Piercing the Fog: Disentangling Key Features for Vision Models in Multi-Degradation Scenarios

Siyu Chen[1], **Shiqiang Ma**[2*], **Fei Guo**[1*]

[1]School of Computer Science and Engineering, Central South University,
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
{csy619,guofei}@csu.edu.cn, sq.ma@siat.ac.cn

Abstract

In natural scenarios, vision models often encounter the challenge of complex degradation scenarios(e.g., rain, snow, fog, or motion blur). These degradations severely corrupt image features, causing existing models to treat rarely seen or unseen degraded images as "unfamiliar", thereby losing their inherent recognition and perception capabilities. To address this challenge, we propose a novel degradation disentanglement model (DDM) aimed at precisely disentangling degraded features from the image. The model enhances its perception of various degradations by controlling the matching of features across different degradation types and further strengthens the cross-correlation of target features by introducing a degradation suppression module. This enables the model to re-identify and re-localize targets while removing degradations. We validated the effectiveness of our method on more challenging few-shot segmentation datasets Degraded-Pascal and Degraded-COCO. Results on them outperform SOTA with 3.71% and 3.69% improvement respectively. The experimental results show that our method significantly improves the performance of vision models in various degradation scenarios and provides new ideas and solutions for visual understanding tasks in complex environments.

Introduction

Complex degradations (rain, snow, fog, motion blur) (Valanarasu, Yasarla, and Patel 2022) significantly alter the statistical properties and visual content of images (noise, blur, reduced contrast, and loss of detail), causing them to deviate substantially from the data distribution (see Figure 1) on which models are trained (Luo et al. 2022). This results in models being unable to effectively extract key features of target, leading to a sharp decline in performance, characterized by a "sense of unfamiliarity" and loss of perceptual ability (Wolf et al. 2021).

These issues are highly prevalent in real-world applications, especially in dynamic and unpredictable environments (Xu et al. 2019; Zhu et al. 2023; Xu et al. 2024). In autonomous driving scenarios, especially those based on pure vision solutions, image degradation not only increases driving risks but can also lead to serious traffic accidents. For example, when the vision model fails to accurately recognize traffic signs or signals, the vehicle may run a red light or enter the wrong lane, thereby causing collisions (Chang et al. 2025). Such failures in visual perception can quickly escalate into catastrophic consequences in complex traffic conditions, especially at high speeds or in dense traffic environments. Similarly, in drone applications, motion blur can distort the boundaries of target objects, making it difficult for models to accurately locate and track them (Wu et al. 2024; Wang, Wang, and Li 2024).

Although image restoration technology can improve overall quality of images, such improvements do not always benefit the segmentation task. ReconDreamer (Ni et al. 2025) proposes a dynamic driving scene reconstruction method for closed-loop simulation in autonomous driving. Its DriveRestorer module and Progressive Data Update Strategy (PDUS) can repair image degradation online. However, deblurring operations will introduce information conflicting with the segmentation task. For example, while enhancing certain features in the image, it also improperly introduce some artifacts or distortions. These undesired effects could mislead the segmentation model, leading to a decrease in the accuracy of the segmentation results.

In addition, most image restoration models are usually optimized for a specific type of degradation (Zhu et al. 2020). LID-Net (Tao et al. 2024) is a lightweight network for dehazing in foggy conditions but may not generalize well to other degradations or complex scenes (Hu et al. 2021). MetaWeather (Kim et al. 2024) performs image restoration only when support and query have same degradation types. Moreover, different types of degradation may require different restoration strategies, further complicating model design and training (Chen et al. 2022). DAMD (Lu et al. 2025) uses continuous learning to handle multiple degradation types in one model, simplifying training but increasing computational demands and complexity.

When these models encounter multi-degradation scenarios with limited training samples, their performance tends to degrade significantly. However, the real world often presents many extreme cases, and models should be capable of uncovering the underlying patterns of the real world from a small number of learnable samples (Yang et al. 2023b). To address this challenge, we propose a novel degradation disentanglement model aimed at precisely separating degraded

*Corresponding authors.

