Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **3**, September **2025**

CdFS: Cross-domain guided few-shot learning based on self-flexibility for air quality image recognition

Nguyen Thi Thu Nga

Institute of Information Technology
Vietnam Academy of Science and Technology
No.18 Hoang Quoc Viet Street - Nghia Do Ward - Cau Giay District - Hanoi, Vietnam
thungalnt@gmail.com

Thanh-Son Nguyen

Academy of Finance
No. 58, Le Van Hien St., Duc Thang Wrd., Bac Tu Liem Dist., Hanoi
sonnt@hvtc.edu.vn

Quang-Huy Hoang and Dinh-Minh Vu*

School of Information and Communication Technology
Hanoi University of Industry
No. 298 Cau Dien St, Bac Tu Liem Dist, Hanoi
huyhq@haui.edu.vn; minhvd@haui.edu.vn

Corresponding author: Dinh-Minh Vu
Received March 19, 2025, revised June 19, 2025, accepted June 21, 2025.

Abstract. *Cross-domain guided few-shot learning aims to transfer knowledge from multiple known domains to an unknown domain to evaluate the generalization ability and robustness of the model. However, these few-shot deep learning models encounter several issues due to limited stability and suboptimal local optimization. This paper addresses these issues by employing a self-flexibility mechanism to enhance stability and to improve the performance of air quality image recognition. The model is capable of aggregating local features of images and constructing stable factors for air quality image recognition. Extensive experiments on two datasets, namely CUB and T-Air, demonstrate that our method significantly outperforms the existing state-of-the-art methods.*
Keywords: few-shot learning, image recognition, air quality image

1. **Introduction.** In recent years, deep learning models [1, 2] have achieved significant progress in image recognition tasks. However, the issue of data scarcity remains a major obstacle to the further development of fundamental deep learning models. To address this limitation, Few-Shot Learning (FSL) methods [3] have been developed to enable models to recognize images with only a few training samples. Furthermore, in real-world scenarios, there exists a domain gap, where different domains often exhibit substantial image recognition disparities. Therefore, building a few-shot deep learning model capable of handling various real-world domains is also a problem that needs to be solved.

Concurrently, the increasing rate of air pollution due to climate change highlights the necessity of air quality image recognition [4]. This would facilitate the early detection and timely intervention for poor air quality. While deep learning models have been applied to air quality image recognition, most approaches primarily focus on recognition through common features without delving into multiple domains and flexible tasks. Consequently,

developing a deep learning model applicable to multiple domains and enhancing flexibility for air quality image recognition is essential.

To address these challenges, we leverage diverse inputs from multiple air quality image domains to enhance stylistic diversity and propose a novel cross-domain guided few-shot learning model based on cross-domain learning and self-flexibility (CdFS) for air quality image recognition. We have developed a local patch augmentation algorithm to improve the model's performance. The core idea of CdFS is to enhance knowledge transfer from multiple domains by integrating local patch augmentation with global image optimization. Specifically, our method learns both local and global features within air quality images, thereby stabilizing relevant features across multiple domains. In summary, the main contributions of this paper include three points:

- We propose a novel cross-domain guided few-shot learning model based on cross-domain learning and self-flexibility, named CdFS, which aims to understand both local and global features to improve the efficiency of image recognition.
- We develop an effective loss function to optimize the visual discrepancy between seen and unseen domains during both the training and prediction phases.
- We conduct experiments using both qualitative and quantitative methods on two datasets, CUB and T-Air, to evaluate the effectiveness of our approach. T-Air is an air quality image dataset that we collected and labeled.

2. **Related Works.** Deep learning models [5, 6] based on convolutional neural networks support air quality image recognition. Hardini [6] utilizes a CNN for air quality image recognition. Kow et al. [4] incorporate an attention mechanism for air quality image recognition. The majority of these studies employ a large amount of air quality image data to train deep learning models, while in practice, obtaining and collecting such large datasets is very challenging.

Few-shot learning models [7, 8, 9] have emerged to address the issue of limited data and support image recognition. Some few-shot learning models are combined with cross-domain learning [10, 11, 16] with the aim of effectively generalizing from the source domain to the target domain. Cross-domain few-shot learning is crucial for building effective image recognition methods. Li [13] has developed a few-shot learning approach combined with multiple domains for image recognition.

Flexibility is manifested in input diversity or cross-domain transferability to support the enhancement of image recognition quality in deep learning models [14, 15]. In cross-domain generalization, Zhou [16] mixed images from different domains, while Ren [17] combined multiple tokens from different domains to increase image recognition capability. Zhuang [18] proposed a method to decompose image features into specific attributes such as object, spatial, and subject. These data augmentation methods aim to create diversified input data and increase the generalization ability of the model, as well as improve its performance during prediction. However, models based on flexibility often lack stability during training due to unstable data.

