-quality class prototypes. (iii) The diverse background elements in remote sensing images render a global background prototype representation unreliable.

In this work, we propose a novel framework for addressing the GFSS task in remote sensing images, called Background-aware Self-mining Prototype Learning (BSPL). Overall, we build on prototypical learning [22], [23], [24] from prior FSS research, using class prototypes as classifiers and deriving segmentation results based on metric space distances. To address base class bias resulting from data volume disparity, we propose a Dynamic Prototype Updating (DPU) mechanism. Each batch is considered a pseudo-episode, from which prototypes of a few samples are extracted to simulate the few-shot scenario and used to update the base classifier, thereby enhancing the model's adaptability in few-shot situations. Considering that a handful of supports cannot cover significant intra-class variations and a single global prototype struggles to represent diverse backgrounds, we design a Background-aware Self-mining (BS) strategy to customize segmentation guidance for each query. Embracing the notion that pixel feature similarity within the same object surpasses that between different objects [25], BS leverages the query prototypes to guide their segmentation, thereby attaining more reliable performance.

It is worth noting that the proposed BS module introduces no additional learnable parameters. Instead, it only adds a minimal amount of basic computation during the inference phase and does not participate in the model's training process. In summary, our contributions are as follows:

1) For the first time, we introduce a task more aligned with practical applications for remote sensing images—Generalized Few-Shot Segmentation (GFSS) and present the BSPL framework as an effective solution.
2) We design a dynamic prototype updating mechanism and background-aware self-mining strategy to address the base class bias and challenges arising from the inherent characteristics of remote sensing images, *i.e.,* intra-class variances and diverse background elements.
3) Extensive experiments demonstrate that the proposed BSPL significantly outperforms traditional FSS approaches in the GFSS task and sets a new state-of-the-art benchmark.

II. Related Works

A. Few-Shot Learning

Few-shot learning (FSL) has been a perennial and classic topic in machine learning, addressing two particular challenges: (i) Modeling a category recognizer with a few labeled samples. (ii) Generalization to unseen categories. Over years of research, metric-based methods [26], [27], [22], [28], [29] have dominated, mapping support and query samples into a shared metric space and predicting class labels based on the distance between query samples and class prototypes. Recently, with the rise of language models [30], [31], [32] researchers have discovered that semantic description knowledge can effectively complement the limitations of a single visual modality, leading to the development of higher-quality classifiers [33], [34], [35], [36]. In the field of remote sensing, FSL has also seen rapid advancements. For instance, SPNet [37] proposes a two-branch structure equipped with prototype self-calibration and inter-calibration mechanisms to optimize the representativeness of prototypes. DA-CAA [38] focuses on the interaction between the source and target domains, transferring class-level distribution information learned from the base classes to the novel classes, thereby mitigating the issue of source domain bias. ProtoConViT [39] posits that limited training samples contain rich spatial contextual information for generalizable discriminative knowledge, introducing a heterogeneous prototype distillation approach to fuse prototypes from CNN and ViT models.

B. Few-Shot Segmentation

Few-shot segmentation (FSS) extends the principles of FSL to a more advanced level by performing pixel-level analysis of images. It typically follows a dual-branch structure [14]: one branch generates segmentation guidance from supports, while the other performs segmentation on queries. According to the segmentation decoding approach, FSS methods are generally categorized into prototype-based [40], [41], [42], [43] and dense comparison-based paradigms [44], [45], [46]. In recent years, advanced FSS methods have proliferated in response to the more challenging nature of remote sensing images. For example, HMRE [47] devises a mutual enhancement strategy for representations to activate shared features between support and query sets. MGCL [48], considering the challenge of low foreground-background contrast in remote sensing images, proposes a mask-guided feature enhancement method. HPR [49] employs a parameter-free adaptive approach to generate multiple representative prototypes, compensating for the limitations of a single prototype's insufficient semantic expression. HSE [50] is the first to propose leveraging class description text information to enrich segmentation guidance, significantly improving performance. DMNet [51] introduces a class-shared semantic mining module that effectively suppresses irrelevant feature contamination by capturing shared semantics between support and query image pairs.

Nevertheless, the FSS task setting still falls short of practical application requirements, as it does not account for base classes during inference and is limited to binary segmentation.

C. Generalized Few-Shot Segmentation

Generalized Few-shot Semantic Segmentation (GFSS), a more challenging task, is first introduced by Tian *et al* [52]. It simultaneously considers the segmentation of both base and novel classes during inference, without the need for paired support-query construction. Currently, relevant research is limited and primarily focuses on natural images. POP [53] orthogonal constraints to encourage prototype orthogonality, thereby mitigating the impact on base classes when updating for novel classes. PCN [54] addresses base class

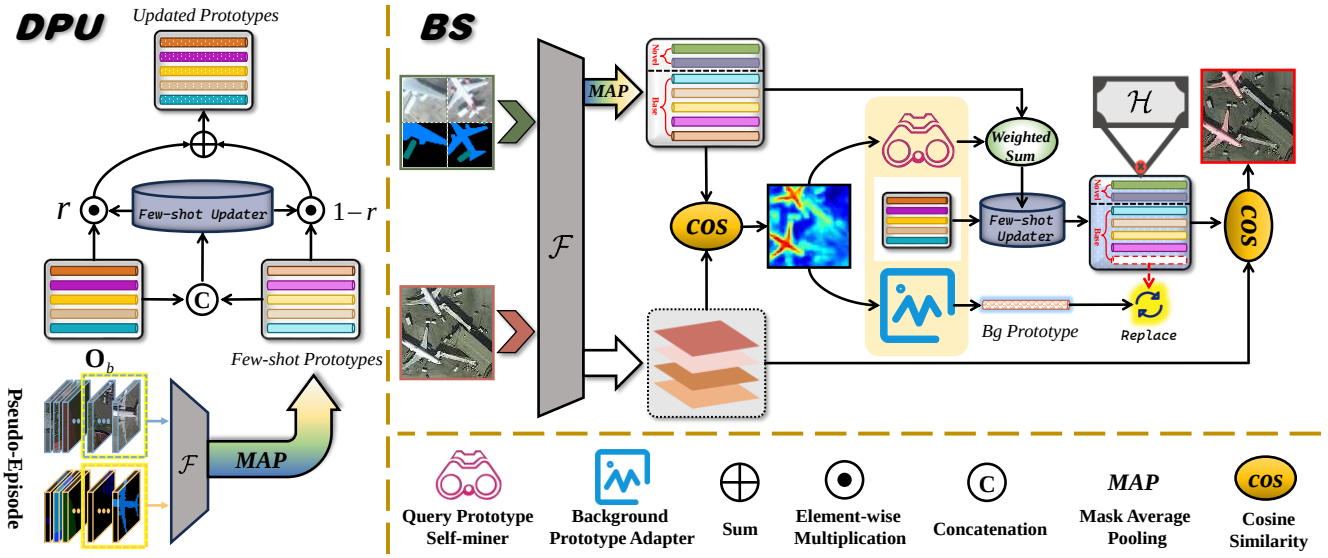

Fig. 2. Illustration of the BSPL framework. We introduce a DPU mechanism in the training process, treating each batch as a pseudo-episode to optimize a few-shot updater. This grants the model adaptability to few-shot scenarios and helps mitigate base class bias. In inference, the proposed BS module customizes segmentation guidance for each query on two levels. First, a query prototype self-miner aggregates high-confidence foreground region features from the samples to mitigate intra-class variation. Additionally, we design a pixel-level background prototype to address the complex background elements in remote sensing images. Finally, we perform a weighted fusion of the classifiers generated at each stage to achieve pixel-level discrimination across all categories.

bias in segmentation results by integrating classifier scores rather than parameters and introduces a prediction calibration module. Visual prompts, as demonstrated in [55], enhance segmentation performance for novel classes and maintain an accurate representation of base classes, even when learned from limited samples. Model bias is thoroughly examined in [56], aligning base and novel class learning in few-shot scenarios, and proposing a class-agnostic foreground-context awareness strategy.

