Supplementary Material:
Fusion or Confusion? A Look at Dataset Pooling for Infrared Object Detection

Stefan Becker ✉, Ann-Kristin Grosselfinger, Jens Bayer,
David Münch, Wolfgang Hübner, Michael Arens
Fraunhofer IOSB, Gutleuthausstr. 1, 76275 Ettlingen, Germany
Fraunhofer Center for Machine Learning
{firstname.lastname}@iosb.fraunhofer.de

Abstract

In this paper, we provide supplementary material to the submitted main paper.

1. Experimental Setup

A more detailed description of the experimental setup can be found in the main paper.

1.1. Detection Model

For object detection, we exemplarily build on a recent variant of the *You Only Look Once* (YOLO) object detection family, in particular on the YOLOv9[1] [11].

1.2. Dataset Pool

In our experiments, the data pool includes seven publicly available infrared datasets with 'vehicle labels.

As a reminder, from the category of *domain-specific* datasets, we use the following three *ground-level* datasets: the FLIR (v2) dataset [2], the M3FD dataset [6], and the MCNet dataset. [12]. A data pool consisting only of datasets from this category is referred to as the *Ground* pool. As representative datasets from the *mid-altitude* category, the DroneVehicle dataset [9], the WIT-UAS dataset [5], the TrafficNight dataset [13], and the HIT-UAV dataset [10] are selected. A data pool only considering datasets from this category is referred to as the *Aerial* pool and we will use this term to describe this category. The full training dataset pool is refereed to as the *All* pool.

The key characteristics of the selected datasets are summarized in Table 1. This includes the considered vehicle sub-categories that are mapped to the super-category 'vehicle'. We sub-sample sequences with a fixed camera position to avoid over-sampling from redundant backgrounds and to balance the datasets. However, while the datasets are not perfectly balanced in terms of individual samples, both single-dataset models and pooled-dataset models utilize the same amount of data from each individual dataset.

Figure 1 shows example images from the used datasets in our experiments.

1.3. Pre-Processing

The following three methods are used in the experiments. The methods are applied to the entire dataset.

- *Histogram equalization* (HE)
- *Contrast limited adaptive histogram equalization* (CLAHE) [7]
- FieldScale [3]

1.4. Design Adaptations

Furthermore, we explore variations in training by fully fine-tuning or freezing the model's backbone. Lastly, we compared standard 'vehicle' detectors with models adapted to predict dataset affinity scores. Unlike [1], where a second classifier head is integrated, our approach employs the default classification head to infer the dataset origin.

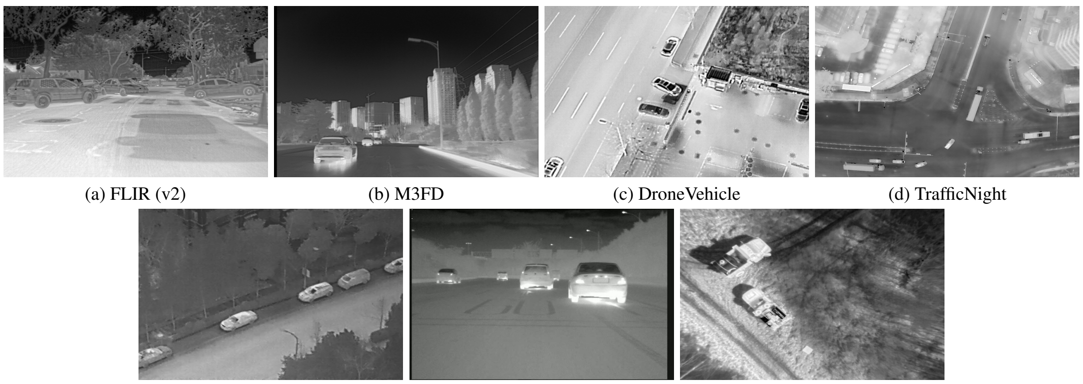
2. Evaluation

This section includes more results from our experiments. Varying factors are the used training data pool, pre-processing methods, fully-finetuned vs. frozen backbone, and DAS vs. default 'vehicle' detector. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. The pre-pocessing method is highlighted with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Fully fine-tuning or frozen backbone where fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

[1] https://github.com/WongKinYiu/yolov9 (accessed 24.02.2025)

Table 1. Key characteristic of prepared datasets used for the experiments. The MCNet [12] and the WIT-UAS [5] datasets are only used for zero-shot evaluation without being added to any training dataset pool.