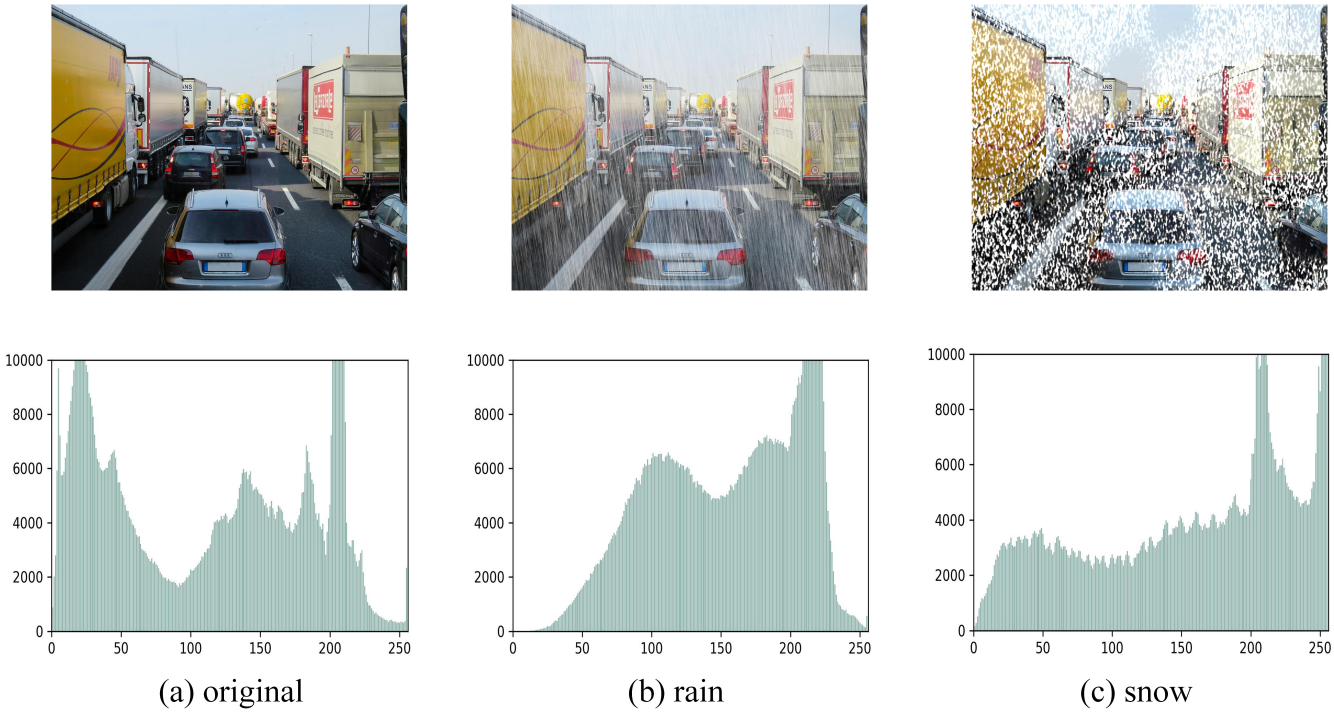

Figure 1: Histogram of pixel distribution under different degradation conditions. The horizontal axis represents the pixel value, and the vertical axis represents the number of pixels. From left to right, it can be seen that different degradation types can greatly change the pixel distribution of the same image and then corrupt encoded image features.

features from the original image features. The model enhances its perception of various degradations by controlling the matching of features across different degradation types and incorporates a degradation suppression module to strengthen the cross-correlation of target features. This enables the model to identify and locate target regions in few-shot, multi-degradation scenarios.

Our main contributions include:

- A novel degradation disentanglement model (DDM) that can precisely separate degraded features from original image features, thereby enhancing the model's robustness and generalization ability in complex multi-degradation scenarios.
- We provide two degraded and challenging few-shot segmentation datasets to validate the model's performance under extreme conditions. It randomly applies 15 distinct degradation types to images (both support and query) and simulates few-shot learning scenarios, establishing a novel benchmark for studying the adaptability of vision models in complex degradation environments.
- Significant performance improvements of our method in various degradation scenarios. This performance enhancement offers new solutions for visual understanding tasks in complex environments, reduces the dependence on large amounts of labeled data, and increases the practicality and adaptability of the model.

Related Work

Degradation Challenges and Solutions in Vision

In real-world scenarios, adverse weather (heavy rain, snow, fog), sensor noise, low light and so on combine to cause degradation during image acquisition, transmission, and storage. In this process, details may be lost, edges may be blurred, colors may be distorted, and contrast may be collapsed. This not only degrades human perception but also severely undermines the accuracy and robustness of downstream tasks such as semantic segmentation, object detection, and medical diagnosis (Li et al. 2024). To alleviate this issue, image restoration techniques have emerged. DerainNet (Fu et al. 2017), DehazeNet(Cai et al. 2016), and DnCNN (Zhang et al. 2017) design end-to-end mapping for rain, haze, and Gaussian noise, respectively. RIDNet (Anwar and Barnes 2019) and MSBDN (Dong et al. 2020) leverage dense skip connections for dehazing and channel attention for denoising. SwinIR (Liang et al. 2021) incorporates the Swin Transformer into image restoration, setting new SOTA on super-resolution and denoising tasks. However, these methods often struggle with mixed degradations in testing, as they are trained on specific, single-degradation scenarios, leading to a significant drop in performance. To solve this problem, Restormer (Zamir et al. 2022) uses efficient Transformer backbone, combining cross-channel global attention and gated depth-wise feed-forward networks to set new SOTA on 16 benchmarks. Diff-Restormer (Zhang et al.

2024) employs diffusion models to model complex degradation distributions, demonstrating promising generalization to unknown degradations.

Few-shot Semantic Segmentation

Traditional few-shot semantic segmentation has concentrated on alleviating data scarcity and improving the model's ability to generalize to unseen categories. SSP (Fan et al. 2022) generates self-support prototypes from high-confidence query predictions. FECANet (Liu et al. 2023) employs a feature-enhancement and context-reconstruction module to filter cross-class noise and incorporate background information. MIANet (Yang et al. 2023a) fuses word-embedded general category prototypes with unbiased multi-scale instance priors, aligning semantic spaces via triplet loss to alleviate intra-class variation and base-class bias.

With the rapid development of vision-language models, new methods have emerged that leverage pre-trained large language models to assist few-shot semantic segmentation (Zhou, Loy, and Dai 2022; Zhou et al. 2023). These methods use fixed text encoders to extract text, offering segmentation models semantic cues aligned with image content. Besides, the rise of large language models has further invigorated few-shot semantic segmentation. Recent research is exploring using these models for locating segmentation targets or generating descriptive text that matches images (Zhu et al. 2024a). However, the above methods all presuppose high-quality training images and ignore how image quality itself affects model performance. This is exactly the core problem we seek to explore.

Vision-language Model

Using superior vision-language alignment, vision-language models have been widely applied to downstream tasks such as zero-shot image classification, open-vocabulary detection and segmentation, cross-modal retrieval, and visual question answering. Classic CLIP (Radford et al. 2021) model, trained on vast image–text pairs, creates a shared embedding space for cross-modal representation. However, CLIP merely offers a general framework that must be tailored to specific tasks.