However, there is currently no research on FSS within the field of remote sensing. This paper introduces an effective solution tailored to the unique characteristics of remote sensing imagery, advancing the practical application of pixel-level analysis in few-shot scenarios.

III. Methodology

A. Preliminaries

1) Standard few-shot segmentation: A standard few-shot segmentation dataset is divided into two class-disjoint sets of classes: $\mathcal{C}_b \cap \mathcal{C}_n = \emptyset$. Under the meta-learning paradigm, training and testing are performed in episodes $\mathcal{T} = \{\mathcal{S}, \mathcal{Q}\}$, each containing paired support $\mathcal{S} = \{x_k, m_k\}_{k=1}^{K}$ and query $\mathcal{Q} = \{x_q\}_{q=1}^{Q}$ sets. Let $x_* \in \mathbb{R}^{H\times W\times 3}$ and $m_* \in \mathbb{R}^{H\times W}$ denote an image (shot) and its corresponding binary mask, respectively. Using limited supervision from the support set, the model is expected to predict the segmentation result $y_q \in [0,1]^{H\times W\times 2}$ for a query sample x_q. It is important to note that the segmentation result y_q contains predictions only for the background class c_{bg} and a specific novel class $c_n \in \mathcal{C}_n$. Furthermore, FSS inherently involves an unreasonable prior assumption that the support set must contain the target classes present in the query.

2) Generalized few-shot segmentation: The generalized setup extends FSS towards practical applications. Specifically, it eliminates the unreasonable support-query pair prior. Through a two-phase process: training on a base set with abundant samples in the first phase, followed by incorporating limited support information for novel classes in the second, the GFSS model achieves prediction across all categories, meaning $y_q \in [0,1]^{H\times W\times(1+|\mathcal{C}_b|+|\mathcal{C}|_n|)}$.

3) The baseline of generalized few-shot segmentation: GFSS models $\mathcal{M}$ follow a common segmentation framework, which can be decomposed into a feature extractor $\mathcal{F}$ and a classifier $\mathcal{H}$:

$$\mathcal{M} = \mathcal{F} \circ \mathcal{H}. \tag{1}$$

The feature extractor maps the input image to a d-dimensional latent space, where prototypes of the same dimensionality serve as the classifier. Base prototypes (including the background) $\boldsymbol{O}_b \in \mathbb{R}^{(|\mathcal{C}_b|+1)\times d}$ are used to predict base classes, while novel prototypes $\boldsymbol{O}_n \in \mathbb{R}^{|\mathcal{C}_n|\times d}$ are used to predict novel classes.

During the training phase, we optimize the base prototypes $\boldsymbol{O}_b$ using the abundant base class data combined with a cross-entropy loss function. The novel class prototypes can be derived from their corresponding support sets through mask average pooling (*MAP*) as follows:

$$p_n^c = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{h,w} m_k^c(h,w)\mathcal{F}(x_k^c)(h,w)}{\sum_{h,w} m_k^c(h,w)}, \tag{2}$$

where $\forall p_n^c \in \boldsymbol{O}_n$, x_k^c and m_k^c represent the k-th sample and its binary mask for class c in the support set. The classifier $\mathcal{H}$ can be represented as:

$$\mathcal{H} = \boldsymbol{O}_b \cup \boldsymbol{O}_n. \tag{3}$$

Finally, we use cosine similarity φ as the metric to obtain the classification results for each pixel in the query sample:

$$y_q(h,w) = \arg\max_c \frac{\exp(\varphi(\mathcal{F}(x_q)(h,w), p^c))}{\sum_{p^c \in \mathcal{H}} \exp(\varphi(\mathcal{F}(x_q)(h,w), p^c))}, \quad (4)$$

where $h \in \{1, \cdots, H\}, w \in \{1, \cdots, W\}$, $c \in \{bg, 1, \cdots, |\mathcal{C}_b| + |\mathcal{C}_n|\}$, and bg denotes the background.

B. Overview

This paper proposes the Background-aware Self-mining Prototype Learning (BSPL) framework to tackle the generalized few-shot semantic segmentation (GFSS) task in remote sensing imagery. This framework includes a Dynamic Prototype Updating (DPU) mechanism III-C and a Background-aware Self-mining (BS) strategy III-D, with the overall training and inference process illustrated in Fig. 2. The left part describes the DPU mechanism. It treats each batch during training as a pseudo-episode, using few-shot prototypes to dynamically update the base prototypes, enhancing the model's adaptability in few-shot scenarios. The right part illustrates the BS strategy, which is only involved in the inference process. It customizes segmentation guidance for each query and refines the background prototypes to better align with the remote sensing image characteristics. Finally, the prototypes generated by the BS strategy are combined with the base class prototypes from the training phase to form the final classifier III-E, which produces the segmentation results.

C. Dynamic Prototype Updating (DPU)

Original base class prototypes $\boldsymbol{O}_b$ learn inclusive category perception from sufficient data. In contrast, prototypes of novel classes $\boldsymbol{O}_n$ are obtained by average pooling a small number of supports. Directly merging the two according to Eq. 3 will inevitably result in a bias toward the base class. To render it manageable, we employ the DPU mechanism during training to endow the classifier with few-shot adaptability. Each batch is treated as a pseudo-episode, we randomly select a handful of samples containing a subset of base classes $\mathcal{C}_{fs}$ and extract their few-shot prototypes $\boldsymbol{O}_{fs}$ through the MAP operation as:

$$\boldsymbol{O}_{fs,c} = \begin{cases} \frac{\sum_{h,w} m^c(h,w)\mathcal{F}(x^c)(h,w)}{\sum_{h,w} m^c(h,w)} & c \in \mathcal{C}_{fs} \\ p^c & \text{Otherwise} \end{cases}, \quad (5)$$

where $c \in \{bg, 1, \ldots, |\mathcal{C}_b|\}$, and $\forall p^c \in \boldsymbol{O}_b$. Subsequently, $\boldsymbol{O}_{fs}$ and $\boldsymbol{O}_b$ are concatenated and fed into a meticulously designed few-shot updater $\mathcal{U}$ to obtain the update coefficient:

$$u = \mathcal{U}([\boldsymbol{O}_b; \boldsymbol{O}_{fs}]) \in \mathbb{R}^{(|\mathcal{C}_b|+1)\times 1}. \quad (6)$$

Thus, dynamically updated prototypes can be expressed as:

$$\boldsymbol{O}_{up} = u * \boldsymbol{O}_b + (1-u) * \boldsymbol{O}_{fs}. \quad (7)$$

Finally, the cross-entropy loss is leveraged to optimize updated prototypes $\boldsymbol{O}_{up}$ and the few-shot updater $\mathcal{U}$.

