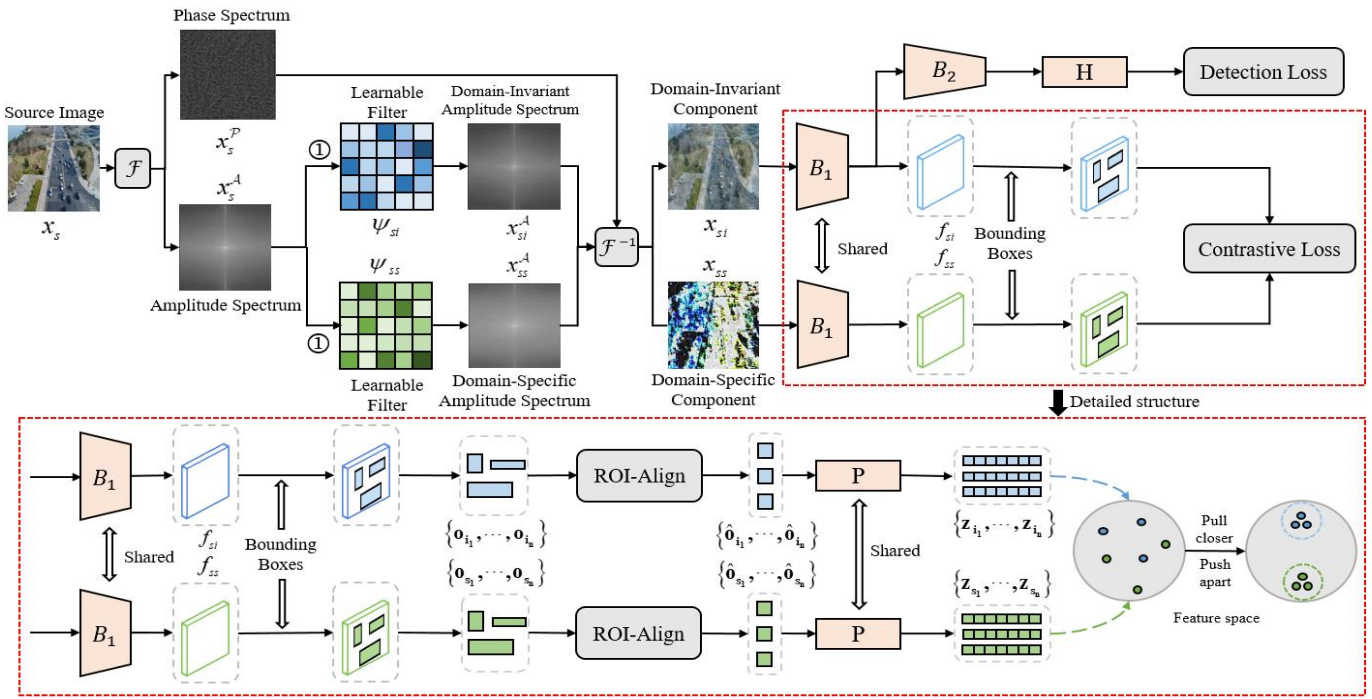
Generalized UAV Object Detection via Frequency Domain Disentanglement