Current CLIP research focuses on two main directions: a) Architectural refinement: SigLIP (Zhai et al. 2023) replaces globally normalized InfoNCE loss with a binary contrastive sigmoid loss, making small batch training more feasible. CLIP-Refine (Yamaguchi et al. 2025) fine-tunes attention-pooling layer parameters of visual encoder, retaining pre-trained priors through residual mixing. b) Task-adaptive extensions of CLIP. Florence-2 (Xiao et al. 2024) unifies CLIP into a multitask vision language model through a shared prompting paradigm, which supports image captioning, question answering, and detection. M2-CLIP (Wang et al. 2024) augments CLIP for video action recognition with a temporal TED-Adapter and a multi-task decoder, achieving spatio-temporal–text alignment.

Method

Figure 2 illustrates the overall architecture of the proposed DDM model with two components: (1) Degradation-Aware Feature Representation Module (DA) and (2) Degradation Reduction Module (DR). Through CLIP, DA module simultaneously extracts two complementary representations: one that retains key feature yet still entangled with degradation feature, and another is degradation-aware. Then DR module disentangle degradation from them through hierarchical cross-attention operation. Our method provides a general model which works on different degradation types, especially when support and query images have different degradation types.

Degradation-aware Feature Representation

Stripping degradation from images is crucial for accurately segmenting targets on degraded images. This process involves two key aspects: recognizing the degradation and eliminating its effects. Previous research has primarily focused on the latter through various methods such as attention mechanisms on the channel dimension (Zamir et al. 2022), diffusion models (Islam et al. 2024), Fourier transforms (Chen et al. 2024) and so on. However, their approach can only work on single and known degradation type because they couldn't recognize the degradation type.

Our method can recognize different degradation types by controlling the matching of features across different degradation types. As we know, powerful CLIP model can provide aligned image and text features through contrastive learning, activating target features in images with relevant image category descriptions. Inspired by this capability, we propose to leverage this image-text matching ability to achieve degradation awareness in our work. Initially, we use CLIP to match image category description features with image features, obtaining original image features which is supposed to be key and degradation feature mixed. To obtain the degradation-aware feature, we creatively optimize the original mage feature and match it with descriptions of degradation types. Through this way, the degradation-aware feature highlight the degradation components in image feature and can be easier to be stripped.

In detail, given an image input $x \in R^{H\times W\times C}$ (support or query images) and degradation type (note that the degradation type is randomly generated), we use a frozen CLIP image and text encoder to get a original image feature F^h, image category feature t^h, and a degradation type feature t^d. Then, we use a small double-layer convolutional network to fine-tune F^h and get a degradation-aware image feature F^d:

$$F^d = ReLu(Conv(F^h)) \tag{1}$$

To ensure the alignment of image features with degradation types, we use contrastive learning as supervision:

$$\mathcal{L}_{con} = InfoNCE(F^d, t^d) \tag{2}$$

Through this module, we further differentiate degradation-aware feature and original image feature and highlight degradation coponents which is useful for subsequent degradation reduction.