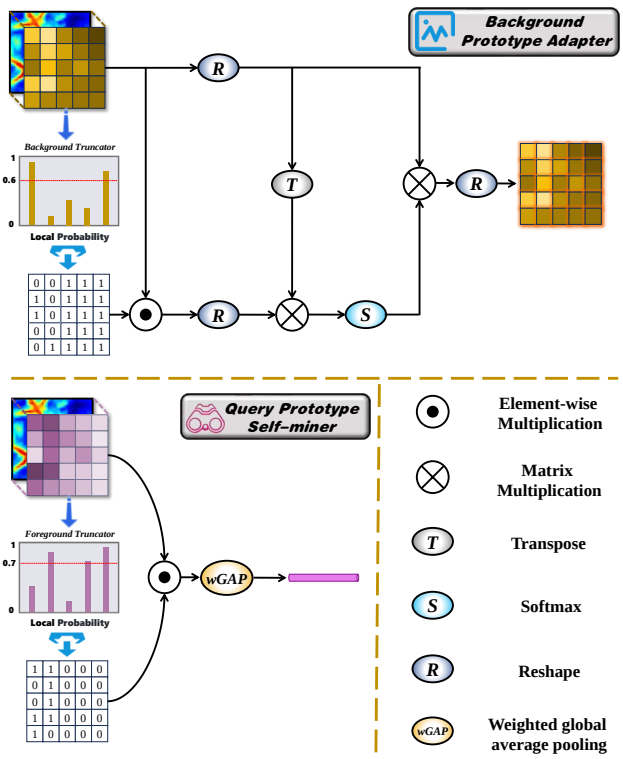

Fig. 3. Implementation process of the query prototype self-miner and background prototype adapter.

D. Background-aware Self-mining (BS)

1) Self-mining query prototypes: Due to significant intra-class variability in remote sensing images, the limited supports are insufficient for comprehensive class representation, hindering reliable segmentation guidance. The classic Gestalt similarity law [57] posits that parts within the same entity are more similar to each other than to parts of different entities within the same category. Inspired by this, we customize segmentation guidance for each query by self-mining its query prototype during inference time, as in the upper part of Fig. 3.

Assuming the support set contains the category subset $\mathcal{C}_{sp}$, which includes all novel classes $\mathcal{C}_n$ and a portion of base classes, the support prototype can be computed using the *MAP* operation as shown in Eq. 2:

$$\boldsymbol{P}_{sp,c} = \begin{cases} MAP(m_k^c, \mathcal{F}(x_k^c)) & c \in \mathcal{C}_{sp} \\ p^c & \text{Otherwise} \end{cases}, \quad (8)$$

where $c \in \{bg, 1, \ldots, |\mathcal{C}_b|, \ldots, |\mathcal{C}_b| + |\mathcal{C}_n|\}$, and $\forall p^c \in \boldsymbol{O}_b$. The coarse-grained probability distribution map for query x_q can be obtained using cosine similarity:

$$\mathcal{D}_0 = softmax(\varphi(\boldsymbol{P}_{sp,c}, \mathcal{F}(x_q))). \quad (9)$$

We can similarly derive the query prototype through the query mask. However, considering the query mask cannot be directly accessed during inference, a soft truncation method is employed on $\mathcal{D}_0$ to generate the estimated query mask:

$$\mathcal{D}_{fg} = \mathbb{I}(\mathcal{D}_0 > \tau_{fg}) * \mathcal{D}_0, \quad (10)$$

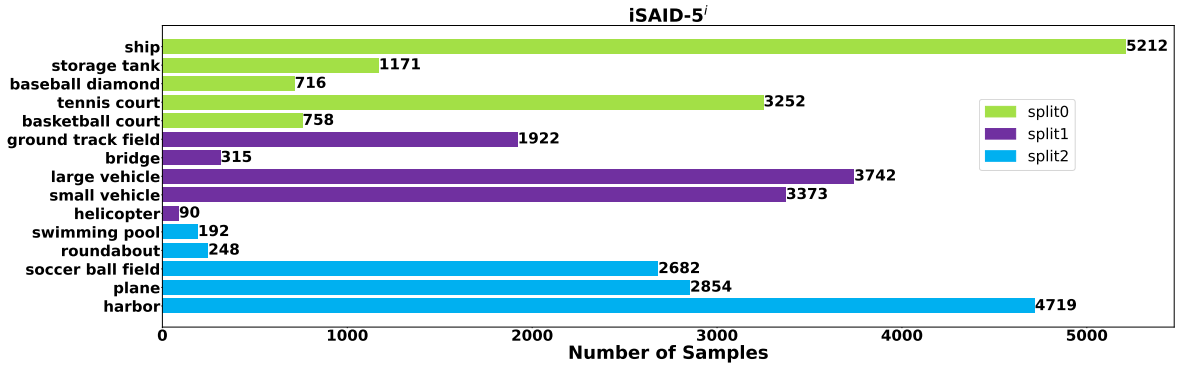

Fig. 4. Category distribution and split division of the iSAID-5^i dataset.

where $\mathbb{I}$ is the indicator function and τ_{fg} is used as the foreground threshold to control the strictness of sampling. Thus, the query prototype is formulated as follows:

$$\boldsymbol{P}_{qr} = MAP(\mathcal{D}_{fg}, \mathcal{F}(x_q)). \tag{11}$$

To fully integrate the support and query prototypes, we apply the weighted summation to obtain the inference prototypes:

$$\boldsymbol{P}_{infer} = r_{infer} * \boldsymbol{P}_{qr} + (1 - r_{infer}) * \boldsymbol{P}_{sp}, \tag{12}$$

where $r_{infer} = max(\varphi(\boldsymbol{P}_{sp}, \boldsymbol{P}_{qr}), 0)$, it serves to assess the reliability of the query prototype. If the correlation between the two prototypes within a specific category is relatively high, the customized segmentation guidance, *i.e.,* query prototype, is given greater weight. Conversely, a lower correlation suggests that the query prototype introduces irrelevant information, thereby reducing its weight.

2) Background prototype adaptation: Remote sensing images are rich in geospatial information, often resulting in complex backgrounds. For example, if the foreground is a car, the background may include diverse elements like buildings, roads, and vegetation. Consequently, a global prototype is inadequate for accurately representing the background class.