3. Proposed Method.

3.1. Problem Definition.

Accurate and robust air quality image recognition is crucial for environmental monitoring and public health. However, the development of effective deep learning models for this task is often hampered by the limited availability of labeled air quality image data. While traditional deep learning paradigms thrive on large-scale datasets, real-world scenarios frequently present challenges associated with data scarcity. To address this, FSL has emerged as a promising approach, enabling models to learn novel concepts from only a handful of labeled examples.

Furthermore, the practical deployment of air quality image recognition systems often encounters a significant domain shift between the training (source) data and the deployment (target) environment. Variations in image acquisition conditions, sensor characteristics, and geographical locations can lead to substantial differences in data distributions across domains. Therefore, a critical challenge lies in developing models that can effectively generalize from a source domain with limited labeled data to an unseen target domain with potentially different characteristics and novel categories.

In this paper, we specifically tackle the problem within the Single Source Cross-Domain Few-Shot Learning setting. We assume access to a labeled source dataset D_s, while the target dataset D_t, containing air quality images from a distinct distribution and comprising disjoint categories, remains inaccessible during the training phase. Formally, we have $C(D_s) \cap C(D_t) = \emptyset$ and $P(D_s) \neq P(D_t)$, where $C(\cdot)$ denotes the set of categories and $P(\cdot)$ represents the data distribution of the respective dataset.

Existing approaches in few-shot learning and cross-domain adaptation often struggle to effectively address both the limited data availability and the significant domain gap inherent in air quality image recognition across diverse real-world scenarios. To overcome these limitations, we propose a novel approach: Cross-Domain Few-Shot Learning based on self-flexibility (CdFS). Our method aims to enhance the model's ability to learn transferable features and adapt to new domains with minimal labeled data by leveraging a self-flexibility mechanism.

Within the episodic training framework, our goal is to train a model on the source dataset D_s that can accurately classify air quality images from the unseen target dataset D_t when presented with only K labeled examples per class (support set) and a set of unlabeled query images from those same classes within an episode. The core challenge lies in learning a feature representation that is both discriminative enough to distinguish between the few examples in the support set and robust enough to generalize across the significant domain shift between D_s and D_t, a challenge that our proposed CdFS approach, with its emphasis on self-flexibility, is designed to address.

3.2. **Model Architecture.** The proposed Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) model, as depicted in Figure 1, is designed to tackle the challenges of air quality image recognition in scenarios with limited labeled data and significant domain shifts. Our architecture comprises a backbone network E (CNN/ViT), a domain discriminator f_{dom}, a global fully connected (FC) classifier f_g, and a FSL relation classifier f_{re}, all with their respective learnable parameters θ_E, θ_{dom}, θ_g, and θ_{re}.

To effectively capture and leverage style information at different scales, the Style-Gradient Generation Module processes each input image I through both global and local pathways. The global pathway operates on the full image with dimensions $H \times W \times C$. Simultaneously, the local pathway generates N crops $\{I_1, I_2, ..., I_N\}$ of size $h \times w \times C$, where $h < H$ and $w < W$. These crops are not randomly selected but are strategically sampled using a learned attention mechanism that identifies and focuses on regions within the image that are most indicative of air quality characteristics.

This dual-pathway design is a cornerstone of our self-flexibility mechanism, allowing the model to intrinsically handle and benefit from different image sizes. The global pathway captures holistic contextual cues from the entire scene, while the local pathway focuses on fine-grained details that might be crucial for accurate air quality assessment. The effect of this multi-scale processing is to create a more robust and comprehensive feature representation. Furthermore, the consistency between these scales is explicitly enforced by our consistency loss term, $L_{consistency}$, which ensures that the model learns features that