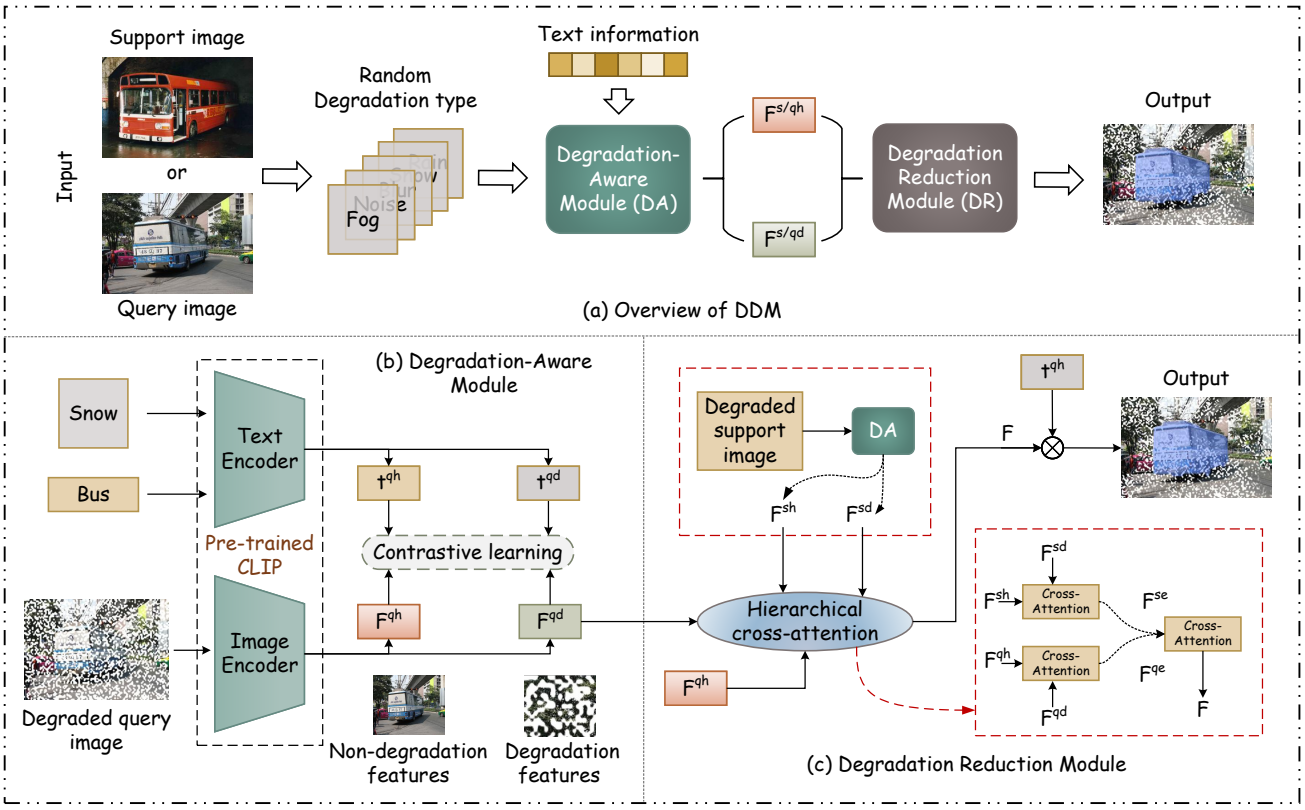

Figure 2: Overview of our model DDM. It enhances the capability of few-shot semantic segmentation on degraded images through disentangling degradation from the target feature. The superscripts **s** and **q** represent support and query, respectively.

Degradation Reduction

Through DA module, we obtain two different image feature F^h and F^d. The information in F^h is a mixture of degraded information and key features while F^d is more degradation-aware. In this module, we use hierarchical cross-attention to gradually eliminate the degradation in F^h based on the degradation information prompted in F^d.

In detail, given F^h and F^d which from DA module. To eliminate the effect of degradation, we reduce the degradation on the extracted degradation-aware feature F^d as follow:

$$F^e = CA(Q, K, V) \tag{3}$$

$$\begin{cases} Q = F^d W_q \\ K = F^h W_k \\ V = F^h W_v \end{cases} \tag{4}$$

where W_q is the learnable weights matrices of F^d, W_k and W_v are two different learnable weights matrices of F^d and CA means cross attention mechanism. Note that we do the above for both the support image and query images to obtain degradation-reduced feature F^{se} and F^{qe}. Through above operations, we highlight the key feature of target by self-focus refinement. Then, we further build mutual-correlation refinement as follow:

$$F = CA(Q_1, K_1, V_1) \tag{5}$$

$$\begin{cases} Q_1 = F^{qe} W_q \\ K_1 = F^{se} W_k \\ V_1 = F^{se} W_v \end{cases} \tag{6}$$

This way, we highlight pixel-relevant components in F^{qe} and achieve degradation-insensitive. After that, we match image category feature t^h with F^{qe} to provide additional supplemental information:

$$Corr = mul(F, t^h) \tag{7}$$

where mul means matrix multiplication. Then we decode $Corr$ to get final predicted mask.

Overall loss

The overall loss function includes the contrastive learning loss $\mathcal{L}_{con}$ and segmentation loss $\mathcal{L}_{seg}$:

$$\mathcal{L} = \mathcal{L}_{seg} + \lambda \mathcal{L}_{con} \tag{8}$$

where λ is hyperparameter for weighting different loss (set 0.2) and $\mathcal{L}_{seg}$ is the cross entropy loss between predicted mask P_{mask} and ground truth gt:

$$\mathcal{L}_{seg} = CrossEntropy(P_{mask}, gt) \tag{9}$$

Methods	Type	1-shot					5-shot				
		fold0	fold1	fold2	fold3	Mean	fold0	fold1	fold2	fold3	Mean
BAM(CVPR'22)	Clear	68.97	73.59	67.55	61.13	67.81	70.59	75.05	70.79	67.20	70.91
HDMNet(CVPR'23)	Clear	71.00	75.40	68.90	62.10	69.40	71.30	76.20	71.30	68.50	71.80
RiFeNet(AAAI'24)	Clear	68.40	73.50	67.10	59.40	67.10	70.00	74.70	69.40	64.20	69.60
Liu et al(AAAI'24)	Clear	68.30	71.30	60.00	60.70	65.10	71.50	74.50	61.50	68.40	68.90
NTRENet++(TCSVT'24)	Clear	69.50	74.80	68.20	61.80	68.60	72.00	75.80	72.40	68.10	72.10
PRFormer(TCSVT'25)	Clear	70.20	75.00	67.30	65.40	69.50	72.40	76.80	70.40	68.30	71.90
BAM(CVPR'22)	Degraded	52.02	60.08	53.32	48.40	53.46	53.07	61.86	55.66	51.99	55.65
HDMNet(CVPR'23)	Degraded	57.34	62.57	57.17	50.83	56.98	59.21	65.04	60.48	52.10	59.21
DSCM(arxiv'24)	Degraded	70.01	76.07	67.21	69.22	70.63	72.91	76.69	69.03	70.09	72.18
DDM(ours)	Degraded	**72.72**	**79.11**	**69.03**	**72.14**	**73.25**	**74.72**	**81.68**	**70.74**	**76.44**	**75.90**

Table 1: Performance comparison on Degraded-Pascal. Results (mIoU%) in bold denote the best performance.

Implementation Details

Dataset

To systematically evaluate the few-shot semantic segmentation capabilities of DDM under degraded conditions, we provide degraded versions of the two mainstream datasets: Pascal (Everingham et al. 2010) and COCO (Lin et al. 2014) as Degraded-Pascal and Degraded-COCO. In detail, we randomly apply fifteen degradations to every image in the dataset, including snow, fog, noise, blur and son on (see Appendix A) (Chen et al. 2024). For each episode, support and query images may have different degradations (the few-shot setting follows (Lang et al. 2022)). In addition, we record the degradation type of each image for contrastive learning. As far as we know, Degraded-Pascal and Degraded-COCO are the first datasets to introduce the "degradation type–degraded image" pair in few-shot semantic segmentation, providing a unified and extensible evaluation platform for degradation robust models.

