DANet: Multi-scale UAV Target Detection with Dynamic Feature Perception and Scale-aware Knowledge Distillation

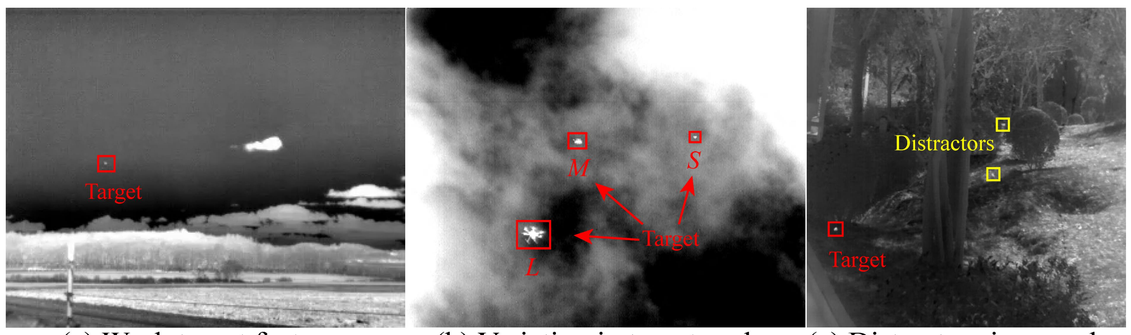
Houzhang Fang, Zikai Liao, Lu Wang, Qingshan Li, Yi Chang, Luxin Yan, Xuhua Wang
Email: lzk773629528@gmail.com, {Houzhangfang, wanglu}@xidian.edu.cn

ACM multimedia
Ottawa, Canada OCT 29 – NOV 3, 2023

XIDIAN UNIVERSITY
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Introduction

Detection of unmanned aerial vehicles (UAVs) using infrared imaging measures usually involves three challenges:
(a) Weak target features (dim in illuminance and small in size, easy to be submerged);
(b) Variation in target scales (dynamic flying creates a varying UAV scale range);
(c) Distractors in complex backgrounds (*i.e.*, birds and leaves similar to UAVs);



(a) Weak target features (b) Variation in target scales (c) Distractors in complex backgrounds

Current state-of-the-art methods either fail to distinguish real UAV targets in the complex backgrounds, or detect poorly for targets of varying scales. This leads to missed detections and false alarms, which are impractical for real-world UAV surveillance.

