

# Supplementary materials of FGAPA

We sincerely thank the two reviewers for their thorough evaluation of our manuscript and for providing constructive and insightful comments. We are also deeply grateful for the reviewers' positive assessment of the novelty of this work, as well as their recognition of the potential and value of the proposed cross-domain hyperspectral image classification method, FGAPA. Below are our specific responses to the reviewers' concerns regarding FGAPA.

## Response to Reviewer 1

1.The experiments provide good initial evidence for the method, but need to go further in order to fully establish the value of the approach. The training and transfer hinges on the selection of the source domain, so only one experiment here is very limiting. The source domain needs some properties of potential alignment with the target domains, and must somehow be sufficiently rich to cover the possibilities. So it isn't clear how robust the method is in general, or how to pick the source domain to begin with, or how the approach can be updated as new domains are sampled.

**To ensure the effectiveness of the proposed method, the source domain is generally required to cover no fewer classes than the target domain, so as to construct a sufficiently representative metric space for downstream tasks. The Chikusei dataset contains 19 diverse land-cover classes, providing rich prior knowledge for cross-domain few-shot learning. In addition, a significant domain shift exists between Chikusei and the target domains, making it well suited for evaluating the generalization capability of few-shot learning methods under challenging conditions. Therefore, Chikusei is selected as the source domain in this study.**

**To further validate the robustness of our method across different source domains, we additionally conduct experiments using the Hanchuan dataset as the source domain. This dataset also contains a rich variety of land-cover classes (16 classes) and exhibits a substantial domain shift from the target domains. Table 1 reports the classification performance of our method on the Indian Pines, Salinas, and Botswana target domains using Hanchuan as the source domain with only five labeled samples per class, in comparison with representative domain adaptation methods, including DCFSL, MLPA, and the contrastive learning-based CTF-SSCL.**

Table 1:Classification results (mean  $\pm$  std) with Hanchuan dataset as source domain

Method	Indian Pines			Salinas			Botswana		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
DCFSL	69.53 $\pm$ 2.85	80.72 $\pm$ 1.56	65.77 $\pm$ 3.19	91.06 $\pm$ 0.99	94.95 $\pm$ 0.62	90.07 $\pm$ 1.10	97.15 $\pm$ 0.83	97.06 $\pm$ 0.96	96.91 $\pm$ 0.90
CTF-SSCL	70.59 $\pm$ 4.22	81.69 $\pm$ 1.75	66.96 $\pm$ 4.52	89.79 $\pm$ 1.70	93.82 $\pm$ 1.01	88.66 $\pm$ 1.89	95.34 $\pm$ 1.397	95.60 $\pm$ 1.36	94.95 $\pm$ 1.51
MLPA	68.13 $\pm$ 3.21	79.61 $\pm$ 1.84	64.24 $\pm$ 3.51	90.99 $\pm$ 1.07	94.61 $\pm$ 0.85	89.98 $\pm$ 1.18	96.38 $\pm$ 1.13	96.31 $\pm$ 1.18	96.08 $\pm$ 1.23
FGAPA	76.37 $\pm$ 2.56	85.76 $\pm$ 1.18	73.34 $\pm$ 2.78	92.08 $\pm$ 0.87	95.22 $\pm$ 1.03	91.19 $\pm$ 0.97	97.65 $\pm$ 0.97	97.58 $\pm$ 1.09	97.45 $\pm$ 1.05

**As shown in Table 1, when Hanchuan is used as the source domain, FGAPA achieves superior OA, AA, and Kappa across all three target datasets. In terms of OA, FGAPA outperforms DCFSL and MLPA by 6.84% and 8.5% on Indian Pines, respectively, with similar gains observed on Salinas and Botswana. Compared with the contrastive learning-based CTF-SSCL, our method benefits from the adversarial domain alignment strategy in the ACA module, resulting in stronger performance, while lower variance further confirms its superior stability. Overall,**

FGAPA demonstrates consistent robustness and reliability across different source domains, including Chikusei and Hanchuan.