Therefore, we plan to adapt pixel-level background segmentation guidance for each query sample. The specific process is illustrated in the lower part of Fig. 3. The idea is to aggregate high-confidence background information probabilistically onto each pixel. Specifically, we first collect pixel-level features that tend to belong to the background category:

$$\boldsymbol{F}_{bg} = \mathcal{R}(\mathcal{F}(x_q) * \mathcal{D}_{bg}) \in \mathbb{R}^{C \times B}, \tag{13}$$

where $\mathcal{D}_{bg} = \mathbb{I}(\mathcal{D}_0 > \tau_{bg})$, $\mathcal{R}$ represents the reshape operation, B denotes the number of collected background pixels, and τ_{bg} is the background truncation threshold. Then, we generate a probability map on x_q for each pixel concerning all background features:

$$\mathcal{A} = softmax(\boldsymbol{F}_{bg} \otimes \mathcal{F}(x_q)) \in \mathbb{R}^{B \times H \times W}, \tag{14}$$

where $\otimes$ stands for the matrix multiplication. The probability map is ultimately used to weigh and aggregate the background features at each pixel, adapting the pixel-level background segmentation guidance:

$$p^{bg} = \boldsymbol{F}_{bg} \otimes \mathcal{A} \in \mathbb{R}^{C \times H \times W}. \tag{15}$$

E. Classifier Integration

For the base class classifier, $\boldsymbol{O}_b$, which undergoes sufficient data optimization during training, tends to be more inclusive. In contrast, $\boldsymbol{P}_{infer}$, generated during inference, is tailored to specific few-shot tasks. Therefore, we leverage the pre-trained few-shot updater $\mathcal{U}$ to achieve a complementary integration of both. Additionally, p^{bg} replaces the background segmentation guidance, generated from Eq. 15. Formally, the base class classifier $\mathcal{H}_b$ can be expressed as:

$$\mathcal{H}_{b,c} = \begin{cases} p^{bg} & c = bg \\ \tilde{u} * \boldsymbol{O}_{b,c} + (1 - \tilde{u}) * \boldsymbol{P}_{infer,c} & \text{Otherwise} \end{cases}, \tag{16}$$

where $\tilde{u} = \mathcal{U}([\boldsymbol{O}_b; \boldsymbol{P}_{infer}[bg, 1, \cdots, |\mathcal{C}_b|]])$ represents the update coefficient and $c \in \{bg, 1, \ldots, |\mathcal{C}_b|\}$. The novel class classifier $\mathcal{H}_n$ is directly derived from the inference prototype generated by Eq. 12:

$$\mathcal{H}_n = \boldsymbol{P}_{infer}[|\mathcal{C}_b| + 1, \cdots, |\mathcal{C}_b| + |\mathcal{C}_n|]. \tag{17}$$

The final classifier can be expressed as:

$$\tilde{H} = \mathcal{H}_b \cup \mathcal{H}_n. \tag{18}$$

IV. Experiments

A. Datasets

We employ the widely used dataset iSAID-5^i [58] in FSS to compare with peers and perform ablation studies to validate the effectiveness of the proposed method. It encompasses 18,076 samples, each with a size of 256×256 pixels. The 15 covered categories are evenly divided into 3 groups, *i.e.,* $i = \{0, 1, 2\}$. The specific categories and quantities of each split are shown in Fig. 4. The experiment employs cross-validation, where two splits are selected as the base set, and the remaining split is used as the novel set.

B. Implementation Details

The baseline architecture of the proposed BSPL strategy is derived from PSPNet [59], utilizing ResNet50 [60] as the backbone. During the initial training phase, the baseline model employs the DPU mechanism III-C and is optimized on the base set with ample samples, using the standard cross-entropy loss function. In each batch, samples from 5 randomly selected

TABLE I
SEGMENTATION (MIOU) COMPARISON ON THE ISAID-5^i DATASET UNDER 1-SHOT AND 5-SHOT SETTINGS. BOLDED VALUES INDICATE THE BEST PERFORMANCE, WHILE THE SECOND-BEST PERFORMANCES ARE UNDERLINED.

Shot	Methods	Split0		Split1		Split2		Average		
		Base	Novel	Base	Novel	Base	Novel	Base	Novel	Total
1	CANet [41]	38.83	13.39	54.47	6.04	45.96	8.11	46.42	9.18	34.01
	PANet [40]	44.24	14.28	61.50	6.68	48.46	10.02	51.40	10.33	37.71
	PFENet [61]	45.12	14.37	60.46	6.77	48.57	8.26	51.38	9.80	37.52
	SCL [62]	45.67	13.27	61.45	5.02	47.89	6.08	51.67	8.12	37.15
	HMRE [47]	42.55	12.25	59.99	5.56	48.59	7.40	50.38	8.40	36.39
	SDM [58]	43.56	14.05	61.25	6.67	49.52	9.28	51.44	10.00	37.63
	CAPL* [52]	47.44	14.77	61.59	6.76	50.43	13.31	53.15	11.61	39.30
	POP* [53]	48.42	15.96	61.66	7.05	50.93	13.29	53.67	12.10	39.81
	BSPL(*ours*)	**50.19**	**19.97**	**62.90**	**9.11**	**51.79**	**14.58**	**54.96**	**14.55**	**41.49**
5	CANet [41]	39.32	14.12	55.23	7.10	46.38	9.37	46.98	10.20	34.72
	PANet [40]	44.98	15.14	62.39	8.01	49.27	11.30	52.21	11.48	38.63
	PFENet [61]	45.88	15.46	60.94	8.21	49.08	10.22	51.97	11.30	38.41
	SCL [62]	45.64	14.55	62.09	7.14	48.60	8.89	52.11	10.19	38.14
	HMRE [47]	43.47	14.57	60.28	7.57	48.96	8.99	50.90	10.38	37.39
	SDM [58]	43.90	15.19	62.10	8.49	50.74	11.44	52.25	11.71	38.74
	CAPL* [52]	48.51	16.59	63.00	8.66	51.01	14.43	54.17	13.23	40.52
	POP* [53]	49.55	18.23	62.70	8.34	51.40	14.98	54.55	13.85	40.98
	BSPL(*ours*)	**50.82**	**21.04**	**62.93**	**9.95**	**52.14**	**15.64**	**55.30**	**15.54**	**42.05**

TABLE II
SEGMENTATION (MIOU) COMPARISON ON THE ISAID-5^i DATASET OF SPECIFIC CATEGORIES UNDER THE 1-SHOT SETTING. THE PERFORMANCE OF NOVEL CLASSES UNDER DIFFERENT SPLITS IS HIGHLIGHTED WITH A YELLOW BACKGROUND.