Evaluation Metric

To validate the effectiveness of our proposed approach, we adopt mean Intersection over Union (mIoU) and Foreground-Background Intersection over Union (FB-IoU) as the primary evaluation metrics. mIoU is the average IoU across all classes in a multi-class segmentation problem, providing a single score that summarizes the performance of the model across all classes. FB-IoU is an extension of IoU that specifically evaluates the segmentation of the foreground objects against the background. It treats all foreground objects as one class and the background as another class, then calculates the IoU between these two classes.

Experiments

We adopt CLIP model with ViT-B/16 for contrastive learning and obtain text embedding. Besides, DINOv2 is adopted for segmentation. Adam is used as the optimizer with a learning rate of 0.001 and trained on four NVIDIA V100 GPUs. Our model is trained for 200 and 50 epochs on Degraded-Pascal and Degraded-COCO respectively, both with a batch size of 20. To evaluate our model's segmentation performance on degraded images, we retrain and make comparisons with previously well-performed FSS methods: BAM (Lang et al. 2022), HDMNet (Peng et al. 2023) and DSCM (Chang, Zhang, and Lu 2024). To further highlight the effectiveness of our method, we also compared segmentation results with RiFeNet (Bao et al. 2024) and so on (Zhu et al. 2024b; Li et al. 2024; Liu et al. 2024; Gao et al. 2025) on clear images.

Segmentation Performance

We analyzed the effectiveness of our methods, Table 1 highlights the significant performance of our model on Degraded-Pascal. In the 1-shot setting, the model must learn target features from a single sample, which is challenging due to the high demands on data quality and feature robustness. Degradations like rain, fog, and motion blur corrupt image details, making accurate feature extraction difficult. The key challenge is to quickly generalize to new categories with limited samples while handling degradation noise. In detail, compared with HDMNet and BAM method with the ResNet framework, our model's mIoU performance has been improved by **+19.79** and **+16.27** respectively, under the 1-shot setting.

In the 5-shot setting, although the model has more samples to learn from, the impact of degradations still exists. The key challenge for the model is to extract consistent target features from multiple samples. The performance of DDM is significantly enhanced, with the average mIoU metric far surpassing that of HDMNet (55.65) and BAM (59.21), reaching 75.90. Similarly, for DSCM with DINO-framework, we also outperforms it with **+2.62** points improvement under the 1-shot setting and **+3.72** points improvement under the 5-shot setting.

BAM, HDMNet, and DSCM have strong few-shot segmentation capabilities, but they experience significant performance reduction when dealing with degraded images. BAM (Lang et al. 2022) focuses on base-class bias but strug-

Methods	Type	1-shot					5-shot				
		fold0	fold1	fold2	fold3	Mean	fold0	fold1	fold2	fold3	Mean
BAM(CVPR'22)	Clear	43.41	50.59	47.49	43.42	46.23	49.26	54.20	51.63	49.55	51.16
HDMNet(CVPR'23)	Clear	43.80	55.30	51.60	49.40	50.00	50.60	61.60	55.70	56.00	56.00
RiFeNet(AAAI'24)	Clear	39.10	47.20	44.60	45.40	44.10	44.30	52.40	49.30	48.40	48.60
Liu et al(AAAI'24)	Clear	40.10	46.80	47.50	41.80	44.10	45.60	53.60	54.80	58.40	53.10
DiffewS(NeurIPS'24)	Clear	47.70	56.40	51.90	48.70	51.20	52.00	63.00	54.50	54.30	56.00
NTRENet++(TCSVT'24)	Clear	44.20	51.90	48.10	45.10	47.30	50.20	55.80	52.70	50.90	52.40
PRFormer(TCSVT'25)	Clear	49.60	50.80	45.20	50.60	49.10	54.30	55.50	49.50	56.00	53.80
BAM(CVPR'22)	Degraded	25.20	32.96	34.25	35.69	32.03	29.09	35.59	37.27	36.53	34.62
HDMNet(CVPR'23)	Degraded	31.58	41.19	40.44	38.45	37.92	34.49	46.57	43.31	43.81	42.05
DSCM(arxiv'24)	Degraded	47.52	50.22	49.82	49.34	49.23	51.24	58.20	56.61	53.95	55.00
DDM(ours)	Degraded	**49.96**	**52.65**	**50.27**	**51.31**	**51.05**	**56.60**	**63.73**	**59.73**	**59.75**	**59.95**

Table 2: Performance comparison on Degraded-COCO. Results (mIoU%) in bold denote the best performance.

Methods	FLOPs ↓	Params(M)↓	FB-IoU(%)↑	
			1-shot	5-shot
BAM(CVPR'22)	2.3T	4.9	62.44	63.43
HDMNet(CVPR'23)	670.9G	4.2	65.14	67.26
DSCM(arxiv'24)	322.6G	0.6	72.00	75.94
DDM(ours)	550.7G	4.1	**73.85**	**78.09**

Table 3: Comparison of FLOPs, the number of learnable parameters and results on Degraded-COCO in terms of FB-IoU.

gles with degraded images. When image details are damaged, BAM cannot accurately recognize target objects or non-target regions. HDMNet (Peng et al. 2023) suffers from disrupted hierarchical features and lacks degradation awareness, leading to inaccurate feature matching and segmentation errors. DSCM (Chang, Zhang, and Lu 2024), while enhancing pixel matching, mistakenly classifies noise as background in degraded images, losing critical features and blurring target boundaries.