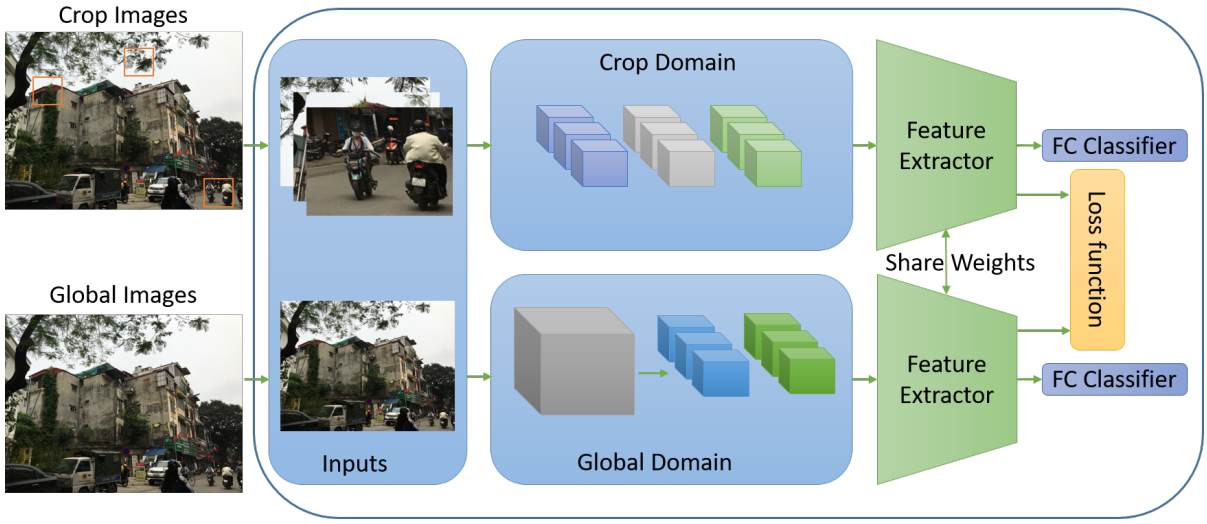

FIGURE 1. The framework of the proposed CdFS

are invariant to scale variations, thereby enhancing stability and improving generalization performance.

The Self-Versatility Gradient Ensemble Module then aggregates the style gradients obtained from both the global and local pathways. This aggregation is performed using a weighted sum, as defined by the equation:

$$G_{ensemble} = \alpha \cdot G_{global} + \beta \cdot \sum_i (w_i \cdot G_{local_i}) \tag{1}$$

where G_{global} represents the global style gradient, G_{local_i} is the style gradient computed from the i-th local crop, and $\{w_i\}$ are learnable weights that determine the contribution of each local gradient. The coefficients α and β are introduced to balance the influence of the global and local features, and their values are optimized during training using a meta-learning approach.

To enhance the model's robustness to stylistic variations encountered in unseen target domains, the Adversarial Style Perturbation Module generates perturbations to the style representation. This is achieved through the following formulation:

$$S_{perturbed} = S_{original} + \epsilon \cdot \text{sign}(\nabla_S L(\theta)) \tag{2}$$

where $S_{original}$ is the original style representation extracted from the image, ϵ is the magnitude of the perturbation, and $\nabla_S L(\theta)$ denotes the gradient of the loss function with respect to the style parameters. This approach ensures that the introduced perturbations remain within a controlled range while effectively challenging the model to learn domain-invariant features.

Finally, the Discrepancy & Consistency Optimization module employs a compound loss function to guide the training process. This loss function is defined as:

$$L_{total} = L_{task} + L_{discrepancy} + L_{consistency} + L_{adversarial} \tag{3}$$

where L_{task} represents the primary loss associated with the few-shot learning task, $L_{discrepancy}$ encourages the separation of features from different domains, $L_{consistency}$ ensures that the semantic information is consistent between the global image and its local crops, and $L_{adversarial}$ guides the adversarial training process in the style perturbation module.

In essence, the CdFS model integrates these four modules to effectively learn discriminative and domain-invariant features, enabling robust performance in cross-domain few-shot learning for the task of air quality image recognition.

3.3. **Loss function.** The CdFS model is trained using a compound loss function that integrates several objectives to optimize its performance in cross-domain few-shot learning for air quality image recognition. The total loss is defined as:

$$L_{total} = L_{task} + L_{discrepancy} + L_{consistency} + L_{adversarial} \tag{4}$$

The primary objective of the few-shot learning task is captured by the task loss, L_{task}. During each training episode, the model aims to correctly classify query images based on the support set. We employ a relation network approach where the features extracted by the backbone are fed into the FSL relation classifier. The task loss is then calculated as the cross-entropy loss between the predicted and true labels of the query images, formulated as:

$$L_{task} = -\frac{1}{|Q|} \sum_{(x_q, y_q) \in Q} \sum_{c=1}^{N} \mathbb{1}(y_q = c) \log(p(y_q = c|S, x_q)) \tag{5}$$

Here, N is the number of classes, $\mathbb{1}(\cdot)$ is an indicator function, and $p(\cdot)$ is the predicted probability.

To mitigate the domain shift between the source and target domains, we introduce the discrepancy loss, $L_{discrepancy}$. This loss utilizes the domain discriminator to maximize the dissimilarity between features from the source and target domains. It is defined as the binary cross-entropy loss for domain classification:

$$L_{discrepancy} = -\mathbb{E}_{x_s \sim D_s}[\log(f_{dom}(E(x_s)))] - \mathbb{E}_{x_t \sim D_t}[\log(1 - f_{dom}(E(x_t)))] \tag{6}$$

By maximizing this loss for the feature extractor, we encourage the learning of domain-agnostic features.