Methods	Split	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
CANet [41]	Split0	28.37	22.67	10.72	2.86	2.32	9.31	39.01	63.11	52.97	24.65	49.41	53.45	1.07	80.10	7.47
	Split1	55.02	74.30	73.25	82.14	60.19	11.65	0.24	10.64	7.68	0.20	43.44	58.45	66.49	16.17	6.41
	Split2	43.25	73.10	73.88	74.24	36.15	29.74	13.24	61.35	44.60	3.03	0.13	0.38	12.92	4.37	22.73
PANet [40]	Split0	23.06	29.73	6.72	9.52	2.37	23.46	33.27	61.40	53.34	31.65	59.29	56.36	24.76	80.77	10.15
	Split1	55.52	71.75	72.21	83.23	59.71	10.45	0.69	11.66	10.19	0.40	49.74	66.81	63.25	75.53	6.85
	Split2	44.05	72.67	75.58	75.69	35.30	35.14	30.00	62.32	47.12	2.03	0.28	2.09	7.30	13.37	27.04
PFENet [61]	Split0	24.00	27.97	8.58	10.25	1.07	32.60	33.20	60.27	50.18	28.56	40.22	41.59	51.89	78.99	24.21
	Split1	51.46	71.15	70.89	82.94	59.24	9.83	0.67	7.74	15.32	0.28	43.83	53.89	67.21	74.35	21.63
	Split2	52.34	70.18	62.53	83.72	33.25	37.44	28.21	60.40	50.10	1.42	0.12	4.24	4.80	4.56	27.61
SCL [62]	Split0	27.25	16.43	8.24	11.97	2.46	33.78	34.52	62.40	55.36	26.78	47.44	57.33	48.26	79.06	2.61
	Split1	55.34	72.06	70.12	83.47	57.35	7.46	0.63	8.97	7.57	0.45	43.52	53.66	64.04	68.60	40.00
	Split2	53.32	67.46	60.23	75.47	33.36	37.62	30.26	58.77	50.35	1.60	0.23	1.35	8.54	2.02	18.25
CAPL* [52]	Split0	14.93	23.50	12.79	18.25	4.37	36.47	32.54	57.89	48.57	35.16	45.64	47.23	56.57	81.90	20.28
	Split1	49.90	69.52	68.77	82.05	55.30	8.11	2.53	9.26	10.73	3.16	45.39	54.40	65.90	70.04	38.36
	Split2	51.28	70.43	70.16	82.36	38.84	40.03	30.26	60.47	50.13	2.12	2.36	3.11	11.40	27.46	22.24
POP* [53]	Split0	23.23	25.00	12.82	17.44	1.33	46.83	21.77	59.67	59.90	27.13	25.44	51.03	59.00	82.85	40.67
	Split1	54.03	70.16	68.28	80.66	55.45	10.31	3.10	9.17	9.91	2.76	46.10	56.12	64.47	68.70	44.36
	Split2	52.37	68.46	72.46	79.94	40.37	36.51	32.67	61.09	54.11	3.26	4.21	8.66	19.30	13.49	20.81
BSPL(*ours*)	Split0	33.41	27.51	12.55	20.45	5.91	40.34	34.45	67.93	55.18	28.71	47.62	54.16	60.99	80.85	23.79
	Split1	56.57	73.95	71.20	87.13	53.73	8.21	8.09	11.21	11.83	6.22	50.69	55.40	66.59	54.68	47.56
	Split2	50.97	71.36	61.64	82.31	40.37	36.68	33.00	65.21	55.95	7.51	7.20	7.36	11.36	24.10	21.88

base classes are chosen, with $|\mathcal{C}_{fs}| = 5$. We freeze the first three layers to enhance the generalization of the backbone network to low-level features. The model is optimized using the SGD optimizer. It has an initial learning rate of 0.01 and is trained for 30 epochs with a batch size of 48. In the BS module III-D, the foreground and background thresholds are set to 0.7 and 0.8, respectively. Their effect on model performance will be evaluated in the ablation experiments.

C. Evaluation Metrics

The evaluation consists of 2000 tasks, where all test set samples are involved in the segmentation assessment of both base and novel classes. Performance for each class is measured using IoU, while the average segmentation performance for base classes, novel classes, and all classes is reported as mIoU. To ensure result reliability, all experimental outcomes are averaged across multiple random seeds.

D. Main Quantitative Results

In Table I, we quantitatively compare BSPL with 6 state-of-the-art FSS algorithms and 2 outstanding GFSS techniques. Considering that the classifiers of the selected FSS methods (CANet [41], PANet [40], PFENet [61], SCL [62], HMRE [47], and SDM [58]) all involve prototypes, during the training