Kunyu Wang, Xueyang Fu, Yukun Huang, Chengzhi Cao, Gege Shi, Zhengjun Zha

University of Science and Technology of China, China

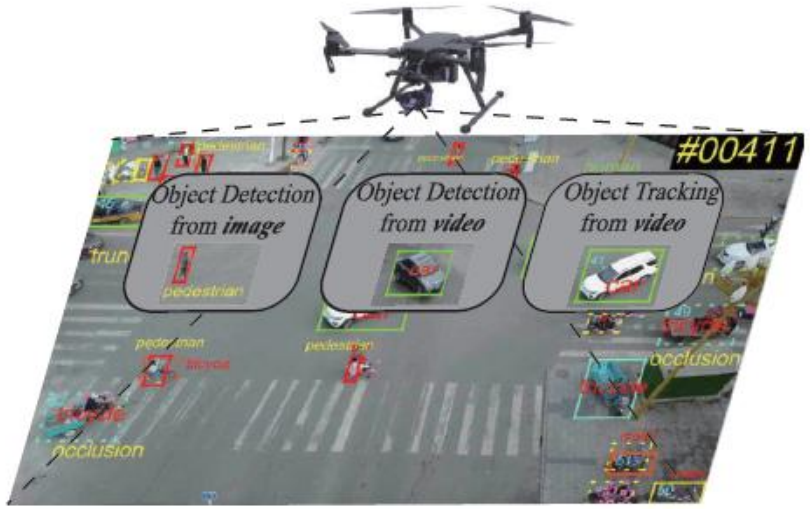
Poster ID: TUE-AM-101

Overview



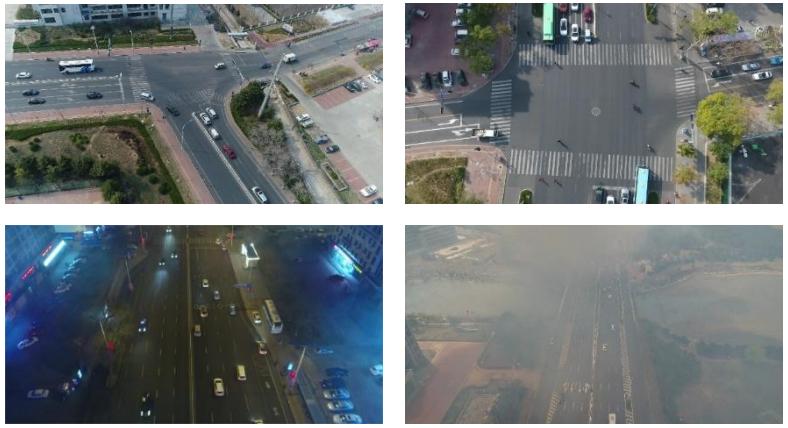
- To our best knowledge, this is the first attempt to learn generalized UAV-OD via frequency domain disentanglement, which provides a novel view angle in this area.
- We propose a new framework that utilizes two learnable filters to extract the domain-invariant and domain-specific spectrum and design an instance-level contrastive loss to guide the disentangling process.
- Extensive experiments on three unseen target domains reveal that our method enables the UAV-OD network to achieve superior generalization in comparison to the baseline and state-of-the-art methods.

Introduction



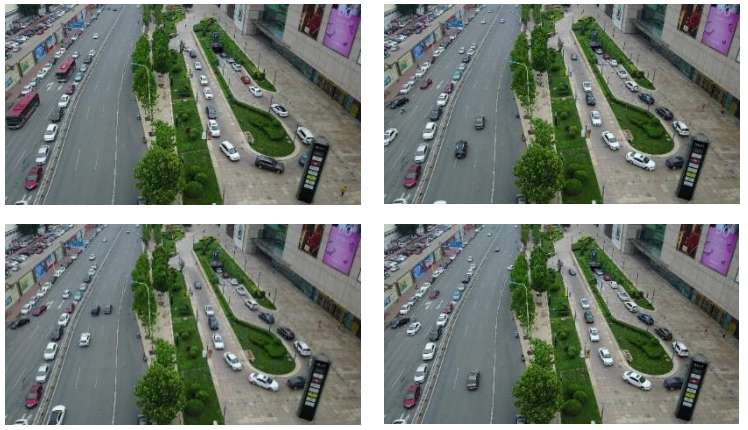
[1] Wu X, Li W, Hong D, et al. Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2021.
[2] Du D, Qi Y, Yu H, et al. The unmanned aerial vehicle benchmark: Object detection and tracking[C]//Proceedings of the European conference on computer vision, 2018.