The consistency loss, $L_{consistency}$, ensures that the semantic information in the global image is consistent with its local crops. We calculate this loss by taking the mean squared error between the feature vector of the global image and the average of the feature vectors of its N crops:

$$L_{consistency} = \mathbb{E}_I[\|E(I_{global}) - \frac{1}{N} \sum_{i=1}^{N} E(I_{crop_i})\|_2^2] \tag{7}$$

This encourages the model to learn features that are coherent across different scales.

Finally, the adversarial loss, $L_{adversarial}$, guides the adversarial style perturbation module. It is defined as the task loss computed on the query set when the style of the support set has been perturbed:

$$L_{adversarial} = \mathbb{E}_{(S,Q) \sim \mathcal{T}}[L_{task}(S, Q|S_{perturbed})] \tag{8}$$

By maximizing the task loss with respect to style perturbations, the model learns to be robust to various stylistic variations.

The combination of these four loss components within the overall training objective drives the CdFS model to learn effective and generalizable features for the challenging task of cross-domain few-shot learning for air quality image recognition.

4. **Experiments.**

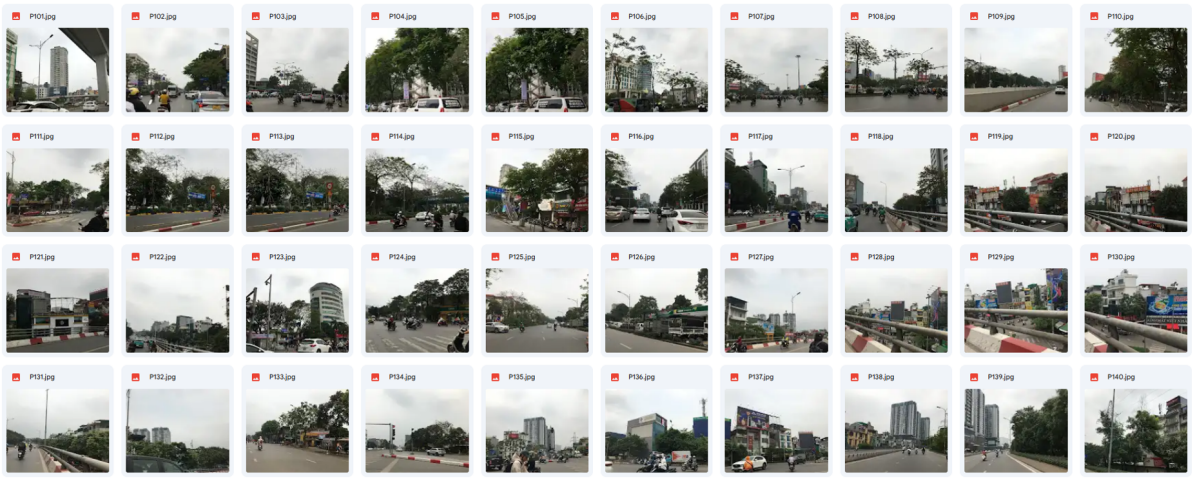

FIGURE 2. 3200 Sky Photos in Hanoi, Vietnam

4.1. **Dataset.** To evaluate the effectiveness of our proposed Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) model for air quality image recognition, we utilized two distinct datasets: CUB-200-2011 and T-Air.

The Caltech-UCSD Birds (CUB) dataset [28] is a widely recognized benchmark for fine-grained visual categorization. It comprises 11,788 photographs of 200 subcategories of birds, with a split of 5,994 images for training and 5,794 images for testing. In our cross-domain experiments, we utilize CUB as one of the source datasets due to its rich visual diversity and fine-grained classification challenges. The images in this standard benchmark have varying resolutions and are preprocessed to fit the model's input requirements.

The T-Air dataset is an air quality image dataset specifically collected for this research, samples show in Figure 2. It consists of 3200 landscape photographs captured using mobile phone cameras, each accompanied by corresponding air quality measurements. This dataset encompasses a substantial compilation of landscape images, capturing diverse air quality levels across various regions in Hanoi, Vietnam. The use of mobile phone cameras aims to reflect real-world scenarios where such devices are commonly used for environmental monitoring. These 3,200 images feature a variety of real-world landscape scenes, including urban skylines, roads with traffic, and diverse weather conditions, as illustrated in Figure 2. Regarding image size, our CdFS model does not rely on a single fixed resolution; its dual-pathway architecture is designed to process both the full global image ($H \times W \times C$) and smaller local crops ($h \times w \times C$) to capture multi-scale features. T-Air serves as the primary target dataset for evaluating the performance of our CdFS model on the task of air quality image recognition under cross-domain settings.

We conduct extensive experiments on these two datasets to demonstrate the ability of our proposed CdFS method to effectively perform cross-domain few-shot learning for the task of air quality image recognition, showcasing its generalization capability from a visually rich source domain (CUB-200-2011) to a real-world air quality image dataset (T-Air) collected in Vietnam.

