Attentive Cross-Domain Few-shot Learning and Domain Adaptation in HSI Classification

1st Rojan Basnet
GSAIS
Kyoto University
Kyoto, Japan
basnet.rojan.55i@st.kyoto-u.ac.jp

2nd Rimsa Goperma
GSAIS
Kyoto University
Kyoto, Japan
rimsa.goperma.53c@st.kyoto-u.ac.jp

3rd Liang Zhao
GSAIS
Kyoto University
Kyoto, Japan
liangzhao@acm.org

***Abstract*—This study explores the application of Attentive Cross-Domain Few-Shot Learning (ACDFSL) in Hyperspectral Image (HSI) Classification, specifically addressing challenges associated with environments possessing limited labeled data. Our approach applies the Squeeze-and-Excitation (SE) attention and Residual elements within a deep learning architecture of four convolution blocks. This innovative strategy of integrating attention mechanisms into few-shot learning models represents a significant departure from traditional practices. After rigorous assessment, the ACDFSL model showcased outstanding results, revealing performance rates of 92.14%, 96.23%, and 91.27% in OA, AA, and Kappa, respectively, on the Salinas dataset. Additionally, the model attained rates of 85.67%, 89.66%, and 85.4% on the University of Pavia (PU) dataset. These results indicate an edge over existing state-of-the-art techniques such as SVM, 3D-CNN, SSRN, and other DFSL variants. This considerable progress emphasizes the potential and applicability of the ACDFSL approach in real-world HSI Classification scenarios, especially where labeled data is sparse, and paves the way for future research in this sphere.**

***Index Terms*—ACDFSL, HSI Classification, SE Attention, FSL, Domain Adaptation**

I. Introduction

Hyperspectral Imaging (HSI), well-known for its precision in capturing data across numerous spectral bands, has found extensive applications in a variety of fields, including but not limited to agriculture, environmental monitoring, and geology [12]. In the face of mounting demand for precise HSI in areas such as environmental conservation and crop yield prediction, the need to bolster HSI classification accuracy is more pressing than ever.

Over time, a multitude of machine learning and deep learning techniques have been employed to improve HSI classification, each presenting its unique challenges. Initial traditional methods paved the way for more advanced techniques, such as the Stacked Autoencoder (SAE) [1], Deep Belief Network (DBN) [8], Convolutional Neural Network (CNN) [2], and 3-D CNNs [19]. While these advancements brought considerable progress, they also introduced issues like heightened computational load, gradient disappearance, and increased demand for extensive labeled data. To counter these problems, ResNet [5] employed Residual Blocks, although this approach necessitated vast amounts of labeled data for optimal model performance.

Data augmentation techniques were subsequently developed to enhance training sample sizes without the need for additionally labeled samples [22]. However, these methods often fell short of boosting the feature representation capabilities of deep models. This drawback led to the development of attention mechanisms [15], [16], [18] aimed at enhancing relevant features while downplaying irrelevant ones. Yet, the requirement for domain adaptation [14] surfaced, aiming to reconcile the disparities between different data distributions in cross-domain learning.

Various strategies like fine-tuning [20] were devised to confront this issue, but they stumbled when encountering substantial data shifts between domains. Semi-supervised and active learning [4] methods were then proposed to decrease the dependence on manual labeling, but these too faced challenges in generalizing for new tasks. This led to the emergence of meta-learning, with FSL [3], [9], [11] surfacing as a promising concept. The technology is indispensable; however, the task of accurately classifying HSI data, often described as the 'small sample size problem, remains a significant challenge. Few-Shot Learning (FSL) presents a promising solution to this conundrum, capitalizing on a limited number of labeled samples [9]. However, traditional FSL methods often presume that source and target class data share identical distributions, which may not always be the case.

This understanding prompted the development of more sophisticated FSL techniques, including FSL with fine-tuning, RN-FSC, DFSL, and DCFSL [3], [7], [9]. These innovative techniques leverage past learning experiences for new tasks and address the lack of genuine HSI samples. Among these, the DCFSL [7] method has emerged as a key tool for HSI classification, skillfully tackling the dual challenges of drastic domain shifts and the scarcity of labeled data. Despite its potential, opportunities for further enhancement and improvement in performance and accuracy persist, making the future of HSI classification an exciting endeavor.