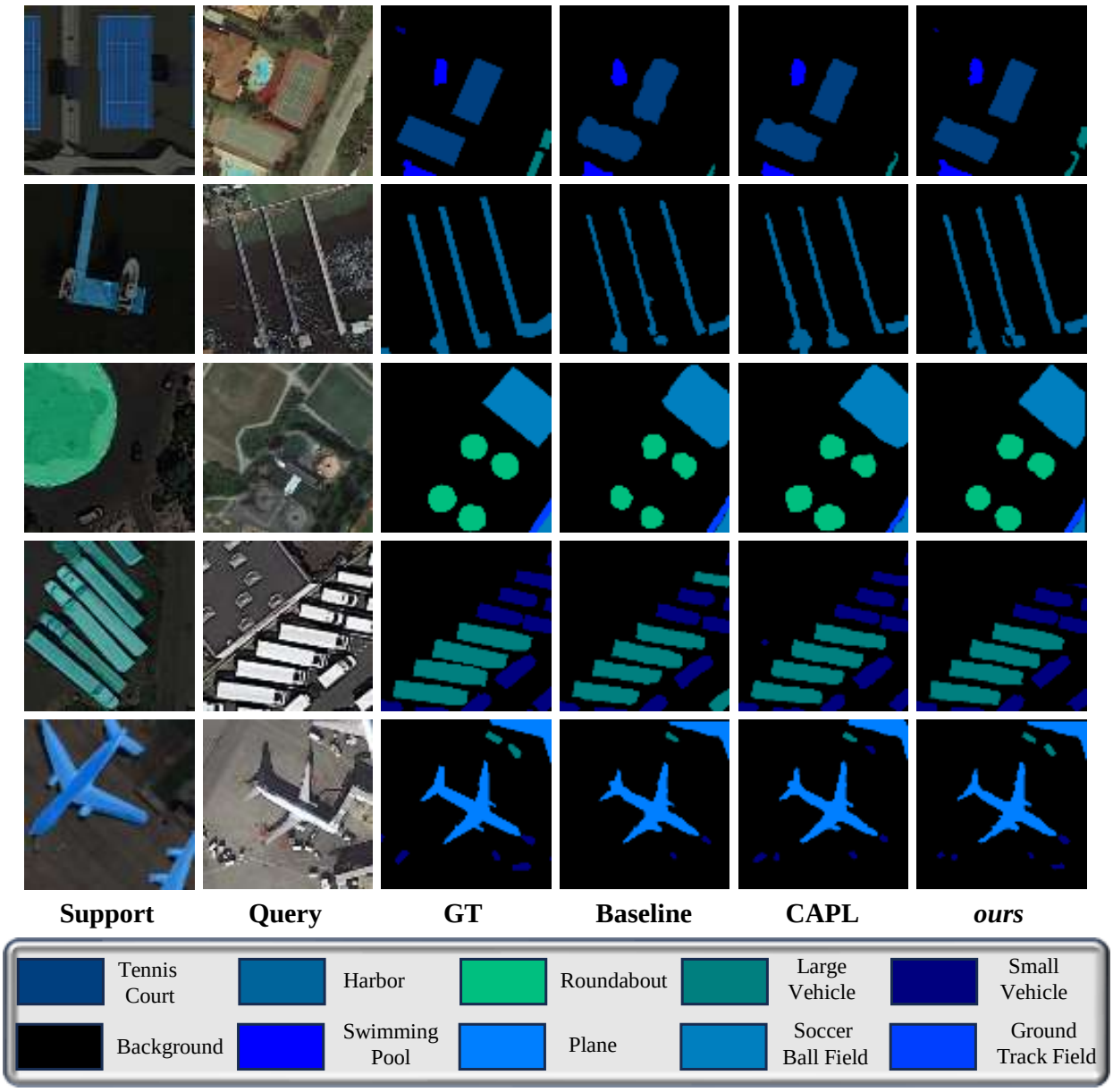

Fig. 5. Visualization of comparison results on the iSAID-5^i dataset for generalized few-shot segmentation under the 5-shot setting.

phase, we use average pooling to obtain the base prototypes. These are then concatenated with the novel prototypes according to Eq. 3 to form the final classifier, enabling the segmentation of base and novel categories. It is important to emphasize that we modified the code solely during the inference stage to update the final classifier without altering the model architecture of the algorithms above. As for GFSS, CAPL [52] and POP [53] are selected as the comparison algorithms (marked with $*$ in the table). We tested them on the dataset using their publicly available code.

As observed in the experiments across all three splits, BSPL achieved a clear advantage for both novel and base classes. Specifically, compared to the similarly strong POP algorithm, BSPL outperformed it by 1.29% in mIoU for base classes and by 2.45% for novel categories, resulting in an overall improvement of 1.68%. Notably, CAPL and POP significantly outperformed the FSS methods. This highlights the limitations of FSS methods in effectively addressing model bias towards base classes through the fusion of two prototypes, making it challenging to handle tasks that require segmentation across all classes. When the shot is expanded to 5, GFSS methods exhibit notable improvements for novel classes but yield only marginal gains for base classes. Specifically, BSPL's mIoU for novel classes increases by 0.99%, while for base classes, the increase is limited to 0.34%. Base prototypes are primarily established during training, leaving additional shots during inference with minimal impact.

The more detailed per-class performance under the 1-shot setting is also reported in Table II. While BSPL's mIoU is comparable to other state-of-the-art methods for certain categories, we observe that those methods perform extremely poorly in some challenging classes. For example, when "C7" is

TABLE III
ABLATION RESULTS ON THE iSAID-5^i DATASET OF KEY COMPONENTS.

Component			1-shot			5-shot		
DPU	SQP	BPA	Base	Novel	Total	Base	Novel	Total
✗	✗	✗	50.46	10.19	37.04	51.68	11.22	38.19
✔	✗	✗	51.12	10.89	37.71	51.70	11.80	38.40
✗	✔	✔	54.32	13.90	40.85	54.29	15.17	41.25
✔	✗	✔	52.24	12.00	38.83	53.42	13.79	40.21
✔	✔	✗	53.07	14.15	40.10	54.15	15.29	41.20
✔	✔	✔	54.96	14.55	41.49	55.30	15.54	42.05

TABLE IV
ABLATION RESULTS ON DESIGN OPTIONS FOR BSPL.

$\mathcal{U}$		r_{infer}		1-shot			5-shot		
Cos	MLP	Cos	MLP	Base	Novel	Total	Base	Novel	Total
✗	✔	✗	✔	54.36	14.51	41.08	55.02	14.99	41.68
✔	✗	✔	✗	53.79	13.82	40.47	54.11	14.33	40.85
✔	✗	✗	✔	54.03	13.94	40.67	54.64	14.85	41.38
✗	✔	✔	✗	54.96	14.55	41.49	55.30	15.54	42.05

a novel class, several methods, including CAPL, achieve only around 3% accuracy. In contrast, BSPL demonstrates more consistent performance across all categories, underscoring its design's robustness and superior generalization capabilities.

E. Visualization Results

Fig. 5 visually demonstrates the advantages of the proposed method through the segmentation results under the 5-shot setting. The first column of the support provides only the mask for the novel class, while the second column requires the segmentation of all categories within the query. The baseline method uses PSPNet as the backbone network and simply concatenates the base and novel prototypes to serve as a pixel-level classifier. Three observations can be made: (i) Due to base class bias, the "baseline method" exhibits poor segmentation capability for novel classes, even resulting in misclassifications. (ii) "CAPL" effectively mitigates base class bias and improves segmentation performance, however, its segmentation of small target edges lacks precision. We attribute this to the foreground-background ambiguity caused by the global background prototype. (iii) BSPL achieves the overall optimal segmentation results, demonstrating not only more balanced performance across all categories but also clear segmentation of small target edges. This underscores the effectiveness of the proposed DPU and BS modules.

F. Ablation Studies

1) Model Component Ablation: We conduct ablation experiments to evaluate the contribution of each component in BSPL to segmentation performance. Here, SPQ (Sec. III-D1) and BPA (Sec. III-D2) refer to the two steps within the BS module. The first row in the experimental setup excludes any additional modules, representing the baseline formed by PSPNet combined with Eq. 3. As shown in Table III, incorporating the DPU strategy to train base class prototypes effectively mitigates base class bias, significantly improving

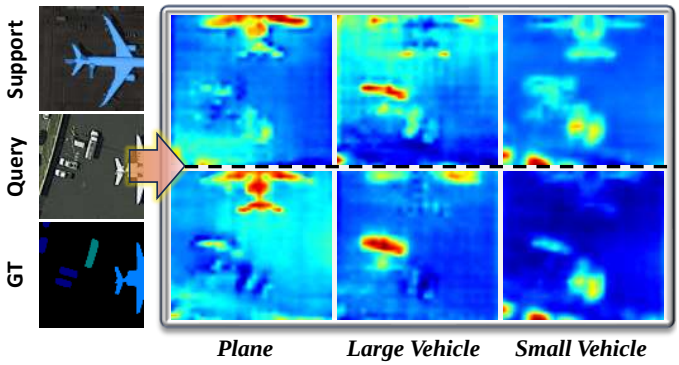

Fig. 6. Impact of different designs in BS. The first row employs the threshold truncation method from [63], while the second row presents our proposed soft threshold truncation strategy.