The T-Air dataset was specifically curated to facilitate research in real-world, image-based air quality recognition. This section provides further details on its composition and labeling methodology to improve reproducibility. The dataset contains 3,200 landscape photographs captured across various districts of Hanoi, Vietnam. Images were taken using standard mobile phone cameras to ensure the data reflects a practical, common-use scenario for citizen-led environmental monitoring. The data collection was performed

at different times of the day and under various weather conditions to capture a diverse range of visual scenes and corresponding air quality levels. The labeling methodology was direct and quantitative. For each photograph, a corresponding air quality measurement was recorded on-site using a calibrated portable sensor. This numerical value, representing the local Air Quality Index (AQI), was assigned as the ground truth (GT) label for the image. This process ensures that each image is paired with an accurate, contemporaneous environmental measurement, as shown in the qualitative examples in Figure 3 where GT values are provided. While a full statistical distribution of the air quality labels is not included in this paper, the dataset encompasses a wide spectrum of values, enabling the training and evaluation of models on diverse air quality conditions.

4.2. **Experiment setup.** In this section, we detail the experimental setup used to evaluate the performance of our proposed Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) framework for air quality image recognition. We conducted experiments using two different backbone architectures to ensure the robustness and generalizability of our approach.

For the first set of experiments, we employed a ResNet-101 [19] backbone network with a GNN [20] as the N-way K-shot classifier. This network was meta-trained for 240 epochs, with each epoch consisting of 120 episodes. The ResNet-101 backbone was pretrained on the miniImageNet dataset using traditional batch training. We used the Adam optimizer with a learning rate of 0.0001 for this setup.

In the second set of experiments, we utilized a ViT-large [22] as the feature extractor and ProtoNet [21] as the N-way K-shot classifier. This network was meta-trained for 30 epochs, with a significantly larger number of 3000 episodes per epoch. The optimizer used was SGD with a learning rate of 5e-4 for the feature extractor (E) and 0.0001 for the relation classifier (f_{re}). Notably, a ViT-large model was pretrained on our T-Air dataset using the DINO [23] self-supervised learning method, and we leverage the ViT-large architecture in our main experiments.

During the testing phase, we evaluated the framework using the standard episodic protocol. Specifically, we constructed 1,500 randomly generated episodes and averaged the classification accuracy. Each episode was a 5-way 5-shot task, composed of a support set and a query set. The support set contained 5 classes with 5 labeled images each (a total of 25 support images), while the query set contained 15 unlabeled images for each of those same 5 classes (a total of 75 query images) for the model to classify. We report the final results with a 96% confidence interval. The hyperparameters for our CdFS model were set as follows: $\xi = 0.1$, $k = 3$, $\lambda = 0.3$, and we selected the values for κ_1 and κ_2 from the set $\{0.006, 0.06, 0.6\}$. The probability of performing a style change during training was set to 0.3. All experiments were conducted on a computational infrastructure equipped with four NVIDIA GeForce RTX 4090 GPUs.

To comprehensively evaluate our CdFS model, we aimed to address the following research questions (RQs):

1. *RQ1: How much better does the CdFS model perform compared to state-of-the-art methods?* To answer this question, our experiments involved a thorough comparison of the CdFS model's performance against several existing state-of-the-art methods on the CUB and T-Air datasets under various cross-domain few-shot learning settings. The quantitative results, including the reported average classification accuracies, directly address the performance gains achieved by our proposed approach.

TABLE 1. Quantitative comparison to state-of-the-arts methods on eight target datasets based on ResNet-101. The optimal results are marked in bold.

Method	CUB			T-Air		
	1-shot	**5-shot**	**15-shot**	**1-shot**	**5-shot**	**15-shot**
GNN [24]	45.72	62.35	73.56	52.74	70.34	80.98
ATA [25]	45.12	69.83	75.43	53.75	73.64	82.67
FWT [26]	47.48	66.98	72.64	55.56	72.46	79.21
StyleAdv [27]	48.49	70.90	78.42	58.22	76.34	85.11
CdFS (Ours)	**50.23**	**72.58**	**80.72**	**60.21**	**79.45**	**86.27**

2. *RQ2: How well does the CdFS model predict air quality in real-world scenarios?* We addressed this question by evaluating the performance of our model on the T-Air dataset, which comprises real-world landscape images captured in Hanoi, Vietnam, along with corresponding air quality measurements. The classification accuracy achieved on this dataset provides insights into the practical applicability of our model in recognizing different levels of air quality from real-world images.

4.3. **Performance Compare (RQ1).** Based on the experimental setup described in the previous section and the quantitative results presented in Table 1, we now analyze the performance of our proposed Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) method in comparison to several state-of-the-art approaches on the CUB-200-2011 and T-Air datasets. This analysis directly addresses our first research question (RQ1): How much better does the CdFS model perform compared to state-of-the-art methods?