In this paper, we introduce an improved approach to FSL for HSI classification, utilizing a deep residual attention-based method that aims to optimize existing methodologies. Our paper strives to bridge this gap by integrating the Squeeze-and-Excitation (SE) attention mechanism into the DCFSL framework. Our research objectives are two-fold: Firstly, to set a new benchmark for HSI classification accuracy by incorporating the SE attention mechanism into the DCFSL framework; and secondly, to improve the handling of domain shifts and reconcile discrepancies between source and target domains in few-shot learning scenarios. The contributions of this paper are three-fold. Firstly, we introduce a

novel SE attention-based framework for domain adaptation in few-shot learning scenarios. Secondly, we demonstrate superior performance in HSI classification, surpassing multiple state-of-the-art algorithms. Finally, we provide a robust method to extract rich domain-invariant representations, even under data scarcity conditions.

The rest of this paper is organized as follows: Section II details the methodology, which includes the integration of the SE attention mechanism into the DCFSL framework. Section III presents our experimental results and compares them with prior studies. Section IV discusses these results and their implications for HSI classification and meta-learning. Finally, Section V concludes the paper.

II. Methodology

A. Attentive-Cross-Domain Few-shot Learning (ACDFSL) Framework

The ACDFSL framework addresses the challenging problem of few-shot learning in scenarios where labeled data is scarce and originates from diverse domains. Unlike in prior studies [7] where the feature embedded extractor struggled to encapsulate inter-dependencies between channels, consequently failing to highlight salient features, our methodology makes a significant breakthrough in this domain. We propose a Squeeze and Excitation (SE) attention mechanism in the residual network with increased depth, a novelty depicted inside the red dotted box in Figure 1. The proposed framework has four major steps mapping, feature extraction, conditional adversarial learning and testing:

Mapping layers: The mapping layers M_s and M_t serve to normalize input dimensions across domains. The primary function of these mapping layers is to process and transform input data to feed in feature extractor network. The equation for the mapping layers is

$$M_s, M_t = f_{ms}(X_s), f_{mt}(X_t). \tag{1}$$

Deep residual Attentive 3-D CNN: This network consists of four residual blocks equipped with squeeze and excitation based attention mechanism enabling rich spatial-spectral embedded features extraction [15]. These extracted domain invariant features serve both source and target Few-Shot Learning (FSL) tasks. The equation for the deep residual 3-D CNN is

$$F_s, F_t = f_{cnn}(M_s), f_{cnn}(M_t). \tag{2}$$

Furthermore, the Euclidean distances between labeled and unlabeled samples are calculated to promote intra-class compactness and inter-class separability, which are crucial for effective classification. The equation for the Euclidean distances squared is

$$D_{ij} = \|F_i - F_j\|^2. \tag{3}$$

Conditional domain discriminator: A conditional domain discriminator is introduced to instill and learn domain invariant latent space extracted by previous network, mitigating the impact of the domain shift. The discriminator, which uses the predicted class labels as conditions, is trained to distinguish between the source and target domains. The equation for the conditional domain discriminator is

$$D_s, D_t = f_{dis}(F_s, Y_s), f_{dis}(F_t, Y_t). \tag{4}$$

The discriminator loss, denoted as L_d, is calculated to measure the alignment of the global data distributions. The equation for the discriminator loss is

$$L_d = \frac{1}{2}\left(\frac{1}{n_s}\sum_{i=1}^{n_s} D_s(i) + \frac{1}{n_t}\sum_{j=1}^{n_t} D_t(j)\right). \tag{5}$$