Motivation



UAV-OD scenarios

Severe changes in global contextual information



Generic OD scenarios

Relatively stable global contextual information

Observation

Reject band	Various Scene			Diverse Illumination			Adverse Weather			Average		
	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP
Null (full band)	66.0	37.6	36.7	11.1	3.4	4.8	42.3	14.9	19.6	39.8	18.6	20.4
$\alpha = 0, \beta = 0.01$	60.0	30.2	32.6	6.4	1.9	2.75	39.5	16.1	19.0	35.3	16.1	18.1
$\alpha = 0.01, \beta = 0.1$	61.4	30.3	32.8	**39.1**	**15.9**	**19.8**	**42.6**	**18.6**	**20.8**	**47.7**	**21.6**	**24.5**
$\alpha = 0.1, \beta = 1$	**70.2**	**35.1**	**37.1**	29.4	10.2	13.6	38.2	10.6	16.7	45.9	18.6	22.5

Table 1. We first conduct preliminary experiments, i.e., exploring whether all spectrum bands contribute equally to the generalization for the UAV-OD task, to gain insight into how to implement our idea. If not, we can extract the spectrum that is conducive to generalization and use it to train the UAV-OD network to enhance its generalization. We can observe that different bands contribute differently to the UAV-OD generalization.

$$\mathcal{F}(x)(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)e^{-j2\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)}.$$

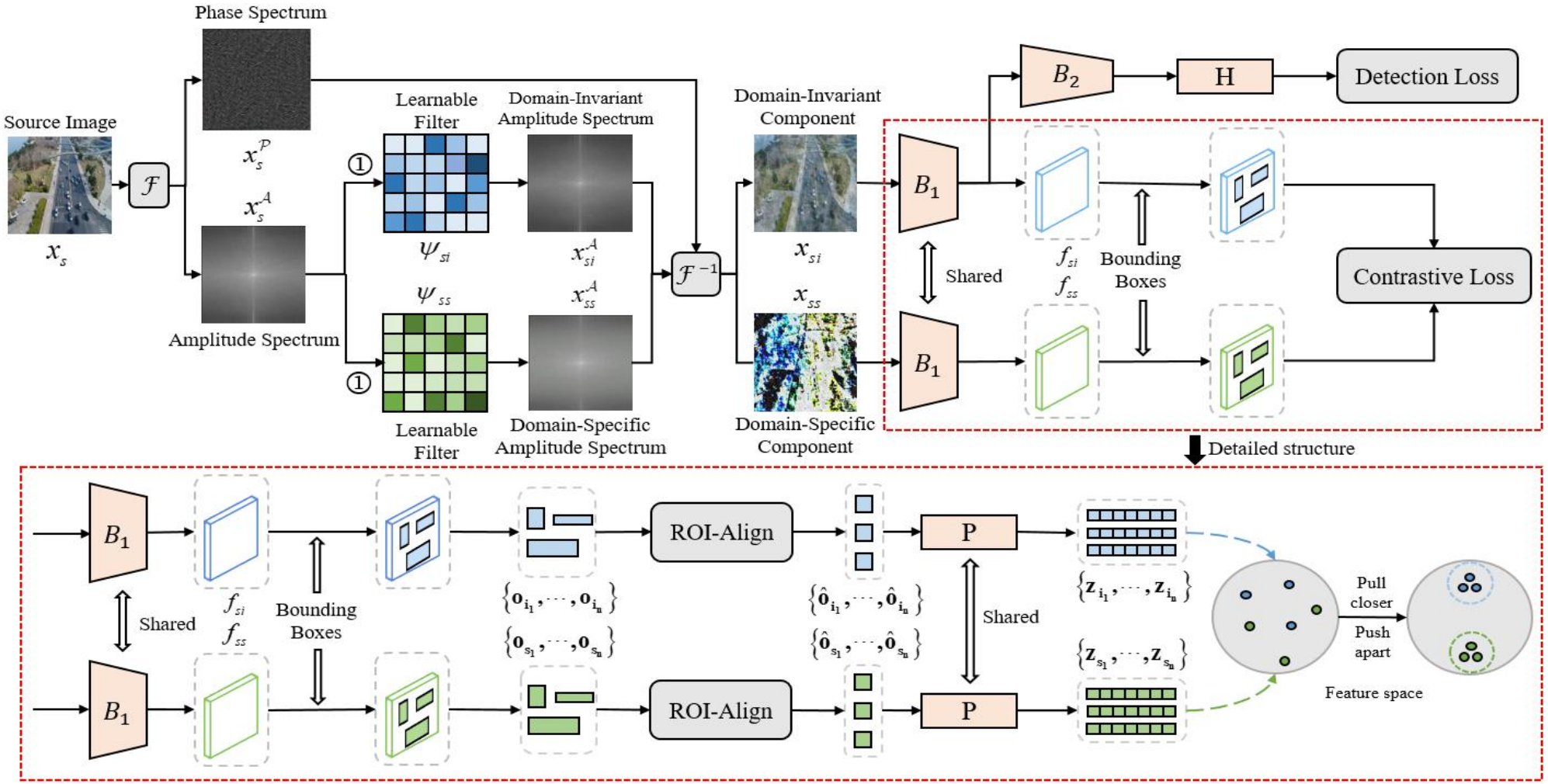
$$\mathcal{A}(x)(u,v) = \left[\mathcal{R}^2(x)(u,v) + \mathcal{I}^2(x)(u,v)\right]^{1/2},$$