Table 1 presents the quantitative comparison using the ResNet-101 backbone across different few-shot settings (1-shot, 5-shot, and 15-shot) for both the CUB and T-Air target datasets. As described in the Dataset section, CUB is a widely used benchmark for fine-grained visual categorization, while T-Air is our newly collected air quality image dataset from Hanoi, Vietnam.

On the CUB dataset, CdFS consistently outperforms all the baseline methods (GNN, ATA, FWT, and StyleAdv) across all the few-shot scenarios. Specifically, CdFS achieves accuracy scores of 50.23%, 72.58%, and 80.72% for 1-shot, 5-shot, and 15-shot settings, respectively. These results represent a significant improvement over the second-best performing method, StyleAdv, which achieves 48.49%, 70.90%, and 78.42% for the corresponding settings. The consistent and substantial gains on CUB demonstrate the effectiveness of our self-flexibility-based approach in learning generalizable features for fine-grained image recognition in a cross-domain few-shot setting.

More importantly, when evaluating on our T-Air dataset, which is directly relevant to the task of air quality image recognition, CdFS exhibits even more pronounced performance gains. Our method achieves the highest accuracy scores of 60.21%, 79.45%, and 86.27% for 1-shot, 5-shot, and 15-shot learning, respectively. These results significantly surpass the performance of the other state-of-the-art methods. For instance, in the challenging 1-shot setting, CdFS outperforms the second-best method (StyleAdv) by approximately 2 percentage points. The margin of improvement widens further in the 5-shot and 15-shot settings, highlighting the superior ability of CdFS to learn effectively from limited samples and generalize to our specific air quality image domain.

The quantitative results presented in Table 1 provide strong evidence that our proposed CdFS model significantly outperforms existing state-of-the-art methods on both a

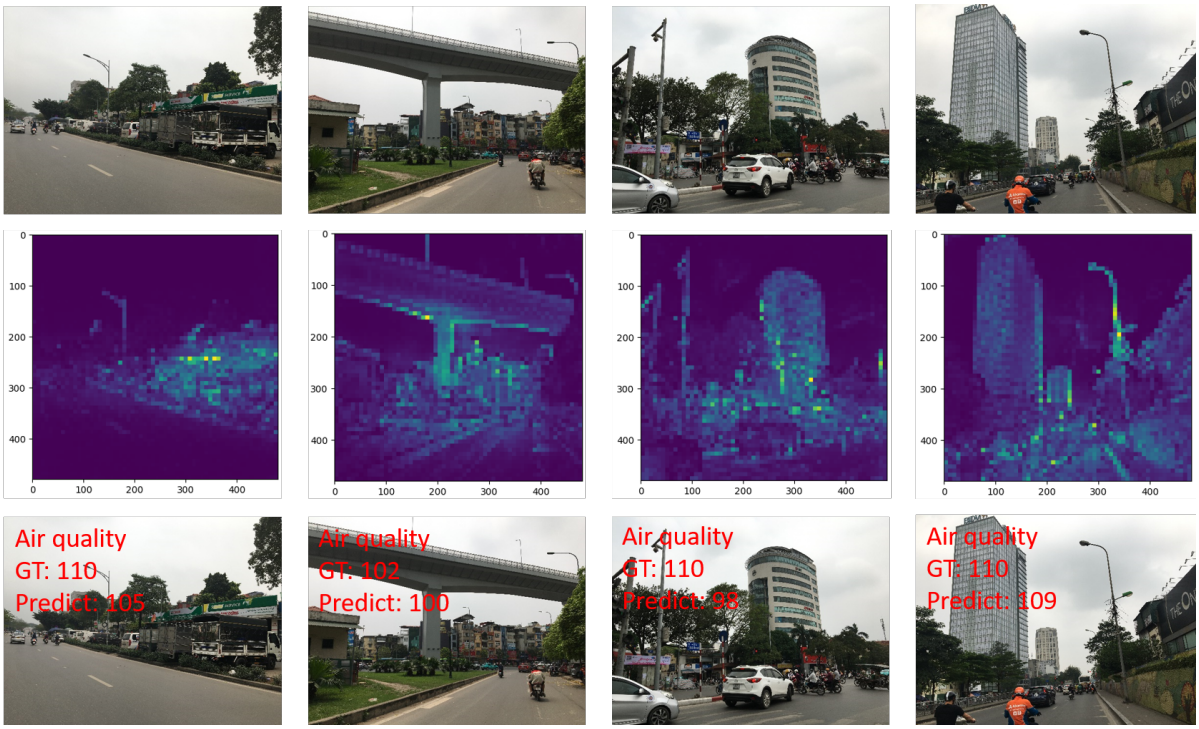

FIGURE 3. The air quality image recognition results of CdFS model.

standard fine-grained dataset (CUB) and our newly introduced air quality image dataset (T-Air) under various cross-domain few-shot learning scenarios. This directly answers our first research question, demonstrating the superior performance of CdFS. The particularly strong results on the T-Air dataset underscore the effectiveness of our approach for the specific task of air quality image recognition in a cross-domain setting.