In addition, the overall loss function for the ACDFSL framework combines the few-shot losses for the source ($L_{s_{fsl}}$) and target ($L_{t_{fsl}}$) domains with the discriminator loss (L_d). The network utilizes the total loss to train in an adversarial way improving both the discriminator network and feature extractor network simultaneously [10]. Therefore, the total loss becomes

$$L_{\text{total}} = L_{s_{fsl}} + L_{t_{fsl}} + L_d. \tag{6}$$

Testing Phase: In this phase, the target mapping layer reduces the input dimensions. The equation for the target mapping layer is

$$M_t = f_{mt}(X_t). \tag{7}$$

After that, the attention based deep residual 3-D CNN trained before extracts distinct domain invariant representations of the target data. The equation for the distinct features is

$$F_t = f_{cnn}(M_t). \tag{8}$$

Then, a nearest neighbor classifier is employed to perform accurate classification. The equation for the nearest neighbor classifier is

$$y = \arg\min_{y_i \in Y_s} \|F_t - F_i\|^2. \tag{9}$$

In summary, the ACDFSL framework effectively leverages conditional adversarial learning to transfer knowledge from the source domain to the target domain, thereby significantly improving the classification performance and reducing the domain shift in the target domain, despite the scarcity of labeled data.

B. Proposed Feature Embedded Network

We propose a 3D Residual Attention Network that combines two fundamental structures: the Squeeze-and-Excitation (SE) block [6] and the Residual Block [13], [15].

1) Squeeze-and-Excitation (SE) Block: This block enables an attention mechanism that captures inter-dependencies between the channels of the convolutional layers accurately. The SE block recalibrates the channel-wise feature responses adaptively, enabling dynamic learning and adjustment of inter-channel relationships according to the input [6]. This helps the network to amplify the relevance of critical features while suppressing less useful ones. The mathematical equations for the SE block are as follows

An input tensor $X \in \mathbb{R}^{C\times D\times H\times W}$ is processed through a 'squeeze' operation in the SE block. Using Global Average Pooling (GAP), it aggregates global spatial information to reduce the input dimensions, generating an output $X_{\text{avg}} \in \mathbb{R}^C$ where

$$X_{\text{avg}} = \frac{1}{D\times H\times W}\sum_{i=1}^{D}\sum_{j=1}^{H}\sum_{k=1}^{W} X_{i,j,k}. \tag{10}$$