Compared with these methods, our DA module generates features that match the degradation types through contrastive learning, enabling the model to better handle complex features in degraded images. The DR module further optimizes these features through hierarchical cross-attention operation, suppressing the noise caused by degradations, thereby demonstrating stronger robustness and generalization capabilities on degraded images.

To further validate the effectiveness of our model on complex datasets, we also conducted experiments on the Degraded-COCO dataset. Compared with Degraded-Pascal, images in Degraded-COCO has multiple objects and some of them are small, making it harder to segment the target. As shown in Table 2, compared with HDMNet, BAM and DSCM, our model has better performance on both 1-shot and 5-shot setting. For example, compared with BAM and HDMNet, the mIoU performance of our DDM model enhanced by **+19.02** and **+13.13** points improvement respectively under 1-shot setting. For 5-shot setting, the mIoU of DDM still surpasses that of BAM (34.62) and HDMNet (42.05), reaching 59.95. For DSCM, our model still outperforms it with **+1.82** (1-shot) and **+4.95** (5-shot) improvements.

Moreover, we compare the mean mIoU of our model with some methods which conducted on clear images in Table 1 and Table 2. As we can see, our model even shows better segmentation performance than some models trained on clear images. Besides, Table 3 gives a comparison in terms of FLOPs, learnable parameters and FB-IoU on Degraded-COCO for 1-shot and 5-shot segmentation (specific FB-IoU results are in Appendix C). These results confirm the precision and robustness of our model under degraded situations, effectively alleviating the negative effect caused by degradation. We completed significance tests on Degraded-Pascal with p-value $2.2e^{-8}$ (see Appendix C).

Visualization

In Figure 3, we show the qualitative results of our model on Degraded-Pascal compared with BAM (Lang et al. 2022) and previous SOTA DSCM(Chang, Zhang, and Lu 2024). We mainly show segmentation results on query images with four different degradation types (fog, blur, snow and rain). Noise will heavily damage the resolution of the support image and fog will change the color and contrast of the query image. However, compared with BAM and DSCM, our model can better learn the key features of objects with degradation decoupling, leading to better segmentation results as shown in the first row of Figure 3. Besides, we found that blurring heavily damages the edge information of images which affect the edge segmentation of the model. However, our model shows excellent results on blurring situation no matter it is applied to support or query images (as shown in the second and last rows of Figure 3). This is because our model cam learn the feature of degradation and disen-

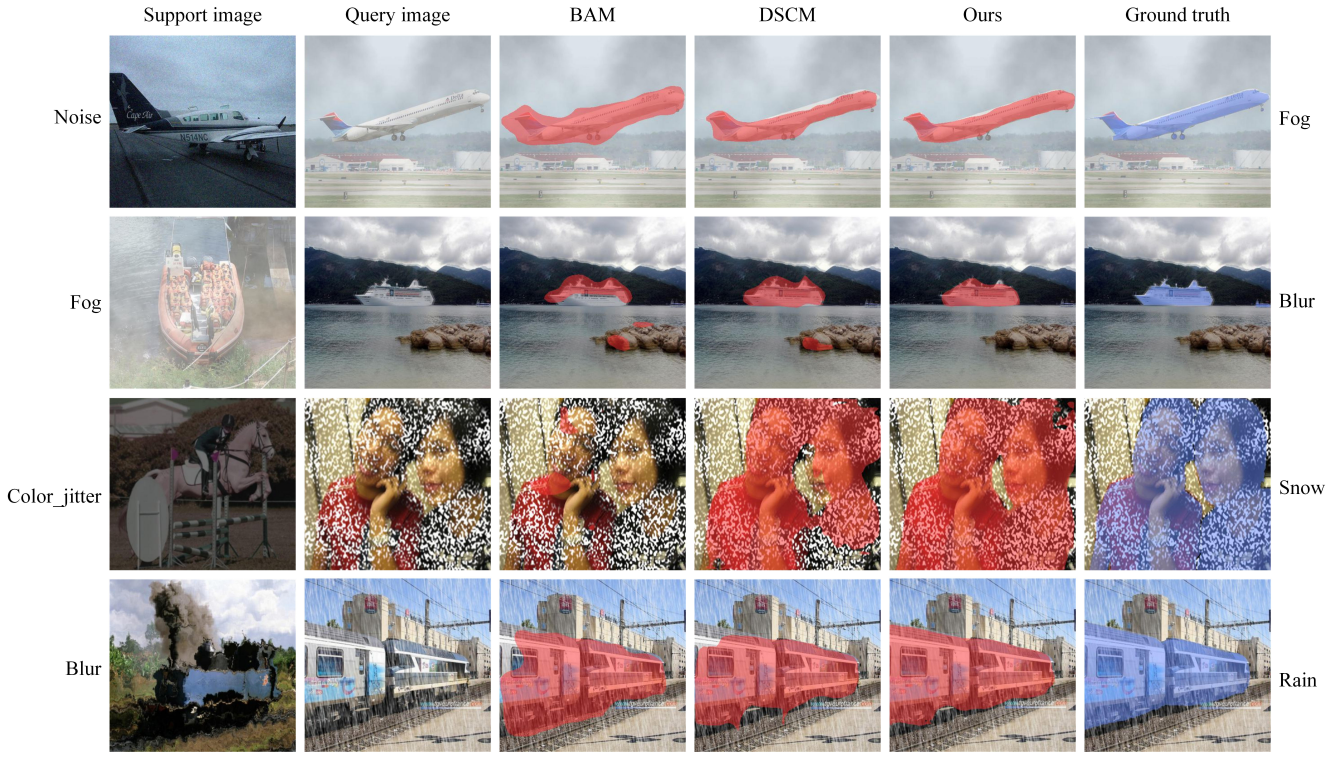

Figure 3: Qualitative comparison on Degraded-Pascal. From left to right: support images, query images, predictions of BAM, DSCM and ours, ground truth on query image. Texts on the left are the degradation types of support images and right are the types of query images.