## Response to Reviewer 2

1. Some definitions are missing in the paragraph about FSL in HSIC. What is FSL? What are source and target domains?

Few Shot Learning (FSL) is fundamentally a meta-learning method that can acquire transferable knowledge from different tasks, enabling rapid adaptation to new tasks. It typically follows an N-way K-shot learning setting, where K labeled samples per class are used to train an N-class classifier. As shown in Figure 1, an FSL task consists of a labeled support set and a query set to be predicted. FSL classifies samples by computing the distance between each query feature and the support features of each class, thereby enabling the model to learn how to infer the categories of query samples based on the given support set. In hyperspectral image classification (HSIC), the same framework can be applied: a small number of labeled samples are selected as the support set, while a large number of unlabeled samples serve as the query set, constructing an FSL task for model learning.

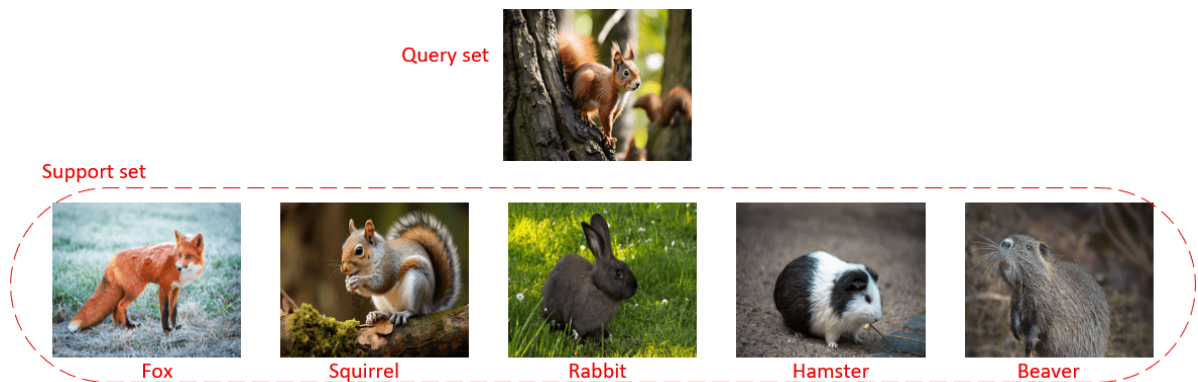


Fig1: Few shot learning

In cross-domain FSL, the source domain refers to a dataset with abundant, fully labeled samples, where the model is initially pre-trained to learn general features and patterns (i.e., to “learn how to learn”). The target domain is the dataset for the downstream task, typically with very few labeled samples and a data distribution and set of classes different from the source domain. The core objective of the entire process is to leverage the prior knowledge acquired from the source domain to quickly adapt to new tasks in the target domain using only a minimal number of labeled samples (e.g., 1–5 per class).

2. Figure 1 is not clear. Acronyms are not defined and do not correspond to the text. What are support features? Why don't we see  $L_{fsl}^t$  and  $L_{fsl}^s$ ? What is the attention score (never mentioned in the text)? Does the prototype bank correspond to  $\mathbf{p}_i$ ?

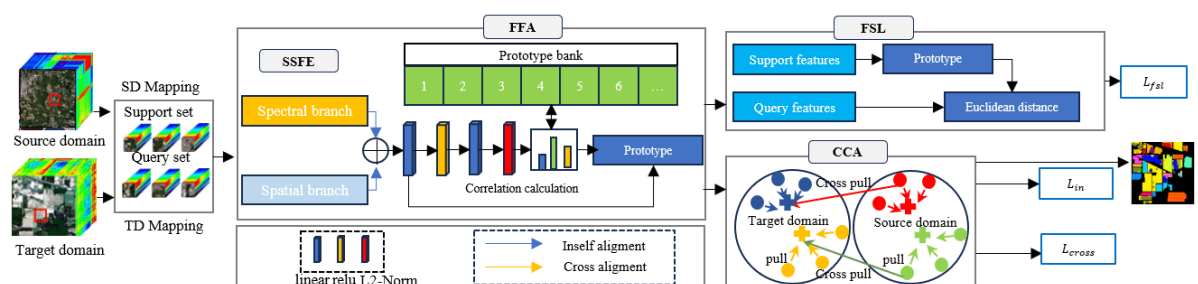


Fig2: Overall framework of FGAPA

Support features and Query features are the feature representations extracted from the support set and the query set, respectively. As shown in Figure 2, the proposed method determines the category of Query features by computing the Euclidean distance between them and the prototypes formed by the Support features of each class.