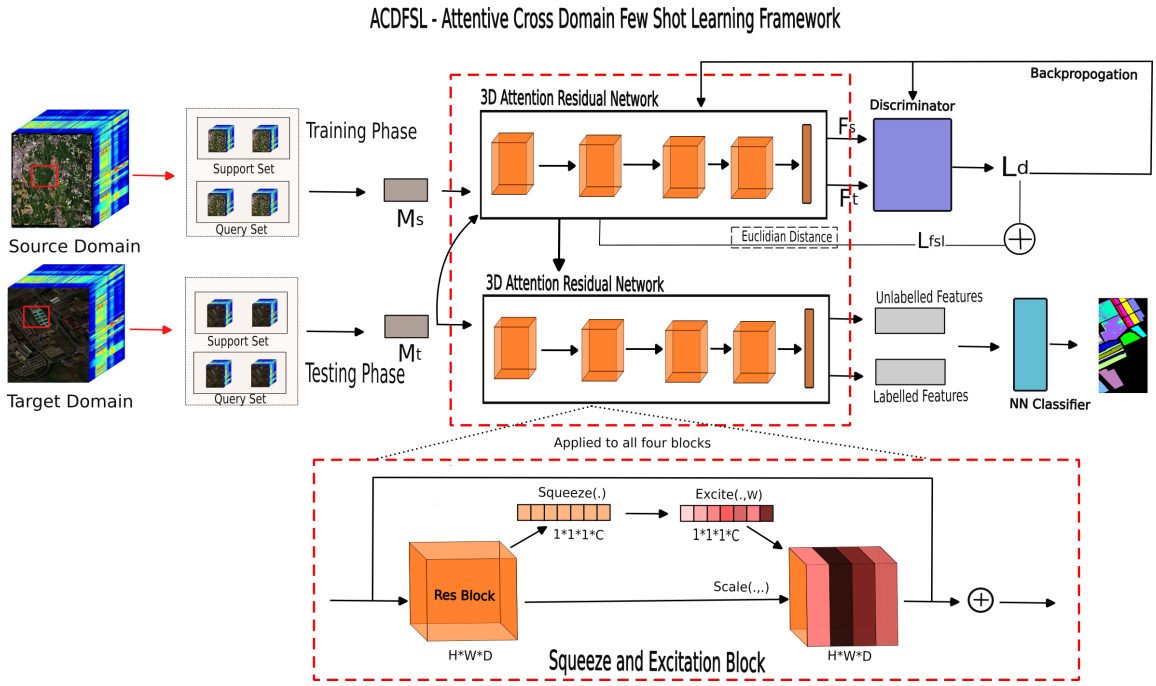


Fig. 1: Illustration of ACDFSL

Following the 'squeeze' operation, an 'excitation' operation is executed via two fully connected (FC) layers. This operation captures channel-wise dependencies effectively and serves as a gating mechanism, hence enhancing the discernibility of channel-wise features. The output $S \in \mathbb{R}^C$ is computed as

$$S = \sigma(W_2\delta(W_1 X_{\text{avg}})). \tag{11}$$

where σ and δ represent the sigmoid function and the ReLU function, respectively. $W_1 \in \mathbb{R}^{C/r \times C}$ and $W_2 \in \mathbb{R}^{C \times C/r}$ are the parameters of the two FC layers.

Finally, the SE block concludes by re-scaling the input tensor with the output of the 'excitation' operation to produce the SE output $Y \in \mathbb{R}^{C \times D \times H \times W}$ as

$$Y = S \cdot X. \tag{12}$$

2) Residual Block: The Residual Block is incorporated with the SE block to enhance the learning abilities of our network. It achieves this by introducing shortcuts or skip connections that enable the training process to bypass one or more layers, resulting in a residual learning framework. This feature eases the training of networks and aids in addressing the problem of vanishing gradient [13], [15]. The mathematical equations for the Residual Block are as follows

An input tensor $X \in \mathbb{R}^{C \times D \times H \times W}$ is passed through a series of transformations F with parameters $\{W_i\}$ within the residual block, and the output $Y \in \mathbb{R}^{C \times D \times H \times W}$ is computed as

$$Y = F(X, \{W_i\}) + W_s \cdot X. \tag{13}$$

Here, F denotes the series of stacked non-linear transformations that include convolutions and the SE attention operation, whereas W_s denotes the transformation in the shortcut connection, utilized when there is a dimension mismatch between the input and output.

Subsequently, the output tensor is passed through a ReLU activation function, imparting non-linearity to the model

$$Y = \max(0, F(X, \{W_i\}) + W_s \cdot X). \tag{14}$$

The integration of the SE block and the Residual Block facilitates efficient extraction and manipulation of meaningful features, boosting network training efficiency through the proposed bottom-up top-down attention structure and the residual connection. This enhances performance in HSI classification tasks by enabling dynamic inter-channel relationship adjustments, fostering efficient feature extraction and manipulation, and improving overall network training efficiency. We have four blocks in our network as the increased depth is closely associated with it's ability to learn complex hierarchical features. While this must be carefully managed to avoid issues such as vanishing gradients, residual blocks mitigates this concern by allowing direct paths from the earlier layers to the later layers.

III. Result

The experimental procedures were executed on an NVIDIA GeForce RTX 3090 GPU, with a memory capacity of 24GB. The programming was implemented using Python 3.6, backed by the PyTorch framework.

A. Datasets:

The research conducted leveraged three distinct hyperspectral imaging datasets, specifically those pertaining to Chikusei, the University of Pavia (UP), and Salinas, the details of which have been collated in Table I. These datasets encompass a range of classifications and present disparate

distribution alignments due to their individual collection contexts in terms of time, location, and environmental conditions. Given this inherent diversity, they provide an ideal composite for assessing the versatility of domain adaptability. For the purpose of this study, the Chikusei dataset was designated as the source, whilst the University of Pavia and Salinas datasets were utilized as target domains.

TABLE I: Summary of HSI Datasets

Dataset	Location	Resolution	Spectral Band Range (Number of Bands)	Number of Classes
Chikusei [21]	Japan	2.5m, 2517x2335	363-1018nm (128)	19
UP	Italy	1.3m, 610x340	430-860nm (103)	9
Salinas	USA	3.7m, 512x217	400-2500nm (204)	16

The Table II provides an overview of the Chikusei dataset.