DA	DR	5^0	5^1	5^2	5^3	mIoU(%)↑	Δ
		68.10	74.20	65.40	67.20	68.73	0
	✓	69.37	76.01	66.64	68.05	70.02	+1.29
✓	✓	72.72	79.11	69.03	72.14	73.25	+4.52

Table 4: Ablation studies of different components on Degraded-Pascal under 1-shot.

tangle it through degradation reduction, making the model focus on the key feature of the segmentation target. From the third row of Figure 3, we found that BAM shows extremely poor segmentation performance when encountered with snow. This is because the model mistakenly regards snow as a part of the segmentation target. Compared with BAM, our model has better performance because we can distinguish between the degradation feature and the key feature by degradation-aware feature representation.

Ablation Study

We conduct a series of ablation studies to investigate the impact of each component on segmentation performance (as shown in Table 4). Note that the experiments in this section are performed on Degraded-PASCAL dataset under the 1-shot setting, using DINOv2 as segmentation baseline. Firstly, we remove DA and DR modules of our model and only use DINOv2 to extract support and query features. Then, we apply DR module on F^h encoded by pres-trained CLIP and match it with t^h. Finally, we integrate DA module, which provides degradation-aware feature representation F^d and refined by DR module. As shown in the second row, DR module integrates context information of the whole image, making model focus on the key feature and improving mIoU from 68.73 to 70.02. Then, the Degradation-aware Feature Representation module significantly improves mIoU from 70.02 to 73.25. This is because DA module better distinguish degradation features from key features, alleviating content interference with the key feature of the target by degraded features.

Conclusion

This paper is the first to reveal the gap in few-shot semantic segmentation research concerning degraded images and proposes an innovative model named DDM, demonstrating remarkable performance on Degraded-Pascal and Degraded-COCO datasets. DDM controls the matching of features across different degradation types and further strengthens the self-correlation of target features by introducing a degradation suppression module. This enables the model to re-identify and re-localize targets while removing degradations. This paper not only thoroughly investigates the impact of degraded images on few-shot semantic segmentation but also provides a more robust few-shot segmentation solution. Besides, it points to a promising direction for future research.

Acknowledgments

This work is supported by grants from the National Natural Science Foundation of China (NSFC 62322215, 62532017, 62402488). This study was also supported in part by the High-Performance Computing Center of Central South University.

References

Anwar, S.; and Barnes, N. 2019. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3155–3164.

Bao, X.; Qin, J.; Sun, S.; Wang, X.; and Zheng, Y. 2024. Relevant intrinsic feature enhancement network for few-shot semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 765–773.

Cai, B.; Xu, X.; Jia, K.; Qing, C.; and Tao, D. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing*, 25(11): 5187–5198.

Chang, S.; Zhang, L.; and Lu, H. 2024. High-Performance Few-Shot Segmentation with Foundation Models: An Empirical Study. *arXiv preprint arXiv:2409.06305*.

Chang, X.; Xue, M.; Liu, X.; Pan, Z.; and Wei, X. 2025. Driving by the rules: A benchmark for integrating traffic sign regulations into vectorized hd map. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 6823–6833.

Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.-H.; Ding, J.-J.; and Kuo, S.-Y. 2022. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17653–17662.

Chen, W.-T.; Vong, Y.-J.; Kuo, S.-Y.; Ma, S.; and Wang, J. 2024. Robustsam: Segment anything robustly on degraded images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4081–4091.

Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; and Yang, M.-H. 2020. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2157–2167.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338.

Fan, Q.; Pei, W.; Tai, Y.-W.; and Tang, C.-K. 2022. Self-support few-shot semantic segmentation. In *European conference on computer vision*, 701–719. Springer.

Fu, X.; Huang, J.; Ding, X.; Liao, Y.; and Paisley, J. 2017. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6): 2944–2956.

Gao, G.; Zhang, A.; Jiao, J.; Liu, C. H.; and Wei, Y. 2025. PRFormer: Matching Proposal and Reference Masks by Semantic and Spatial Similarity for Few-Shot Semantic Segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*.

Hu, X.; Zhu, L.; Wang, T.; Fu, C.-W.; and Heng, P.-A. 2021. Single-image real-time rain removal based on depth-guided non-local features. *IEEE Transactions on Image Processing*, 30: 1759–1770.

Islam, M. T.; Alam, I.; Woo, S. S.; Anwar, S.; Lee, I.; and Muhammad, K. 2024. Loli-street: Benchmarking low-light image enhancement and beyond. In *Proceedings of the Asian Conference on Computer Vision*, 1250–1267.

Kim, Y.; Cho, Y.; Nguyen, T.-T.; Hong, S.; and Lee, D. 2024. Metaweather: Few-shot weather-degraded image restoration. In *European Conference on Computer Vision*, 206–222. Springer.

Lang, C.; Cheng, G.; Tu, B.; and Han, J. 2022. Learning what not to segment: A new perspective on few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8057–8067.

Li, J.; Shi, K.; Xie, G.-S.; Liu, X.; Zhang, J.; and Zhou, T. 2024. Label-efficient few-shot semantic segmentation with unsupervised meta-training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 3109–3117.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1833–1844.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755. Springer.