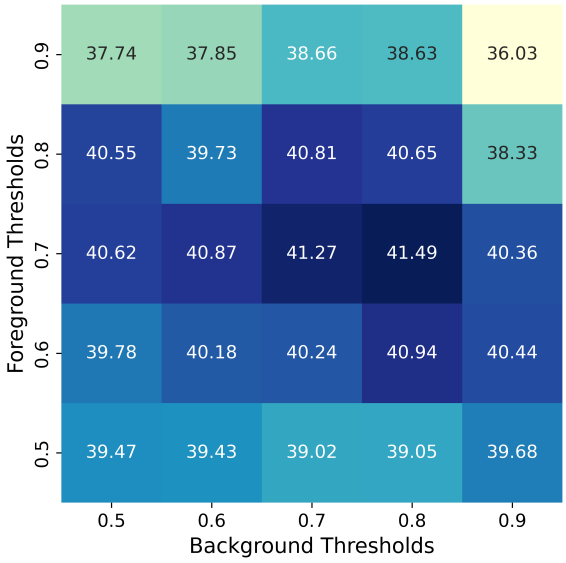

Fig. 7. Impact of the two thresholds in the BS module on the average performance of all categories under the 1-shot scenario.

segmentation performance for novel classes. Building on this, adding either the SQP or BPA components individually results in substantial performance improvements, with SQP playing a more prominent role. This confirms the benefit of customizing foreground-background segmentation guidance for each query in GFSS tasks. Ultimately, when all components are utilized, the model achieves optimal performance.

2) Design Options for BSPL: In BSPL, which involves the construction of the few-shot updater $\mathcal{U}$ (6) and the generation of the weight factor r_{infer} (Eq. 12), we explore how different design approaches affect segmentation performance. In Table IV, "Cos" refers to directly measuring the cosine similarity between inputs, while "MLP" represents concatenating the inputs along the channel dimension and mapping the result using a multi-layer perception. The model achieves optimal performance when $\mathcal{U}$ and r_{infer} are computed using "MLP" and "Cos", respectively. We present two reasons: (i) The few-shot updater $\mathcal{U}$ requires a complementary integration of $\boldsymbol{O}_b$, which represents broad information, and the few-shot prototype $\boldsymbol{O}_{fs}$. MLP is well-suited for this task due

to its strong nonlinear modeling capability. (ii) The weight factor r_{infer} essentially measures the correlation between the support and query prototypes. Low correlation indicates that the query prototype contains category-irrelevant information, necessitating a reduction in its weight. Since cosine similarity can quickly assess high-dimensional sparse data, we use it to compute r_{infer}.

Additionally, SSP [63] similarly employs a threshold truncation method to extract high-confidence foreground information from query samples:

$$\mathcal{D}_{fg} = \mathbb{I}(\mathcal{D}_0 > \tau_{fg}). \tag{19}$$

In contrast, BSPL enhances this by using a soft threshold truncation (Eq. 10) to preserve the probability differences among different pixels. Fig. 6 visualizes the heatmaps of various categories in the query samples to validate the advancement of this approach. The soft threshold truncation strategy in the second row activates the target category regions more accurately, demonstrating that retaining probability differences is beneficial for customizing high-quality segmentation guidance.

3) Ablation Study of Hyperparameters: The BS module involves foreground and background thresholds τ_{fg} and τ_{bg}, both of which assist in collecting high-confidence category information and subsequently influence the model's segmentation performance. We conduct a series of experiments on various hyperparameters under the 1-shot setting, with the results shown in Fig. 7. When $\tau_{fg} \in [0.6, 0.7]$ and $\tau_{fg} \in [0.7, 0.8]$, the model achieves the best performance. This is because the model is typically more adept at identifying background classes, requiring a stricter background threshold to obtain reliable background information.

V. Conclusion

To address the limitations of few-shot segmentation in practical remote sensing applications, we introduce the generalized few-shot segmentation task for the first time, eliminating reliance on manually constructed support-query pairs and enabling segmentation for both base and novel categories. For this task, the proposed Background-aware Self-mining Prototype Learning (BSPL) framework mitigates base class bias caused by data imbalance through a dynamic prototype updating mechanism. Furthermore, considering the inherent intra-class differences and complexity of background elements in remote sensing images, we develop a background-aware self-mining strategy to customize segmentation guidance for each query, enhancing the model's segmentation performance across all categories in few-shot scenarios. Extensive experiments on the iSAID-5^i dataset validate the effectiveness of the proposed design, achieving optimal overall performance.

References

[1] Y. Zhang, P. David, and B. Gong, "Curriculum domain adaptation for semantic segmentation of urban scenes," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2020–2030.

[2] B. Forster, "An examination of some problems and solutions in monitoring urban areas from satellite platforms," *International journal of remote sensing*, vol. 6, no. 1, pp. 139–151, 1985.

[3] S. Liu, Y. Ma, X. Zhang, H. Wang, J. Ji, X. Sun, and R. Ji, "Rotated multi-scale interaction network for referring remote sensing image segmentation," *arXiv preprint arXiv:2312.12470*, 2023.

[4] C. Ru, S.-B. Duan, X.-G. Jiang, Z.-L. Li, Y. Jiang, H. Ren, P. Leng, and M. Gao, "Land surface temperature retrieval from landsat 8 thermal infrared data over urban areas considering geometry effect: Method and application," *IEEE Transactions on geoscience and remote sensing*, vol. 60, pp. 1–16, 2021.

[5] W. Qiao, L. Shen, J. Wang, X. Yang, and Z. Li, "A weakly supervised semantic segmentation approach for damaged building extraction from postearthquake high-resolution remote-sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[6] E. Macioszek and A. Kurek, "Extracting road traffic volume in the city before and during covid-19 through video remote sensing," *Remote Sensing*, vol. 13, no. 12, p. 2329, 2021.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2015.7298965

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 834–848, Apr 2018. [Online]. Available: http://dx.doi.org/10.1109/tpami.2017.2699184

[9] J. Qu, P. Yang, W. Dong, X. Zhang, and Y. Li, "A semi-supervised multiscale convolutional sparse coding-guided deep interpretable network for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[10] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, "Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1144–1158, 2017.

[11] X. Zeng, Z. Wang, Y. Wang, X. Rong, P. Guo, X. Gao, and X. Sun, "Semipscn: Polarization semantic constraint network for semi-supervised segmentation in large-scale and complex-valued polsar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, 2024.

[12] X. Zeng, T. Wang, Z. Dong, X. Zhang, and Y. Gu, "Superpixel consistency saliency map generation for weakly supervised semantic segmentation of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[13] R. Zhou, W. Zhang, Z. Yuan, X. Rong, W. Liu, K. Fu, and X. Sun, "Weakly supervised semantic segmentation in aerial imagery via explicit pixel-level constraints," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[14] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots, "One-shot learning for semantic segmentation," in *Procedings of the British Machine Vision Conference 2017*, Jan 2017. [Online]. Available: http://dx.doi.org/10.5244/c.31.167

[15] M. Siam, B. Oreshkin, and M. Jagersand, "Amp: Adaptive masked proxies for few-shot segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019. [Online]. Available: http://dx.doi.org/10.1109/iccv.2019.00535

[16] B. Yang, C. Liu, B. Li, J. Jiao, and Q. Ye, "Prototype mixture models for few-shot semantic segmentation," in *Computer Vision – ECCV 2020,Lecture Notes in Computer Science*, Jan 2020, p. 763–778. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-58598-3_45

[17] K. Nguyen and S. Todorovic, "Feature weighting and boosting for few-shot segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019. [Online]. Available: http://dx.doi.org/10.1109/iccv.2019.00071

[18] K. Huang, F. Wang, Y. Xi, and Y. Gao, "Prototypical kernel learning and open-set foreground perception for generalized few-shot semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 256–19 265.