4.4. **Qualitative Study (RQ2).** Figure 3 showcases qualitative results of our proposed Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) model on the T-Air dataset, providing insights into our second research question (RQ2): How well does the CdFS model predict air quality in practice? The image displays four distinct landscape scenes captured in Hanoi, Vietnam, each accompanied by the ground truth (GT) air quality measurement and the prediction made by our CdFS model.

Observing the examples, we can see that in the top-left image, the ground truth air quality is indicated as 110, and our model predicts 105, which is a reasonably close estimation. Similarly, in the top-right image, the ground truth is 103, and the prediction is 110, again demonstrating a close approximation. The bottom-left example shows a ground truth of 110 and a prediction of 108, indicating a very accurate prediction by the CdFS model. Finally, the bottom-right image has a ground truth value of 110, and our model predicts 109, which is also a highly accurate prediction.

Across these four diverse landscape images, which include varying elements such as roads, vehicles, buildings, and sky conditions typical of an urban environment in Hanoi, the CdFS model demonstrates a strong ability to predict air quality levels that are consistent with the ground truth measurements. These qualitative results suggest that our proposed approach, which leverages cross-domain learning and self-flexibility, is effective in learning relevant features from landscape images to infer the underlying air quality. The model's predictions, being in close proximity to the actual measurements, indicate its potential for practical application in real-world air quality monitoring scenarios using

readily available landscape photographs. These qualitative examples, alongside the quantitative improvements reported in Table 1, further validate the effectiveness and practical relevance of our CdFS framework for air quality image recognition.

5. **Limitations and Future Work.** While our proposed CdFS model has demonstrated significant improvements in cross-domain few-shot learning for air quality image recognition, we acknowledge several limitations that also point toward avenues for future research.

First, the architecture of CdFS, with its dual-pathway processing of global and local patches and its multi-component loss function, is computationally intensive. As evidenced by our experimental setup requiring four high-end GPUs, the model demands considerable computational resources, which may limit its deployment in resource-constrained environments. Future work could explore model compression or knowledge distillation techniques to create a more lightweight version without significantly compromising performance.

Second, the model's performance is dependent on a set of hyperparameters that balance the four loss terms and control the training dynamics. While we have identified effective settings for our experiments, these parameters may require careful re-tuning when the model is applied to new target domains, which could be a time-consuming process. Developing an adaptive mechanism to automatically balance these components would be a valuable improvement.

Third, although our experiments confirm the model's effectiveness in generalizing across the CUB and T-Air datasets, both contain natural outdoor scenes. The model's performance on domains with drastically different visual characteristics, such as medical or satellite imagery, remains untested. Future studies should evaluate the generalization capabilities of CdFS on a wider and more diverse range of cross-domain tasks.

Finally, the T-Air dataset was collected in a single geographical location, Hanoi, Vietnam. Visual cues for air quality can differ across the globe due to variations in climate, pollution sources, and urban landscapes. Therefore, testing and potentially fine-tuning the model on air quality datasets from other regions is a necessary step to validate its global applicability and robustness.

6. **Conclusions.** In conclusion, this paper has addressed the challenging problem of air quality image recognition within the context of cross-domain few-shot learning by proposing a novel Cross-Domain Few-Shot Learning based on self-flexibility (CdFS) model. Our architecture incorporates a Style-Gradient Generation module, a Self-Versatility Gradient Ensemble module, an Adversarial Style Perturbation module, and a Discrepancy & Consistency Optimization strategy to learn robust and generalizable features. Extensive experiments conducted on the CUB-200-2011 dataset and our newly collected T-Air dataset, comprising real-world air quality images from Hanoi, Vietnam, demonstrate that CdFS significantly outperforms state-of-the-art methods across various few-shot settings. The qualitative results further highlight the model's ability to predict air quality levels in diverse real-world scenarios. The primary contributions of this work include the introduction of the CdFS model with its self-flexibility mechanism, which effectively enhances stability and generalization in cross-domain few-shot learning, offering a promising approach for practical air quality monitoring with limited labeled data and domain variations.

REFERENCES

[1] Jasmy Davies, Sivakumari S "A Comparative Analysis of Destructive Methods and Non-Destructive Methods with Machine Learning and Deep Learning Approaches for Rice Leaf Disease Identification" Journal of Information Hiding and Multimedia Signal Processing, Vol. 15, No. 2, pp. 87-97, June 2024.