$$\mathcal{P}(x)(u,v) = \arctan\left[\frac{\mathcal{I}(x)(u,v)}{\mathcal{R}(x)(u,v)}\right],$$

$$f_s(i,j) = \begin{cases} 1, & i \in [\frac{\alpha H}{2}, \frac{\beta H}{2}] \cup [\frac{(1-\alpha)H}{2}, \frac{(1-\beta)H}{2}] \\ & j \in [\frac{\alpha W}{2}, \frac{\beta W}{2}] \cup [\frac{(1-\alpha)W}{2}, \frac{(1-\beta)W}{2}] \\ 0, & \text{otherwise} \end{cases}$$

$$\mathcal{A}(x) = \hat{\mathcal{A}}(x) \otimes f_s,$$

Framework


Phase Spectrum
Source Image
Learnable Filter
Domain-Invariant Amplitude Spectrum
Domain-Invariant Component
Amplitude Spectrum
Learnable Filter
Domain-Specific Amplitude Spectrum
Domain-Specific Component
Shared
Bounding Boxes
Detection Loss
Contrastive Loss
Detailed structure
ROI-Align
ROI-Align
Shared
Pull closer
Push apart
Feature space

Experiments

Method	Various Scene			Diverse Illumination			Adverse Weather			Average		
	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP
Baseline	66.0	37.6	36.7	11.1	3.4	4.8	42.3	14.9	19.6	39.8	18.6	20.4
JiGen [2]	61.3	26.4	35.8	33.5	12.9	15.9	45.5	19.5	**22.7**	46.8	19.6	24.8
RSC [13]	73.3	48.7	44.3	14.6	6.2	7.3	47.1	16.6	21.2	45.0	23.8	24.3
StableNet [39]	75.0	48.8	44.9	18.6	9.0	9.5	47.5	17.0	21.1	47.0	24.9	25.2
Single-DGOD [30]	73.7	49.3	43.6	27.5	11.9	13.8	47.3	18.7	22.8	49.5	26.6	26.7
Ours	**75.1**	**49.7**	**45.3**	**39.0**	**18.5**	**20.7**	**48.0**	**17.2**	22.3	**54.0**	**28.4**	**29.4**

Table 2. Comparisons of the domain generalization results. All methods are trained on daylight images from UAVDT and tested on unseen images with various scene structures, diverse illumination conditions, and adverse weather conditions from UAVDT. The average generalization performance across three unseen target domains is "average".

Method	Various Scene			Diverse Illumination			Adverse Weather			Average		
	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP
Baseline	53.4	21.8	26.0	31.0	13.6	15.4	37.6	7.6	14.5	40.7	14.3	18.6
JiGen [2]	62.8	28.6	32.3	36.5	10.1	15.8	39.5	**11.8**	17.4	46.3	16.8	21.8
RSC [13]	65.3	25.5	31.3	33.5	11.6	15.3	39.3	9.0	16.3	46.0	15.4	21.0
StableNet [39]	64.3	25	30.6	31.7	15.2	16.3	**40.7**	9.9	16.8	45.6	16.7	21.2
Single-DGOD [30]	62.9	26.3	30.7	**36.8**	16.7	18.5	34.8	7.3	13.9	44.8	16.8	21.0
Ours	**65.8**	**33.9**	**35.4**	36.4	**19.5**	**19.3**	40.6	11.3	**18.4**	**47.6**	**21.6**	**24.4**

Table 3. Comparisons of the cross-dataset domain generalization results. All methods are trained on daylight images from VisDrone2019 and tested on unseen images with various scene structures, diverse illumination conditions, and adverse weather conditions from UAVDT.