TABLE II: Overview of Chikusei Dataset

No.	Class	Samples
1	Water	2845
2	Bare soil (school)	2859
3	Bare soil (park)	286
4	Bare soil (farmland)	4852
5	Natural plants	4297
6	Weeds in farmland	1108
7	Forest	20516
8	Grass	6515
9	Rice field (grown)	13369
10	Rice field (first stage)	1268
11	Row crops	5961
12	Plastic house	2193
13	Manmade (non-dark)	1220
14	Manmade (dark)	7664
15	Manmade (blue)	431
16	Manmade (red)	222
17	Manmade grass	1040
18	Asphalt	801
19	Paved ground	145

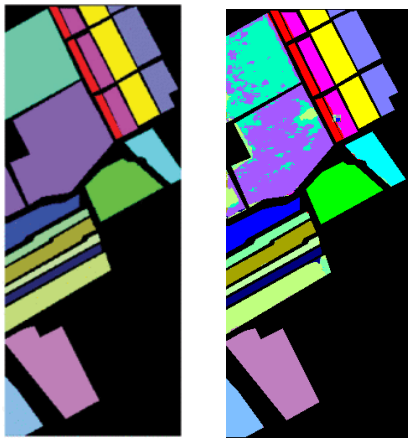
B. Experimental Design and Comparative Study

The Table III provides the parameter settings for our experimentation.

TABLE III: Training and Evaluation Details

Parameter	Value
Training Method	ACDFSL
Training Samples/Class	5
Optimizer	Adam
Iterations	20,000
Learning Rate	0.001
Spatial Size	9×9
Evaluation Metrics	Overall Accuracy (OA), AA, Kappa

1) Performance Analysis: Salinas Dataset: The Salinas dataset is analyzed in terms of algorithmic performances and detailed metrics, as presented in Table IV [11], [17]. This table provides a comprehensive comparison of various algorithms on the Salinas dataset, showcasing their classification results for different classes. In accordance to the result, the SVM performs well on classes 1 and 16 but struggles with complex classes like class 8. The 3D-CNN shows impressive results in some instances, but falls short for others. The ACDFSL algorithm, however, shows consistently superior performance across most classes, with a slight dip in class 15. Figure 2 visually represents the ground truth and predicted images for the Salinas Data. Furthermore, the



(a) Ground Truth (b) Predicted image

Fig. 2: Salinas Data

ACDFSL algorithm achieves the highest overall accuracy (OA) at 92.14%, closely followed by the DT-FSL at 91.03%. The ACDFSL also excels in average accuracy (AA) with a score of 96.23%, and in the kappa statistic, which measures classification agreement, scoring 91.27

This evidence shows that the ACDFSL algorithm outperforms the others in terms of classification accuracy on the Salinas dataset.

2) Performance Analysis: UP Dataset: Similarly, we conducted a performance analysis on the UP dataset. Figure 3 shows the ground truth and predicted images for the UP Data. The classification results, reported in percentages, are presented in Table V [11], [17] for various approaches in HSI classification.

The algorithms' performance on the UP dataset shows some variation across different classes. For instance, the SVM and 3D-CNN algorithms show a high degree of variability, excelling in some classes but underperforming in others. However, the ACDFSL algorithm consistently delivers high performance across most classes. Figure 3 visually represents the ground truth and predicted images for the Salinas Data. The ACDFSL also outshines other algorithms in terms of overall accuracy (OA), average accuracy (AA), and Kappa statistics, scoring 92.14%, 96.23%, and 91.27% respectively.

IV. Discussion

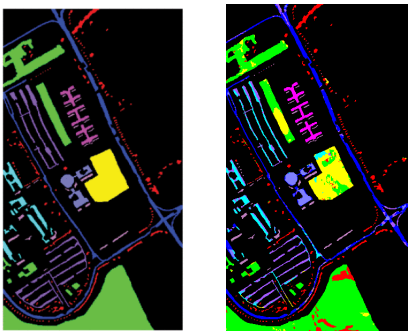
This study implemented the Attentive Cross-Domain Few-Shot Learning (ACDFSL) method for hyperspectral image (HSI) classification across three datasets: Chikusei, University of Pavia (UP), and Salinas. Evidenced by evaluation metrics including Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient, ACDFSL consistently surpassed traditional machine learning algorithms such as SVM,