The FSL task is conducted simultaneously on both the source and target domains. Its loss function  $L_{fsl}$  consists of two components,  $L_{fsl}^t$  and  $L_{fsl}^s$ , which are collectively referred to as  $\mathcal{L}_{fsl}$  in the text. In the original Figure 2, the term "attention score" corresponds to the "correlation calculation" described in the paper, and  $p_i$  denotes the prototype of the  $i$ -th class in the prototype bank. Figure 2 has been updated accordingly to provide a clearer illustration of the overall workflow of FGAPA.

3. Equation (5) and (6): what is the dimensionality of the weights and biases in the MLP?

In FGAPA, the feature extractor outputs features of dimension  $128 \times 1$ . As specified in Equations (5) and (6), the MLP layers are configured as follows: the first layer has weights of size (128, 64) and biases of size 64, and the second layer has weights of size (64, 128) and biases of size 128.

4. Why do you need to transform the features and prototypes with MLP?

Since features and prototypes originate from different semantic levels, the MLP layer maps them into a comparable feature space. Meanwhile, the ReLU activation function within MLP further helps the model to learn more complex nonlinear relationships, thereby enhancing its expressive capability. Therefore, transforming both features and prototypes through the MLP allows for more effective correlation computation in the feature space and facilitates the learning of richer, more expressive representations.

5. Why is it a cosine similarity (normalized dot product) in equation (7) and not in equation (10)?

Both Equation (7) and Equation (10) employ cosine similarity for correlation computation. This approach is naturally aligned with prototype learning, better suited for high-dimensional embedding spaces, and maintains optimization stability. Moreover, the temperature coefficient adjusts the sharpness of the softmax distribution, further enhancing training stability. To clarify the methodological details, Equation (10) will be revised accordingly to avoid potential ambiguity.

6. What is the total optimization objective of the model in the end? It never appears.

The overall loss function of our method consists of three components: the few-shot learning loss  $L_{fsl}$ , the intra-domain alignment loss  $L_{in}$ , and the cross-domain alignment loss  $L_{cross}$ . It is defined as follows:

$$\text{Loss} = L_{fsl} + \lambda_1 L_{in} + \lambda_2 L_{cross}$$

Here,  $\lambda_1$  and  $\lambda_2$  are weighting hyperparameters for  $L_{in}$  and  $L_{cross}$ , respectively. Their values are tuned within the range [0, 1] with a step size of 0.1. After optimization, the best-performing values are set as  $\lambda_1 = 0.5$  and  $\lambda_2 = 0.3$ .

7. You talk about 4 datasets. Where do they come from? Is it public data? Is it published? How do you remove noisy bands?

This study conducts experiments on four publicly available hyperspectral datasets. Specifically, Chikusei is used as the source domain, while Indian Pines, Salinas, and Botswana serve as the target domains. The detailed information, data sources, and noise removal procedures for each dataset are as follows:

- Chikusei

The Chikusei dataset was captured by the Headwall Hyperspec-VNIR-C sensor over Chikusei City, Japan. It is publicly available at <https://naotoyokoya.com/Download.html> and consists of  $2517 \times 2335$  pixels with a spatial resolution of 2.5 m. The data provides 128 spectral bands covering wavelengths from 343 to 1018 nm and includes 19 land-cover classes. Figure 3 shows the pseudo-color image, the ground truth, and the corresponding class labels along with the number of samples for each class.