Experiments

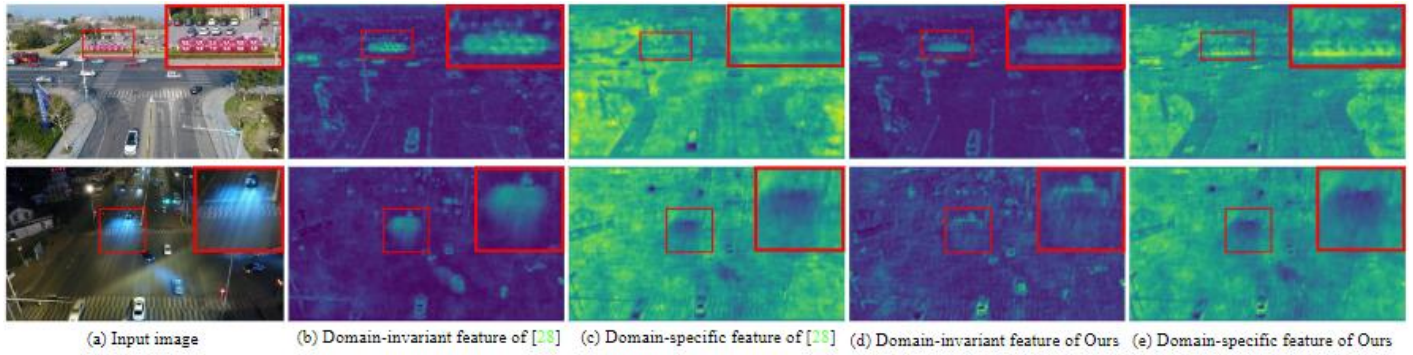


Figure 1. Comparisons of domain-invariant and domain-specific features of Single-DGOD and our method. Our method achieves a more thorough disentanglement.

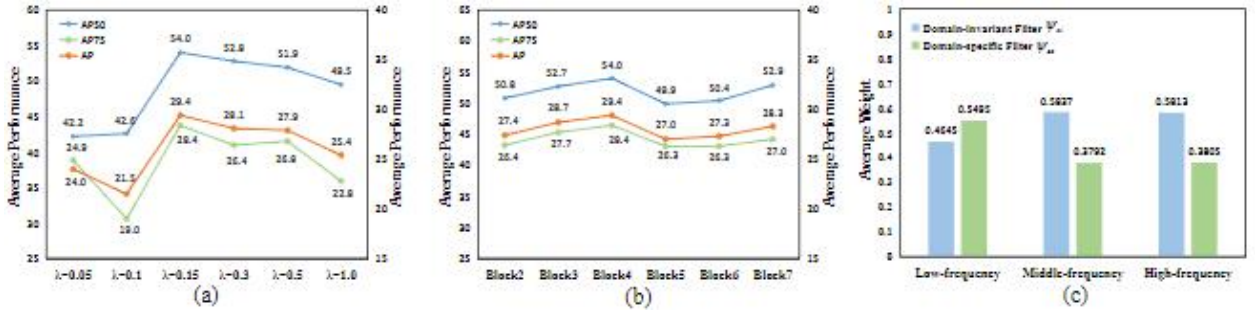


Figure 2. (a) Ablation analysis of the hyper-parameter (b) Ablation analysis of the backbone division. (c) Statistic analysis of two learnable filters We can conclude that middle- and high-frequency components contain more domain-invariant information than the low-frequency component.


1958
University of Science and Technology of China


JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thank you!