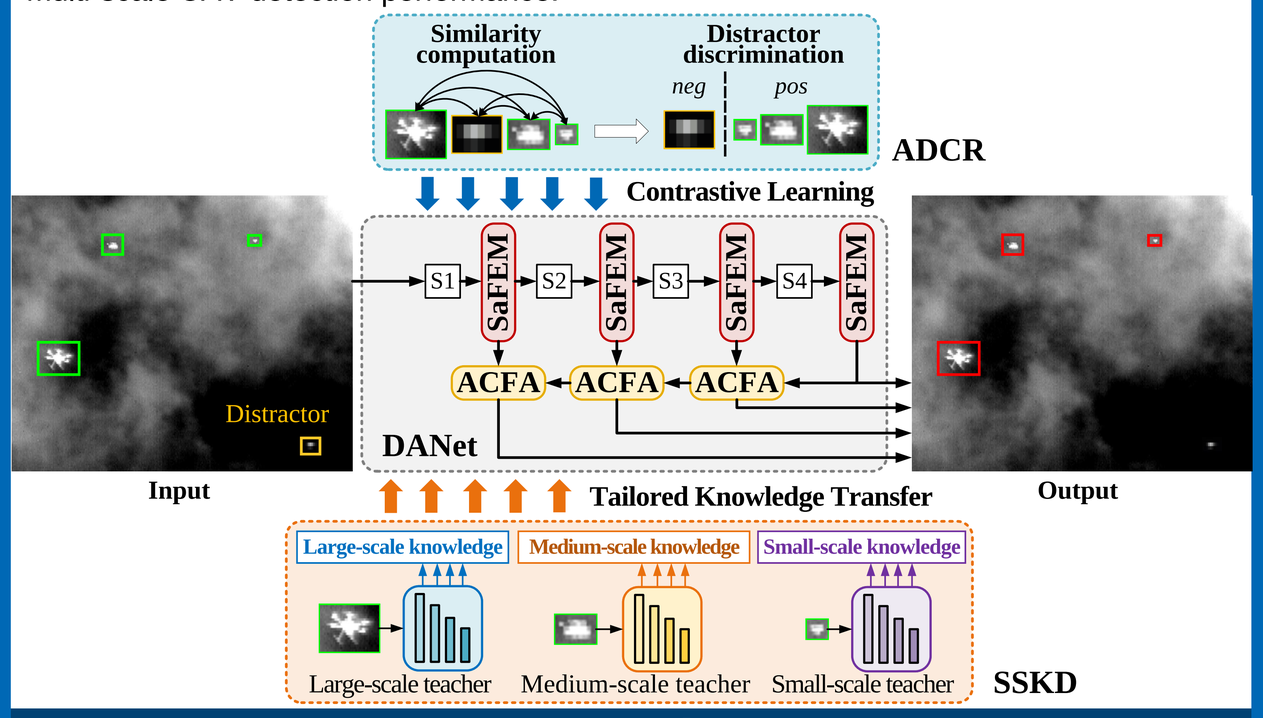
Overview of Our Approach

According to the three challenges in infrared UAV detection task, we propose DANet to address the problems, as the main novelties of our paper are:

1. Scale-adaptive Feature Enhancement Mechanism (SaFEM) and Attention-guided Cross-weighting Feature Aggregator (ACFA). We use the former to enhance critical multi-scale UAV features via separable deformable convolution, which generates attention weights by precisely perceiving target regions with dynamic receptive fields. The latter exchange critical spatial and semantic properties from shallow and deep layers for the accurate representation of UAV features from different network levels.

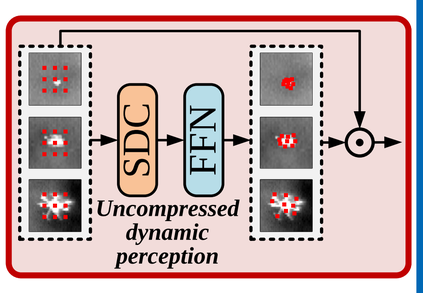
2. Anti-Distractor Contrastive Regularization (ADCR). To reduce the false alarm rate, we extracts features from regions of distractors in the background and real UAV targets, and enforce similarity computation on them to improve the discrimination ability of our model. This mechanism is only employed during training, and removed during inference.

3. Scale-Specific Knowledge Distiller (SSKD). To further increase the multi-scale UAV detection performance with no extra model complexity, we use three teacher models (small-, medium-, and large-scale teacher, each individually trained on small-, medium-, and large-scale dataset) to simultaneously train our DANet as the student model, which transfer tailored knowledge of infrared UAV targets of different scales, and thus improves multi-scale UAV detection performance.



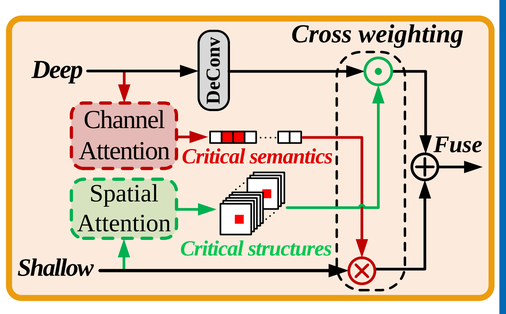
SaFEM and ACFA

SaFEM: We use deformable convolution to obtain attention weights. To reduce complexity, we separate it to a depthwise part (SDC) and a pointwise part (FFN). SDC is a grouped deformable convolution to extract multi-scale spatial properties, and FFN is a two-layer feed-forward network that projects channel information. Without any dimensionality reduction, SaFEM can dynamically perceive and highlight multi-scale UAVs accurately.



SaFEM

ACFA: To fully leverage cross-level features from different network layers to complement feature details, we use channel / spatial attention to highlight semantic / structural properties on shallow / deep feature maps. SDC is also employed in both attentions to ensure the accurate extraction of multi-scale UAV features. Guided by exchanged critical semantic / structural information, multi-scale UAV features can be represented precisely.