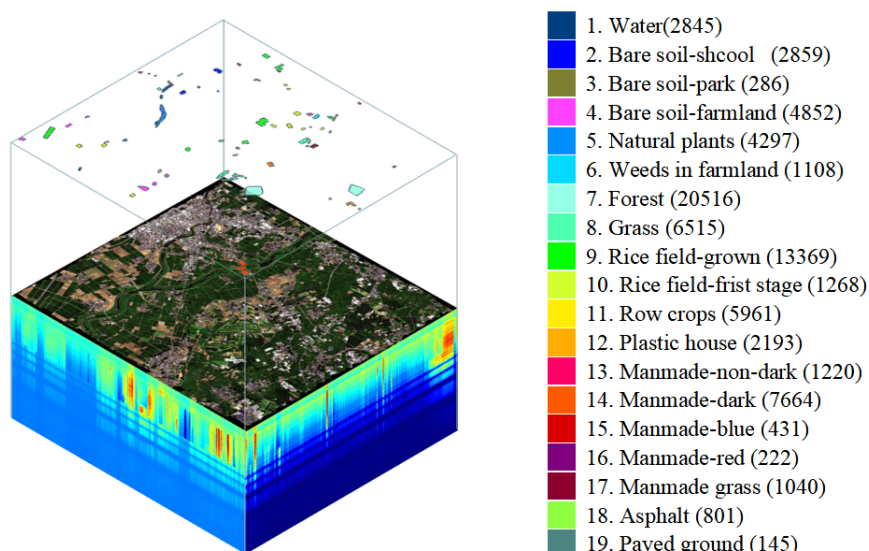


Figure 3: The land cover types and the number of samples for the Chikusei dataset

- Indian Pines

The Indian Pines dataset was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Indiana, USA. It is publicly accessible at [https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scenes](https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes) and comprises  $145 \times 145$  pixels with a spatial resolution of approximately 20 m. After removing 20 water absorption bands (104–105, 150–163, and 220), 200 spectral bands covering the wavelength range from 400 to 2500 nm are used. Figure 4 shows the pseudo-color image, the ground truth, and the corresponding class labels along with the number of samples for each class.

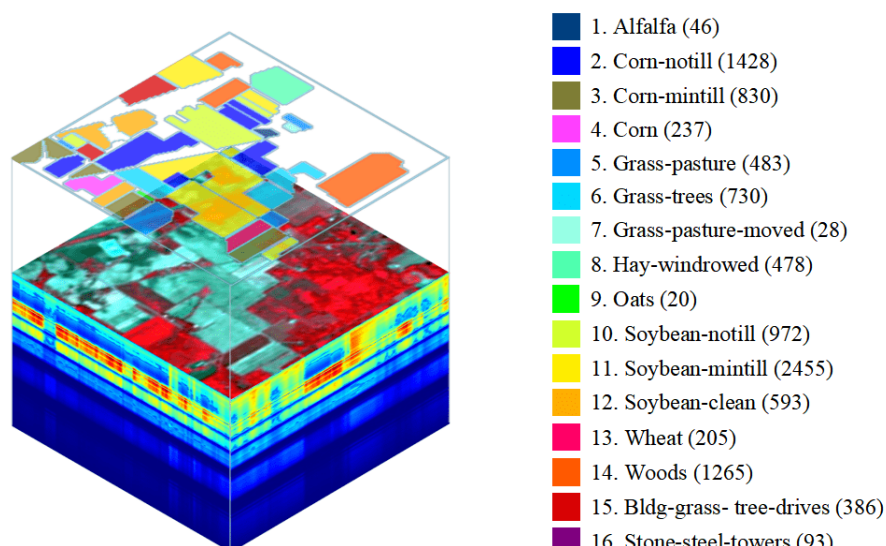


Fig4: The land cover types and the number of samples for the Indian Pines dataset

- Salinas

The Salinas dataset was collected by the AVIRIS sensor over the Salinas Valley in California, USA. It is publicly available at [https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scenes](https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes) and consists of  $512 \times 217$  pixels with a spatial resolution of approximately 3.7 m. A total of 204 spectral bands covering the range of 400–2500 nm are used, divided into 16 land-cover classes. Figure 5 shows the pseudo-color image, the ground truth, and the corresponding class labels along with the number of samples for each class.