TABLE IV: Classification results (%) on the Salinas dataset with various algorithms

Class	SVM	3D-CNN	SSRN	DFSL+NN	DFSL+SVM	RN-FSC	DCFSL	DT-FSL	ACDFSL
Brocoli_green_weeds_1	93.33	87.01	78.68	97.31	93.46	76.58	99.43	98.32	**100**
Brocoli_green_weeds_2	74.59	92.40	99.89	99.06	94.16	79.19	99.51	99.52	**100**
Fallow	66.46	72.59	69.19	87.97	87.74	88.86	91.75	91.06	**97.31**
Fallow_rough_plow	72.31	96.87	97.90	98.98	86.99	98.23	99.09	98.10	**99.71**
Fallow_smooth	91.23	92.25	93.57	89.04	81.77	88.15	92.65	94.38	**94.5**
Stubble	60.48	99.30	98.94	97.82	88.07	99.36	98.46	97.54	**100**
Celery	81.59	97.15	95.56	99.04	89.45	99.28	**99.66**	99.49	98.94
Grapes_untrained	33.62	52.38	69.43	66.79	71.56	73.36	72.83	79.92	**81.15**
Soil_vinyard_develop	91.13	95.79	90.52	94.76	90.58	93.74	99.28	99.59	**99.9**
Corn_senesced_green_weeds	60.31	73.16	81.44	74.89	88.55	62.17	85.87	85.47	**94.65**
Lettuce_romaine_4wk	80.36	71.94	91.79	85.36	88.90	91.65	99.12	97.42	**99.53**
Lettuce_romaine_5wk	91.11	83.48	96.15	98.49	98.65	95.44	99.74	99.22	**99.74**
Lettuce_romaine_6wk	85.18	81.51	95.31	98.90	95.65	98.84	99.23	99.28	**99.34**
Lettuce_romaine_7wk	73.05	93.14	97.46	98.08	94.96	97.78	**99.44**	99.64	98.97
Vinyard_untrained	53.14	44.38	71.92	75.36	65.04	68.68	77.32	**81.65**	77.27
Vinyard_vertical_trellis	**99.51**	83.97	91.09	83.14	86.05	81.17	91.84	89.90	98.67
OA	66.47	75.61	84.25	85.29	82.80	82.78	89.15	91.03	**92.14**
AA	75.46	82.33	88.68	90.31	87.60	87.03	94.08	94.41	**96.23**
Kappa	63.54	73.14	82.53	83.69	80.95	80.88	87.96	90.03	**91.27**

TABLE V: Classification results (%) on the UP dataset with various algorithms

Class	SVM	3D-CNN	SSRN	DFSL+NN	DFSL+SVM	RN-FSC	DCFSL	DT-FSL	ACDFSL
Asphalt	91.85	76.29	75.87	69.19	73.43	73.98	82.20	**92.68**	86.34
Meadows	84.01	71.53	66.95	84.63	89.25	88.80	87.74	84.55	**90.43**
Gravel	29.72	58.41	61.30	57.47	48.09	52.07	67.46	71.60	**74.79**
Trees	51.51	56.73	80.65	89.99	84.72	90.64	93.16	91.28	**97.97**
Painted metal sheets	93.90	99.48	99.24	**100.00**	99.65	98.94	99.49	99.58	98.81
Bare Soil	36.54	38.32	61.57	71.90	66.49	78.65	75.51	**80.44**	78.83
Bitumen	58.49	89.71	86.62	97.08	92.61	93.79	98.42	96.25	**99.11**
Self-Blocking Bricks	43.06	32.63	59.13	65.97	61.94	70.73	79.05	79.86	**80.73**
Shadows	81.55	94.60	92.17	96.35	96.38	98.61	99.78	100	**100**
OA	66.88	68.61	73.22	77.18	76.19	78.47	82.33	85.47	**87.67**
AA	63.40	68.42	75.78	81.40	79.17	83.14	86.48	88.48	**89.66**
Kappa	61.81	63.66	69.42	74.03	72.76	75.23	79.51	83.11	**85.49**



(a) Ground Truth (b) Predicted image

Fig. 3: UP Data

3D-CNN, SSRN, and other DFSL variants. This outcome demonstrates ACDFSL's effective application to HSIs of various complexities.