[19] S. Hajimiri, M. Boudiaf, I. Ben Ayed, and J. Dolz, "A strong baseline for generalized few-shot semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 269–11 278.

[20] Z. Tian, X. Lai, L. Jiang, S. Liu, M. Shu, H. Zhao, and J. Jia, "Generalized few-shot semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 563–11 572.

[21] M. R. I. Hossain, M. Siam, L. Sigal, and J. J. Little, "Visual prompting for generalized few-shot segmentation: A multi-scale approach," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23 470–23 480.

[22] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[23] K. Huang, J. Geng, W. Jiang, X. Deng, and Z. Xu, "Pseudo-loss confidence metric for semi-supervised few-shot learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8671–8680.

[24] H. Zhang and H. Ding, "Prototypical matching and open set rejection for zero-shot semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6974–6983.

[25] Q. Fan, W. Pei, Y.-W. Tai, and C.-K. Tang, "Self-support few-shot semantic segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 701–719.

[26] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal, "Improved few-shot visual classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 493–14 502.

[27] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2, no. 1. Lille, 2015, pp. 1–30.

[28] C. Zhang, Y. Cai, G. Lin, and C. Shen, "Deepemd: Differentiable earth mover's distance for few-shot learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5632–5648, 2022.

[29] Y. Jia, J. Gao, W. Huang, Y. Yuan, and Q. Wang, "Exploring hard samples in multi-view for few-shot remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[30] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[31] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.

[32] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[33] W. Chen, C. Si, Z. Zhang, L. Wang, Z. Wang, and T. Tan, "Semantic prompt for few-shot image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 23 581–23 591.

[34] Y. Jia, Q. Zhou, W. Huang, J. Gao, and Q. Wang, "Like humans to few-shot learning through knowledge permeation of vision and text," *arXiv preprint arXiv:2405.12543*, 2024.

[35] H. Zhang, J. Xu, S. Jiang, and Z. He, "Simple semantic-aided few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 28 588–28 597.

[36] C. Xing, N. Rostamzadeh, B. Oreshkin, and P. O. O Pinheiro, "Adaptive cross-modal few-shot learning," *Advances in neural information processing systems*, vol. 32, 2019.

[37] G. Cheng, L. Cai, C. Lang, X. Yao, J. Chen, L. Guo, and J. Han, "Spnet: Siamese-prototype network for few-shot remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[38] Y. Guo, B. Fan, Y. Feng, X. Jia, and M. He, "Distribution-aware and class-adaptive aggregation for few-shot hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[39] Y. Zhuang, Y. Liu, T. Zhang, L. Chen, H. Chen, and L. Li, "Heterogeneous prototype distillation with support-query correlative guidance for few-shot remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[40] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9197–9206.

[41] C. Zhang, G. Lin, F. Liu, R. Yao, and C. Shen, "Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5217–5226.

[42] Y. Liu, X. Zhang, S. Zhang, and X. He, "Part-aware prototype network for few-shot semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 142–158.

[43] B. Yang, C. Liu, B. Li, J. Jiao, and Q. Ye, "Prototype mixture models for few-shot semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. Springer, 2020, pp. 763–778.

[44] X. Yang, B. Wang, K. Chen, X. Zhou, S. Yi, W. Ouyang, and L. Zhou, "Brinet: Towards bridging the intra-class and inter-class gaps in one-shot segmentation," *arXiv preprint arXiv:2008.06226*, 2020.

[45] C. Zhang, G. Lin, F. Liu, J. Guo, Q. Wu, and R. Yao, "Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9587–9595.

[46] J. Min, D. Kang, and M. Cho, "Hypercorrelation squeeze for few-shot segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 6941–6952.

[47] Y. Jia, J. Gao, W. Huang, Y. Yuan, and Q. Wang, "Holistic mutual representation enhancement for few-shot remote sensing segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[48] S. Li, F. Liu, L. Jiao, X. Liu, P. Chen, and L. Li, "Mask-guided correlation learning for few-shot segmentation in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[49] W. Shen, A. Ma, J. Wang, Z. Zheng, and Y. Zhong, "Adaptive self-supporting prototype learning for remote sensing few-shot semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[50] Y. Jia, W. Huang, J. Gao, Q. Wang, and Q. Li, "Embedding generalized semantic knowledge into few-shot remote sensing segmentation," *arXiv preprint arXiv:2405.13686*, 2024.

[51] H. Bi, Y. Feng, Z. Yan, Y. Mao, W. Diao, H. Wang, and X. Sun, "Not just learning from others but relying on yourself: A new perspective on few-shot segmentation in remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–21, 2023.

[52] Z. Tian, X. Lai, L. Jiang, S. Liu, M. Shu, H. Zhao, and J. Jia, "Generalized few-shot semantic segmentation." in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2022. [Online]. Available: http://dx.doi.org/10.1109/cvpr52688.2022.01127

[53] S.-A. Liu, Y. Zhang, Z. Qiu, H. Xie, Y. Zhang, and T. Yao, "Learning orthogonal prototypes for generalized few-shot semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 319–11 328.

[54] Z. Lu, S. He, D. Li, Y.-Z. Song, and T. Xiang, "Prediction calibration for generalized few-shot semantic segmentation," *IEEE transactions on image processing*, vol. 32, pp. 3311–3323, 2023.

[55] M. R. I. Hossain, M. Siam, L. Sigal, and J. J. Little, "Visual prompting for generalized few-shot segmentation: A multi-scale approach," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 23 470–23 480.

[56] K. Huang, F. Wang, Y. Xi, and Y. Gao, "Prototypical kernel learning and open-set foreground perception for generalized few-shot semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 19 256–19 265.

[57] K. Koffka, *Principles of Gestalt psychology*. routledge, 2013.

[58] X. Yao, Q. Cao, X. Feng, G. Cheng, and J. Han, "Scale-aware detailed matching for few-shot aerial image semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[59] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[61] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia, "Prior guided feature enrichment network for few-shot segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 2, pp. 1050–1065, 2020.

[62] B. Zhang, J. Xiao, and T. Qin, "Self-guided and cross-guided learning for few-shot segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8312–8321.

[63] Q. Fan, W. Pei, Y.-W. Tai, and C.-K. Tang, "Self-support few-shot semantic segmentation," *ECCV*, 2022.

Yuyu Jia received the B.E. degree and the M.S. degree in control theory and engineering from the Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree at the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include few-shot learning, deep learning, and remote sensing.

Jiabo Li received his B.E. degree in Cyberspace Security from Xiamen University, Xiamen, Fujian, China, in 2024. He is currently pursuing his M.E. degree at the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, Shaanxi, China. His research interests focus on few-shot learning and remote sensing.

Qi Wang (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, machine learning, pattern recognition, and remote sensing. For more information, visit the link (https://crabwq.github.io/).