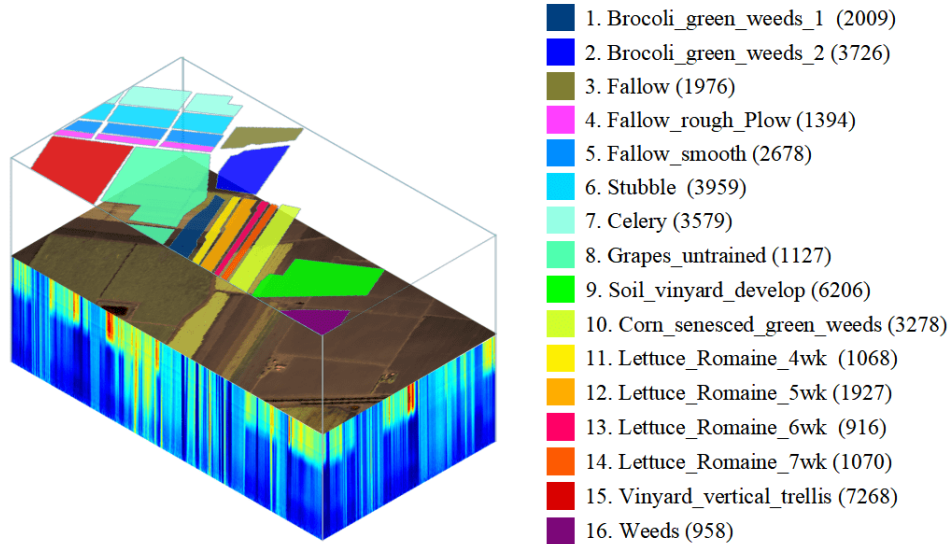


Fig5: The land cover types and the number of samples for the Salinas dataset

- Botswana

The Botswana dataset was acquired by NASA's EO-1 satellite over the Okavango Delta in Botswana. It is publicly available at [https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scenes](https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes). The image size is  $1476 \times 256$  pixels, with a spatial resolution of approximately 20 m. Out of the original 242 spectral bands (400–2500 nm), 145 bands are used after removing noisy ones (specifically bands 1–9, 56–81, 98–101, 120–133, and 165–186). Figure 6 shows the pseudo-color image, the ground truth, and the corresponding class labels along with the number of samples for each class.

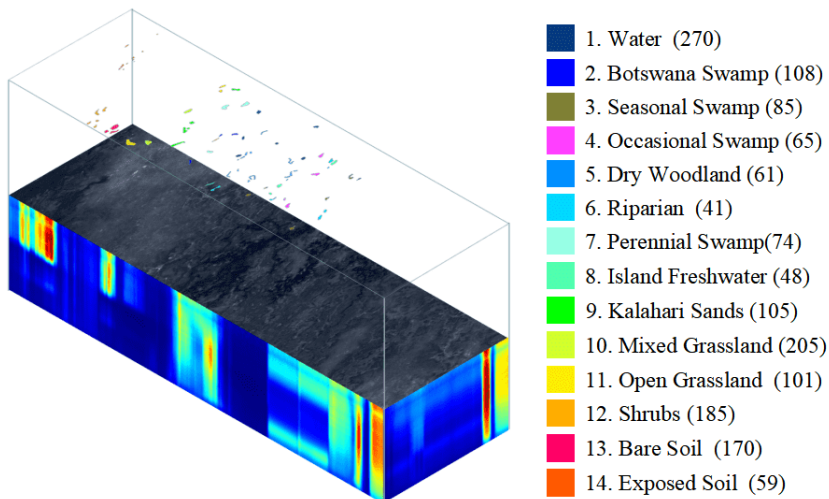


Fig6: The land cover types and the number of samples for the Botswana dataset

8. In the experimental protocol, you have to detail the metrics more in detail. What do they measure? How are they calculated? Why are they relevant for the task?

**Three commonly used metrics for evaluating classification performance are employed in this study: Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient. OA measures the overall classification correctness, AA calculates the average accuracy across all classes to avoid dominance by classes with a large number of samples, and Kappa provides a more reliable assessment of agreement between the classification results and ground truth by correcting for random consistency. The definitions of these metrics are as follows:**

- **OA:**

$$OA = \frac{\sum_{i=1}^c TP_i}{N}$$

where  $TP_i$  denotes the number of correctly classified samples for the  $i$ -th class,  $N$  is the total number of samples, and  $c$  represents the total number of classes.