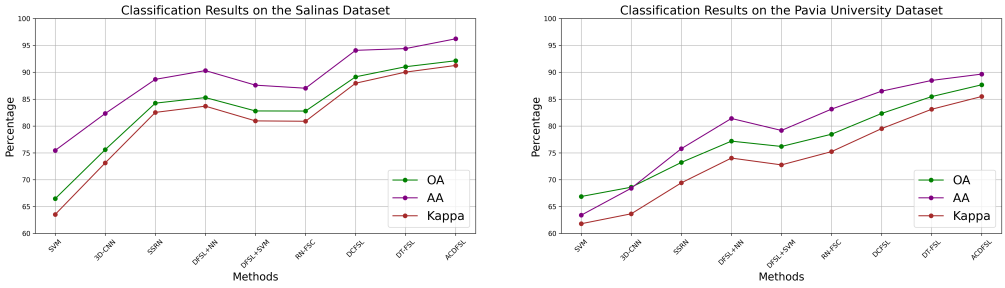
Our ACDFSL method employs an attention mechanism that allows the model to concentrate on task-relevant features. This leads to enhanced accuracy, especially when handling high dimensional data, where the relevance of features can significantly vary. This approach stands in contrast to how a standard CNN operates, where each layer indiscriminately learns to react to specific features present in the input data, without considering their relevance to the specific task at hand.

In terms of performance, ACDFSL exhibited outstanding results for specific classes in the Salinas and UP datasets as seen in Figure 4. For instance, in the Salinas dataset, ACDFSL achieved perfect classification accuracy for classes 1, 2, and 6, and near-perfect results for class 9, 11, 12, and 13 far exceeding the inconsistent outcomes of SVM and 3D-CNN. Similarly, in the UP dataset, ACDFSL maintained high performance across most classes, particularly excelling in classes 2, 3, 4, 7, 8 and 9. These findings suggest that ACDFSL effectively captures the spectral and spatial features of HSI, thereby ensuring superior classification accuracy.

Furthermore, our study highlighted ACDFSL's proficiency in classifying specific classes like Shadows, Brocolli, Lettuce and Meadows. This suggests that these classes might possess unique spectral features or more uniformity, facilitating simpler classification. These findings reflect the potency of ACDFSL's feature learning approach and attention mechanism in tackling complex classification tasks, even in scenarios where classes have closely related or overlapping spectral characteristics.

Despite the significant findings, this study has some limitations. The focus primarily on Salinas and UP datasets narrows its applicability to other potential target areas. A comprehensive analysis considering computational cost and model complexity—essential aspects for real-world ACDFSL application—was not conducted. Additionally, our method excels at detecting edges but struggles to perform well in larger areas. This could be attributed to factors such as limited receptive field of the network, SE attention bias, and data distribution.

Given these findings and identified limitations, we propose several directions for future research. First, the inclusion of more diverse HSI datasets would validate the broad applicability of ACDFSL. Techniques such as dilated convolutions



(a) Comparative results in Salinas Data

(b) Comparative results in UP Data

Fig. 4: OA(%), AA(%) and Kappa of nine HSI classification methods on two target datasets. (a) Salinas (b) UP.

and spatial-aware attention could be explored for making the model gather more contextual information of larger areas. Also, the exploration of enhancements to the ACDFSL model, through robust data augmentation techniques, source dataset variations, or a dynamic discriminator network, is worth investigating.

V. Conclusion

In this research, we propose a novel model ACDFSL to address the challenge of HSI classification under the constraints of cross-domain and few-shot learning. The model is engineered to concurrently learn a classification algorithm for target classes using a limited number of labeled samples while minimizing domain shifts incorporating the Squeeze-and-Excitation attention mechanism into the ACDFSL framework, resulting in marked improvement in HSI classification accuracy. Experiments conducted on three diverse datasets Chikusei, UP, and Salinas demonstrate the superior performance of the ACDFSL method. Not only does it outpace other deep learning methodologies when limited labeled samples are available, but it also surpasses existing HSI FSL methods. These findings confirm the effectiveness of the ACDFSL model, marking a significant step forward in the field of HSI classification. Further developments and improvements of the ACDFSL model, as well as its application to an expanded range of HSI datasets, will be the subject of future work. For reference and reuse, the source code for our work is publicly accessible on our GitHub repository at: https://github.com/Rojan119/ACDFSL-Future-Wisdom.git.