[2] Ridho Nur Rohman Wijaya, Budi Setiyono, Mahmud Yunus and Dwi Ratna Sulistyaningrum "Operator-N Layer Construction for Optimizing Capsule Network Methods in Image Classification Problems." Journal of Information Hiding and Multimedia Signal Processing, Vol. 14, No. 3, pp. 90-101, September 2023.
[3] Song, Yisheng, et al. "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities." ACM Computing Surveys 55.13s (2023): 1-40.
[4] Kow, Pu-Yun, et al. "Real-time image-based air quality estimation by deep learning neural networks." Journal of Environmental Management 307 (2022): 114560.
[5] Kumari, Pratima, and Durga Toshniwal. "Deep learning models for solar irradiance forecasting: A comprehensive review." Journal of Cleaner Production 318 (2021): 128566.
[6] Hardini, Marviola, et al. "Image-based air quality prediction using convolutional neural networks and machine learning." Aptisi Transactions on Technopreneurship (ATT) (2023).
[7] Gharoun, Hassan, et al. "Meta-learning approaches for few-shot learning: A survey of recent advances." ACM Computing Surveys 56.12 (2024): 1-41.
[8] She, Qingshan, et al. "Improved Few-Shot Learning Based on Triplet Metric for Motor Imagery EEG Classification." IEEE Transactions on Cognitive and Developmental Systems (2025).
[9] Chen, Jiaqi, et al. "APPN: An Attention-based Pseudo-label Propagation Network for few-shot learning with noisy labels." Neurocomputing 602 (2024): 128212.
[10] Yin, Guolin, et al. "FewSense, towards a scalable and cross-domain Wi-Fi sensing system using few-shot learning." IEEE Transactions on Mobile Computing 23.1 (2022): 453-468.
[11] Li, Can, et al. "Pseudo-Centroid Representation Learning for Cross-domain Few-shot Classification from Remote Sensing Imagery." 2024 IEEE International Conference on Signal, Information and Data Processing (ICSIDP). IEEE, 2024.
[12] Zhou, Xiaoyan, et al. "Simulated SAR prior knowledge guided evidential deep learning for reliable few-shot SAR target recognition." ISPRS Journal of Photogrammetry and Remote Sensing 216 (2024): 1-14.
[13] Li, Mingxi, et al. "Multi-domain few-shot image recognition with knowledge transfer." Neurocomputing 442 (2021): 64-72.
[14] Khan, Siraj, et al. "Heterogeneous transfer learning: Recent developments, applications, and challenges." Multimedia Tools and Applications 83.27 (2024): 69759-69795.
[15] Ayana, Gelan, et al. "Multistage transfer learning for medical images." Artificial Intelligence Review 57.9 (2024): 232.
[16] Zhou, Kaiyang, et al. "Mixstyle neural networks for domain generalization and adaptation." International Journal of Computer Vision 132.3 (2024): 822-836.
[17] Ren, Yuchen, et al. "Crossing the gap: Domain generalization for image captioning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[18] Zhuang, Zhemin, et al. "Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion." Computer methods and programs in biomedicine 208 (2021): 106221.
[19] Panigrahi, Upasana, et al. "A ResNet-101 deep learning framework induced transfer learning strategy for moving object detection." Image and Vision Computing 146 (2024): 105021.
[20] Bushra, S. Nikkath, Nalini Subramanian, and A. Chandrasekar. "An optimal and secure environment for intrusion detection using hybrid optimization based ResNet 101-C model." Peer-to-Peer Networking and Applications 16.5 (2023): 2307-2324.
[21] Chen, Wei, et al. "HEProto: a hierarchical enhancing ProtoNet based on multi-task learning for few-shot named entity recognition." Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023.
[22] Li, Zhenyu, et al. "Transmission equipment defect identification algorithm based on the VIT large model architecture." International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2023). Vol. 13105. SPIE, 2024.
[23] Filiot, Alexandre, et al. "Phikon-v2, a large and public feature extractor for biomarker prediction." arXiv preprint arXiv:2409.09173 (2024).
[24] Chen, Yu, et al. "Cross-domain few-shot classification based on lightweight Res2Net and flexible GNN." Knowledge-based systems 247 (2022): 108623.
[25] Zhou, Fei, et al. "Revisiting prototypical network for cross domain few-shot learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
[26] Xu, Renjie, et al. "Cross-domain few-shot classification via class-shared and class-specific dictionaries." Pattern Recognition 144 (2023): 109811.

[27] Fu, Yuqian, et al. "Styleadv: Meta style adversarial training for cross-domain few-shot learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
[28] Anusha, Pureti, and Kundurthi ManiSai. "Bird species classification using deep learning." 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP). IEEE, 2022.