- **AA:**

$$AA = \frac{1}{c} \sum_{i=1}^c \frac{TP_i}{N_i}$$

where  $TP_i$  denotes the number of correctly classified samples for the  $i$ -th class,  $N_i$  represents the total number of samples in the  $i$ -th class, and  $c$  is the total number of classes.

- **Kappa:**

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where  $P_o$  represents the probability that the model and a random classifier assign the same class labels,  $P_e$  denotes the theoretical probability of agreement between the model and a random classifier, calculated based on their respective marginal distributions.

9. As stated in the global comment, the results are missing statistical significance tests to make sure your method is statistically better on these datasets.

**Due to space constraints, some important content in the main text could not be fully elaborated. To address this, we have supplemented the performance results of FGAPA on the three target domains—Indian Pines, Salinas, and Botswana—with Chikusei as the source domain, including standard deviations to demonstrate statistical significance. Tables 2-4 present the performance statistics of FGAPA under 1-5 shot settings on the three target domains, respectively. All experiments were independently run 10 times, and the results are reported as averages.**

Table 2: Classification results (mean  $\pm$  standard deviation) on Indian Pines using 1-5 labeled samples per class

Indian Pines	Method							
	DFSL+NN	DCFSL	HFSL	DM-MRN	FSCF-SSL	CTF-SSCL	MLPA	FGAPA
<b>1-shot</b>	40.93 $\pm$ 4.82	41.32 $\pm$ 5.75	42.95 $\pm$ 6.74	41.65 $\pm$ 5.36	46.44 $\pm$ 5.10	43.52 $\pm$ 4.84	41.83 $\pm$ 5.51	<b>52.16 <math>\pm</math> 4.48</b>
<b>2-shot</b>	50.98 $\pm$ 3.09	51.60 $\pm$ 4.53	53.78 $\pm$ 5.59	55.41 $\pm$ 3.88	60.14 $\pm$ 4.84	56.82 $\pm$ 4.92	53.47 $\pm$ 4.82	<b>64.81 <math>\pm</math> 4.32</b>
<b>3-shot</b>	56.47 $\pm$ 2.44	56.03 $\pm$ 3.45	63.65 $\pm$ 4.13	60.85 $\pm$ 3.28	67.72 $\pm$ 1.62	63.61 $\pm$ 3.5	57.65 $\pm$ 2.63	<b>71.19 <math>\pm</math> 4.18</b>
<b>4-shot</b>	60.44 $\pm$ 2.82	61.96 $\pm$ 3.98	69.96 $\pm$ 3.93	63.95 $\pm$ 3.07	72.92 $\pm$ 5.56	67.29 $\pm$ 3.04	63.31 $\pm$ 2.05	<b>76.12 <math>\pm</math> 2.76</b>
<b>5-shot</b>	63.16 $\pm$ 2.92	65.74 $\pm$ 2.57	74.03 $\pm$ 2.71	69.28 $\pm$ 3.30	76.96 $\pm$ 2.74	70.86 $\pm$ 3.22	66.59 $\pm$ 2.81	<b>79.65 <math>\pm</math> 2.76</b>

Table 3: Classification results (mean  $\pm$  standard deviation) on Salinas using 1–5 labeled samples per class

Salinas	Method							
	DFSL+NN	DCFSL	HFSL	DM-MRN	FSCF-SSL	CTF-SSCL	MLPA	FGAPA
<b>1-shot</b>	75.65 $\pm$ 2.46	74.81 $\pm$ 4.12	68.41 $\pm$ 4.76	76.77 $\pm$ 5.72	68.30 $\pm$ 4.83	74.19 $\pm$ 3.14	73.37 $\pm$ 3.08	<b>79.47 <math>\pm</math> 3.30</b>
<b>2-shot</b>	82.16 $\pm$ 2.64	82.04 $\pm$ 3.26	80.18 $\pm$ 2.76	84.89 $\pm$ 3.23	81.40 $\pm$ 2.99	81.79 $\pm$ 3.26	81.63 $\pm$ 2.86	<b>87.55 <math>\pm</math> 3.13</b>
<b>3-shot</b>	84.99 $\pm$ 1.53	84.88 $\pm$ 1.84	85.28 $\pm$ 2.65	87.48 $\pm$ 3.23	84.33 $\pm$ 3.52	85.57 $\pm$ 2.17	85.37 $\pm$ 2.01	<b>90.70 <math>\pm</math> 1.86</b>
<b>4-shot</b>	88.61 $\pm$ 1.53	88.05 $\pm$ 1.90	86.98 $\pm$ 2.57	90.04 $\pm$ 2.03	88.11 $\pm$ 3.66	88.58 $\pm$ 1.59	89.04 $\pm$ 1.50	<b>91.83 <math>\pm</math> 1.09</b>
<b>5-shot</b>	89.74 $\pm$ 1.26	90.12 $\pm$ 1.22	89.14 $\pm$ 2.09	91.08 $\pm$ 2.39	90.96 $\pm$ 2.06	89.96 $\pm$ 1.43	89.94 $\pm$ 1.30	<b>93.15 <math>\pm</math> 1.00</b>