References

[1] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., vol. 7, pp. 2094-2107, 2014.

[2] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," IEEE Trans. Geosci. Remote Sens., vol. 54, pp. 6232-6251, 2016.

[3] K. Gao, B. Liu, X. Yu, J. Qin, P. Zhang, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," Remote Sens., vol. 12, pp. 923, 2020.

[4] M. Gao, Z. Zhang, G. Yu, S. Ö. Arık, L. S. Davis, and T. Pfister, "Consistency-Based Semi-supervised Active Learning: Towards Minimizing Labeling Cost," Computer Vision – ECCV 2020, vol 12355, pp. 510-526, 2020.

[5] 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 (CVPR), Seattle, WA, USA, June 27-30, 2016, pp. 770-778.

[6] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018, pp. 7132-7141.

[7] Z. Li, M. Liu, Y. Chen, Y. Xu, W. Li, and Q. Du, "Deep cross-domain few-shot learning for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 60, Article no. 5501618, 2022.

[8] T. Li, J. Zhang, and T. Zhang, "Classification of hyperspectral image based on deep belief networks," in Proceedings of the IEEE International Conference on Image Process (ICIP), Paris, France, October XX-XX, 2014, pp. 5132-5136.

[9] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 57, pp. 2290-2304, 2019.

[10] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," Advances in neural information processing systems, vol. 31, 2018, pp. 1647–1657.

[11] Q. Ran, Y. Zhou, D. Hong, M. Bi, Li. Ni, X. Li, and M. Ahmad, "Deep transformer and few-shot learning for hyperspectral image classification," CAAI Transactions on Intelligence Technology, 2023, doi: 10.1049/cit2.12181.

[12] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, J. A. Benediktsson, "Feature extraction for hyperspectral imagery: the evolution from shallow to deep: overview and toolbox," IEEE Trans. Geosci. Remote Sens. Mag., vol. 8, no. 4, pp. 60-88, 2020.

[13] D. Tran, J. Ray, Z. Shou, S.-F. Chang, and M. Paluri, "Convnet architecture search for spatiotemporal feature learning," arXiv preprint arXiv:1708.05038, 2017.

[14] Z. Wang, B. Du, Q. Shi, and W. Tu, "Domain adaptation with discriminative distribution and manifold embedding for hyperspectral image classification," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 7, pp. 1155-1159, Jul. 2019.

[15] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual Attention Network for Image Classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 6450-6458.

[16] S. Woo, J. Park, J. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Proceedings of the European Conference on Computer Vision (ECCV), September 2018, pp. 3–19.

[17] Y. Xu, Y. Zhang, T. Yue, C. Yu, and H. Li, "Graph-Based Domain Adaptation Few-Shot Learning for Hyperspectral Image Classification," Remote Sensing, vol. 15, no. 4, Article no. 1125, 2023.

[18] Y. Xu, Y. Zhang, C. Yu, C. Ji, T. Yue, and H. Li, "Residual spatial attention kernel generation network for hyperspectral image classification with small sample size," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 5529714, 2022.

[19] X. Yang, Y. Ye, X. Li, R.Y.K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," IEEE Trans. Geosci. Remote Sens., vol. 56, pp. 5408-5423, 2018.

[20] J. Yang, Y. Zhao, and J. C. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 8, pp. 4729-4742, Aug. 2017.

[21] N. Yokoya and A. Iwasaki, "Airborne Hyperspectral Data over Chikusei," Space Application Laboratory, University of Tokyo, 2016. [Online]. Available: https://opendatalab.com/Chikusei_Dataset

[22] C. Zheng, N. Wang, and J. Cui, "Hyperspectral image classification with small training sample size using superpixel-guided training sample enlargement," IEEE Trans. Geosci. Remote Sens., vol. 57, pp. 7307-7316, 2019.