ACFA

Our Related Works

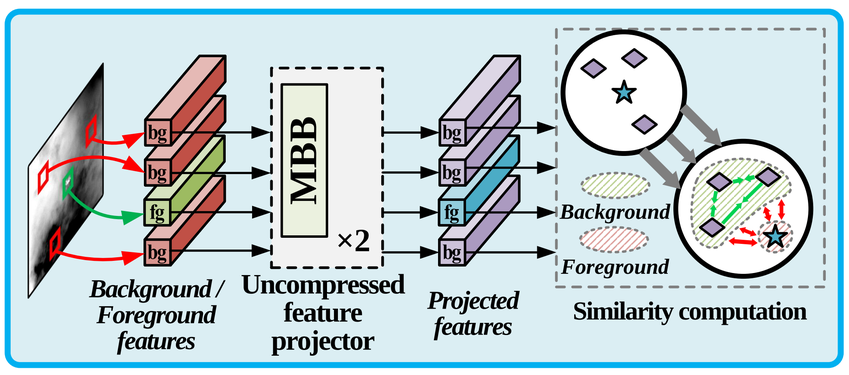
[1] Houzhang Fang, Zikai Liao, et. al., "**Differentiated Attention Guided Network Over Hierarchical and Aggregated Features**", *IEEE TII*, 2023
[2] Houzhang Fang, Lan Ding, et. al., "**Infrared Small UAV Target Detection based on Depthwise Separable Residual Dense Network and Multiscale Feature Fusion**", *IEEE TIM*, 2022
[3] Houzhang Fang, Mingjiang Xia, et. al., "**Infrared Small UAV Target Detection based on Residual Image Prediction via Global and Local Dilated Residual Networks**", *IEEE GRSL*, 2022
[4] Houzhang Fang, Xiaolin Wang, et. al., "**A Real-Time Anti-Distractor Infrared UAV Tracker with Channel Feature Refinement Module**", *ICCVW*, 2021

Anti-Distractor Contrastive Regularization

- **Positive / negative samples:** We take regions of real UAV target as the positive samples, and those of distractors as the negative samples. These regions are extracted from outp-ut features maps after SaFEMs at each stage, and are determined by the labels as well as the downsampling strides of the network.
- **Feature projection:** To avoid losing valuable information by feature compression, we propose the uncompressed feature projector (UFP) to project features with two multi-branch blocks (MBBs). It maintains original dimensionality of the features to preserve crucial details of UAVs and distractors for a better discrimination of them.
- **Loss function:** We enforce similarity computation on features of real UAV targets and distractors, and calculate its contrastive loss as:

$$\mathcal{L}_{ADCR} = \sum_{a=1}^{N_p} \frac{-1}{N_p} \sum_{p=1}^{N_p} \log \frac{\Phi(s_a, s_p)}{\sum_{p=1}^{N_p} \Phi(s_a, s_p) + \sum_{n=1}^{N_n} \Phi(s_a, s_n)}$$

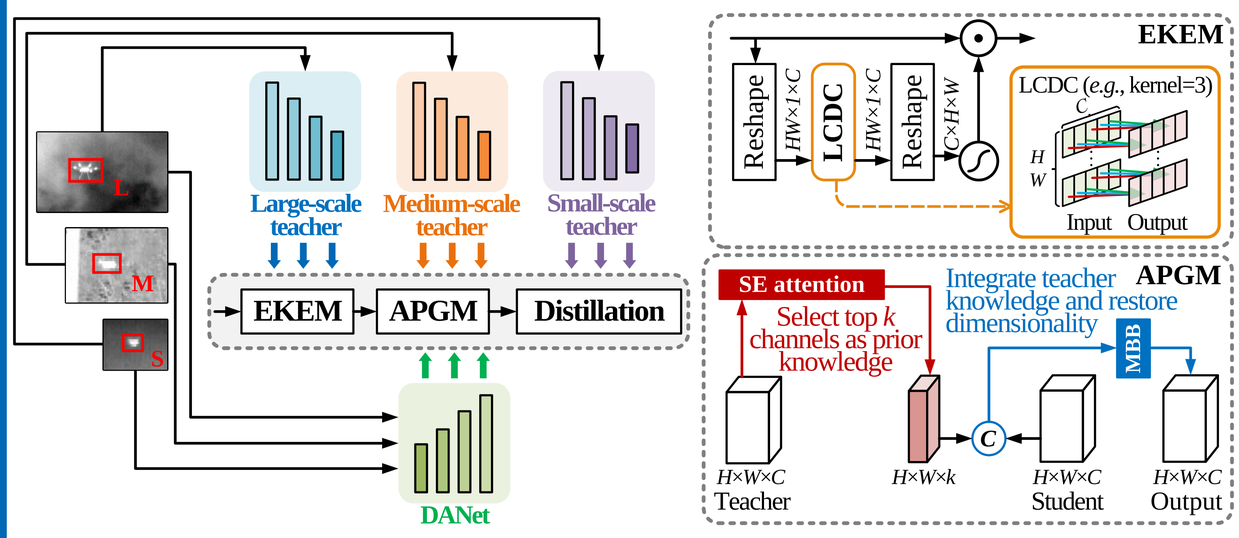
$$\Phi(x, y) = \exp((x \cdot y)/\mu)$$



Scale-Specific Knowledge Distiller

We establish the Scale-Specific Knowledge Distiller based on a "divide-and-conquer" strategy.