Table 4: Classification results (mean  $\pm$  standard deviation) on Salinas using 1–5 labeled samples per class

Botswana	Method							
	DFSL+NN	DCFSL	HFSL	DM-MRN	FSCF-SSL	CTF-SSCL	MLPA	FGAPA
<b>1-shot</b>	86.88 $\pm$ 3.37	85.14 $\pm$ 2.96	68.85 $\pm$ 5.00	59.91 $\pm$ 4.66	68.76 $\pm$ 4.89	79.20 $\pm$ 3.64	82.72 $\pm$ 2.81	<b>90.62 <math>\pm</math> 1.71</b>
<b>2-shot</b>	91.64 $\pm$ 2.69	92.79 $\pm$ 1.92	82.84 $\pm$ 3.75	81.07 $\pm$ 3.70	81.55 $\pm$ 3.14	89.67 $\pm$ 3.57	92.63 $\pm$ 1.55	<b>94.95 <math>\pm</math> 2.15</b>
<b>3-shot</b>	94.44 $\pm$ 1.94	94.67 $\pm$ 1.36	88.83 $\pm$ 3.45	86.91 $\pm$ 1.93	90.19 $\pm$ 3.24	93.31 $\pm$ 1.65	94.42 $\pm$ 1.98	<b>96.90 <math>\pm</math> 1.31</b>
<b>4-shot</b>	95.54 $\pm$ 1.46	95.98 $\pm$ 1.23	94.19 $\pm$ 2.49	88.82 $\pm$ 2.46	93.32 $\pm$ 2.36	94.93 $\pm$ 0.85	95.30 $\pm$ 1.34	<b>97.16 <math>\pm</math> 1.35</b>
<b>5-shot</b>	96.48 $\pm$ 0.84	96.90 $\pm$ 1.03	94.56 $\pm$ 1.61	92.55 $\pm$ 1.79	95.81 $\pm$ 1.43	96.05 $\pm$ 1.32	96.65 $\pm$ 1.04	<b>98.22 <math>\pm</math> 1.31</b>

As shown in Tables 2-4, under the 1–5 shot settings, FGAPA consistently outperforms all comparative methods across the three datasets: Indian Pines, Salinas, and Botswana. Specifically, under the 5-shot condition, the collaborative work of its ACA and FFA modules extracts discriminative features while reducing inter-domain differences, leading to outstanding performance—surpassing the second-best method by 2.69%, 3.07%, and 1.32% on the three datasets, respectively. Under the 1-shot condition, the discriminative features captured by the FFA module better characterize class distributions and enhance generalization, achieving advantages of 5.72%, 2.7%, and 3.74%, respectively. Moreover, FGAPA generally exhibits lower standard deviations than other methods across all experimental scenarios, indicating strong stability under varying sample conditions. Overall, these results fully demonstrate the statistically significant superiority of FGAPA.

10. There are temperature parameters in equations (7) and (10). How do you tune these? This information is lacking. Same for hyperparameters in the losses: how are the weights selected? On which data? What impact on the performance of the model? What is the batch size used? (This is necessary for reproducibility.)