Liu, H.; Peng, P.; Chen, T.; Wang, Q.; Yao, Y.; and Hua, X.-S. 2023. Fecanet: Boosting few-shot semantic segmentation with feature-enhanced context-aware network. *IEEE Transactions on Multimedia*, 25: 8580–8592.

Liu, Y.; Liu, N.; Wu, Y.; Cholakkal, H.; Anwer, R. M.; Yao, X.; and Han, J. 2024. Ntrenet++: Unleashing the power of non-target knowledge for few-shot semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*.

Lu, X.; Xiao, J.; Zhu, Y.; and Fu, X. 2025. Continuous adverse weather removal via degradation-aware distillation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 28113–28123.

Luo, Z.; Huang, Y.; Li, S.; Wang, L.; and Tan, T. 2022. Learning the degradation distribution for blind image super-resolution. *arXiv preprint arXiv:2203.04962*.

Ni, C.; Zhao, G.; Wang, X.; Zhu, Z.; Qin, W.; Huang, G.; Liu, C.; Chen, Y.; Wang, Y.; Zhang, X.; et al. 2025. Recondreamer: Crafting world models for driving scene reconstruction via online restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1559–1569.

Peng, B.; Tian, Z.; Wu, X.; Wang, C.; Liu, S.; Su, J.; and Jia, J. 2023. Hierarchical dense correlation distillation for few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 23641–23651.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.

Tao, F.; Chen, Q.; Fu, Z.; Zhu, L.; and Ji, B. 2024. LID-Net: A lightweight image dehazing network for automatic driving vision systems. *Digital Signal Processing*, 154: 104673.

Valanarasu, J. M. J.; Yasarla, R.; and Patel, V. M. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2353–2363.

Wang, M.; Xing, J.; Jiang, B.; Chen, J.; Mei, J.; Zuo, X.; Dai, G.; Wang, J.; and Liu, Y. 2024. M2-clip: A multimodal, multi-task adapting framework for video action recognition. *arXiv preprint arXiv:2401.11649*.

Wang, P.; Wang, Y.; and Li, D. 2024. Dronemot: Drone-based multi-object tracking considering detection difficulties and simultaneous moving of drones and objects. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 7397–7404. IEEE.

Wolf, V.; Lugmayr, A.; Danelljan, M.; Van Gool, L.; and Timofte, R. 2021. Deflow: Learning complex image degradations from unpaired data with conditional flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 94–103.

Wu, Y.; Wang, X.; Zeng, D.; Ye, H.; Xie, X.; Zhao, Q.; and Li, S. 2024. Learning motion blur robust vision transformers with dynamic early exit for real-time UAV tracking. *arXiv preprint arXiv:2407.05383*.

Xiao, B.; Wu, H.; Xu, W.; Dai, X.; Hu, H.; Lu, Y.; Zeng, M.; Liu, C.; and Yuan, L. 2024. Florence-2: Advancing a unified representation for a variety of vision tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4818–4829.

Xu, C.; Guan, Z.; Zhao, W.; Wu, H.; Niu, Y.; and Ling, B. 2019. Adversarial incomplete multi-view clustering. In *IJCAI*, volume 7, 3933–3939.

Xu, C.; Si, J.; Guan, Z.; Zhao, W.; Wu, Y.; and Gao, X. 2024. Reliable conflictive multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 16129–16137.

Yamaguchi, S.; Feng, D.; Kanai, S.; Adachi, K.; and Chijiwa, D. 2025. Post-pre-training for modality alignment in vision-language foundation models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 4256–4266.

Yang, Y.; Chen, Q.; Feng, Y.; and Huang, T. 2023a. Mianet: Aggregating unbiased instance and general information for few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7131–7140.

Yang, Z.; Huang, J.; Chang, J.; Zhou, M.; Yu, H.; Zhang, J.; and Zhao, F. 2023b. Visual recognition-driven image restoration for multiple degradation with intrinsic semantics recovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14059–14070.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zhai, X.; Mustafa, B.; Kolesnikov, A.; and Beyer, L. 2023. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, 11975–11986.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7): 3142–3155.

Zhang, Y.; Zhang, H.; Chai, X.; Cheng, Z.; Xie, R.; Song, L.; and Zhang, W. 2024. Diff-restorer: Unleashing visual prompts for diffusion-based universal image restoration. *arXiv preprint arXiv:2407.03636*.

Zhou, C.; Loy, C. C.; and Dai, B. 2022. Extract free dense labels from clip. In *European conference on computer vision*, 696–712. Springer.

Zhou, Z.; Lei, Y.; Zhang, B.; Liu, L.; and Liu, Y. 2023. Zegclip: Towards adapting clip for zero-shot semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11175–11185.

Zhu, L.; Chen, T.; Ji, D.; Ye, J.; and Liu, J. 2024a. Llafs: When large language models meet few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3065–3075.

Zhu, L.; Deng, Z.; Hu, X.; Xie, H.; Xu, X.; Qin, J.; and Heng, P.-A. 2020. Learning gated non-local residual for single-image rain streak removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(6): 2147–2159.

Zhu, M.; Liu, Y.; Luo, Z.; Jing, C.; Chen, H.; Xu, G.; Wang, X.; and Shen, C. 2024b. Unleashing the potential of the diffusion model in few-shot semantic segmentation. *Advances in Neural Information Processing Systems*, 37: 42672–42695.

Zhu, Y.; Wang, T.; Fu, X.; Yang, X.; Guo, X.; Dai, J.; Qiao, Y.; and Hu, X. 2023. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 21747–21758.