- **Divide:** We separate multi-scale UAV detection task into three single-scale UAV detection tasks, *i.e.*, small-, medium-, and large-scale. To this end, three individual models are designed to fulfill these three-scale detection tasks, respectively, and pretrained to their optimal to ensure the best knowledge to learn.
- **Conquer:** During distillation, we ensure the best knowledge transfer by adopting the proposed element-wise knowledge enhancement module (EKEM) and attentional prior-knowledge guidance module (APGM). EKEM can enhance UAV features and suppress background clutters for effective distillation, and APGM can alleviate learning difficulty due to model capacity gaps by integrating key prior knowledge from teachers to student before distillation. With EKEM and APGM, discrete knowledge from three teachers can be effectively transferred, and thus improving detection performance on each target scale of our DANet.



Experiments

We conduct experiments on real UAV dataset consisting of small-, medium-, and large-scale datasets, collected from our datasets of 31,329 images and public datasets (IRSTD-1k, NUAA-SIRST, and "A dataset for multi-sensor drone detection") of 10,712 images.

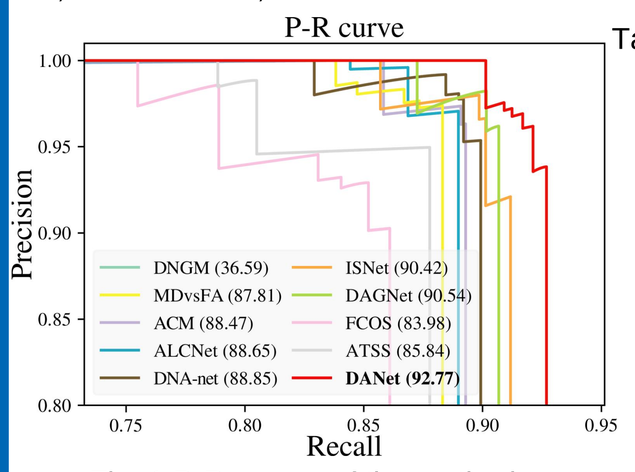


Fig. 1 P-R curves of the methods.

Table 1 Detection performance of the methods.

Method	Evaluation metrics			
	P	R	F1	FPS
DNGM	37.17	40.02	38.54	14.99
MDvsFA	85.69	88.18	86.92	28.84
ACM	89.14	88.68	88.91	**36.18**
ALCNet	88.58	90.29	89.40	34.51
DNA-Net	89.54	87.06	88.28	27.59
ISNet	91.23	90.61	90.92	33.16
DAGNet	92.75	90.09	91.40	34.95
FCOS	86.88	82.97	84.88	26.10
ATSS	87.63	90.46	89.03	29.77
Ours	**95.68**	**92.51**	**94.07**	34.57

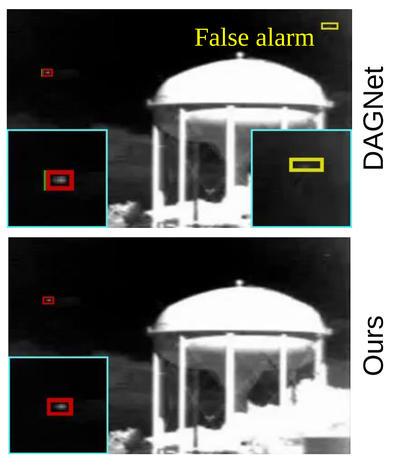


Fig. 2 Detection results visualization.

Table 2 Ablation study of the proposed method.

Model				Metrics			
SaFEM	ACFA	ADCR	SSKD	P	R	F1	FPS
-	-	-	-	90.07	87.93	88.98	**38.86**
✓	-	-	-	90.88	89.95	90.41	31.59
-	✓	-	-	90.39	88.57	89.47	34.83
-	-	✓	-	92.10	89.02	90.53	34.91
-	-	-	✓	91.93	90.44	91.18	35.85
✓	✓	-	-	93.85	91.76	92.79	35.85
✓	✓	✓	-	94.11	91.59	92.83	30.29
✓	✓	-	✓	93.87	92.19	93.02	31.58
✓	✓	✓	✓	**95.68**	**92.51**	**94.07**	31.57