**In this experiment, the hyperparameters include the temperature parameter  $\tau$  in Equations (7) and (10), the weighting coefficients  $\lambda_1$  and  $\lambda_2$  in the total loss, and the batch size used during model training. The specific settings for these parameters are described below.**

- For the temperature parameter  $\tau$ , we set it to a universal value of 0.1 based on experience.
- In the final loss, two hyperparameters controlling the dual-domain adversarial strategy are included, as specifically shown below:

$$Loss = L_{fsl} + \lambda_1 L_{in} + \lambda_2 L_{cross}$$



where  $\lambda_1$  and  $\lambda_2$  represent the weighted hyperparameters for  $L_{in}$  and  $L_{cross}$ . Their values range from 0 to 1 with increments of 0.1. They were tuned using the OA, AA, and Kappa metrics on three datasets, and the optimal parameters obtained were  $\lambda_1 = 0.5$  and  $\lambda_2 = 0.3$ .

- Furthermore, we set the batch size in FGAPA to 64, which determines the number of samples used in each parameter update during training. Throughout the training process, the data are divided into multiple batches, each containing 64 samples. This configuration not only improves training efficiency but also helps suppress noise and enhances the stability of model training.

11. The last paragraph of the introduction does not contain the appropriate acronym (FFE instead of FFA).

Figure 3 is too small.

Following the suggestions, we have corrected all instances of FFE to FFA in the text. Furthermore, Figure 3 has been enlarged to enable readers to more clearly observe the performance differences among the methods.

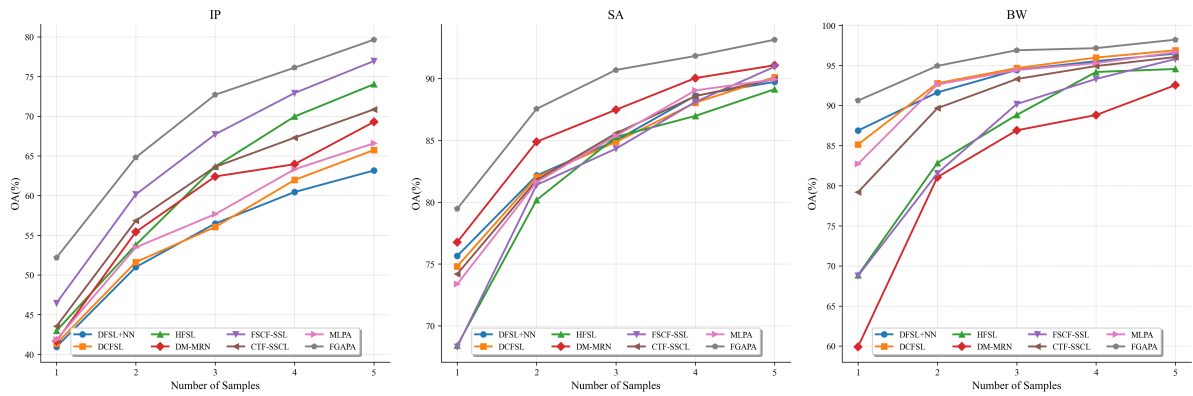


Figure 7: Revised version of figure 3 in FGAPA

12. I am not an expert in hyperspectral image processing, and it is difficult to understand the task from the text. It would be highly beneficial to add a paragraph that clearly presents the task. For example, what does Figure 2 represent? The bibliography is also somewhat limited.

We sincerely thank the reviewer for the valuable suggestions. In the revised manuscript, we will provide clearer explanations of the notations and core concepts to improve readability. Additionally, we will include relevant references to better reflect the current research landscape.

## The ending

We sincerely appreciate the reviewers' positive assessment of the novelty of this work and their recognition of the potential and practical value of the proposed cross-domain hyperspectral image classification method within the overall research framework and methodological motivation. We fully acknowledge the main concerns raised by the reviewers and have provided comprehensive responses regarding the robustness of FGAPA to new source and target domains, the readability of its fundamental concepts, the setting of experimental hyperparameters, methodological details, and the statistical significance of the experimental results. We hope that this supplementary material addresses the reviewers' concerns. Once again, we thank the reviewers for their valuable suggestions, which have helped us identify key areas for improvement in the manuscript. We believe that incorporating this feedback will significantly enhance the quality and clarity of the paper.