Dataset	Reference	Dataset Type	Resolution / Pixel	# Images Test	# Images Train	# Categories	# Vehicle Categories (child category names)	# Instances Test	# Instances Train
FLIR	[2]	domain-specific, ground	640 × 512	3493	9824	15	3 ('car', 'bus', 'truck')	30517	73622
M3FD	[6]	domain-specific, ground	1024 × 768	613	2386	6	3 ('car', 'bus', 'truck')	4154	15407
DroneVehicle	[9]	domain-specific, aerial	640 × 512	4465	5495	5	5 ('car', 'freight_car', 'truck', 'bus', 'van')	81629	115701
HIT-UAV	[10]	domain-specific, aerial	640 × 512	272	1013	5	2 ('Car', 'OtherVehicle')	1373	5355
TrafficNight	[13]	domain-specific, aerial	1280 × 1024	366	1759	6	6 ('car', 'buses', 'trucks', 'tractor', ' (empty) + semi-trailers')	5762	29852
WIT-UAS	[5]	domain-specific, aerial	640 × 512	857	✗	2	1 ('vehicle')	2396	✗
MCNet	[12]	domain-specific, ground	720 × 576	665	✗	10	3 ('car', 'truck', 'bus')	2424	✗



(a) FLIR (v2) (b) M3FD (c) DroneVehicle (d) TrafficNight

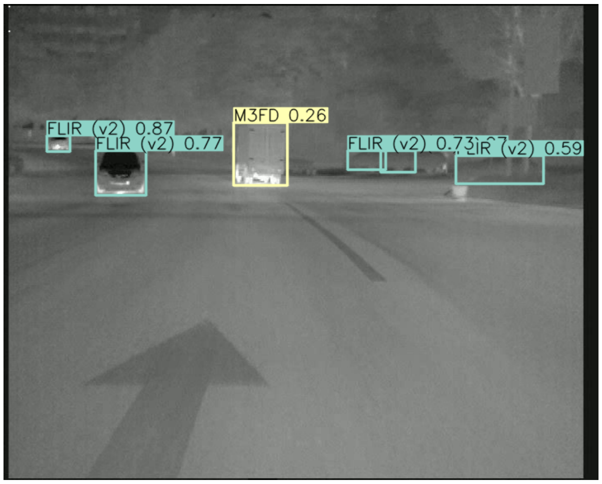
(e) HIT-UAV (f) MCNET (g) WIT-UAS

Figure 1. Example images from the used datasets in our experiments.

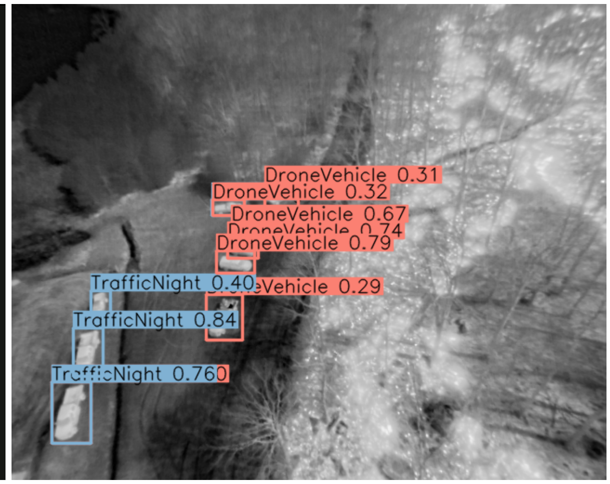
Table 2 includes the results for the FLIR (v2) dataset [2]. Table 3 includes the results for the M3FD dataset [6]. Table 4 includes the results for the DroneVehicle dataset [9]. Table 5 includes the results for the TrafficNight dataset [13]. Table 6 includes the results for the HIT-UAV dataset [10]. Table 7 includes the results for the MCNet [12]. Table 8 includes the results for the WIT-UAS dataset [5].

Figure 2 shows qualitative results on the MCNET and WIT-UAS datasets with using DAS to assign training datasets. The colors of the bounding boxes encode the assigned dataset (color coding: Flir (v2), M3FD, DroneVehicle, TrafficNight, HIT-UAV).

To asses the change in domain gaps after applying different pre-processing methods, we extract CLIP-based [8] image features and project them into a lower-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [4]. Figure visualizes the low-dimensional embedding for the original images and after applying HE, CLAHE or Fieldscale. there is no significant change in the pairwise distances between dataset clusters. It should be noted that due to random sampling of dataset images and afterwards utilizing t-SNE, there is always variation and the cluster position in the low dimensional embedding change.

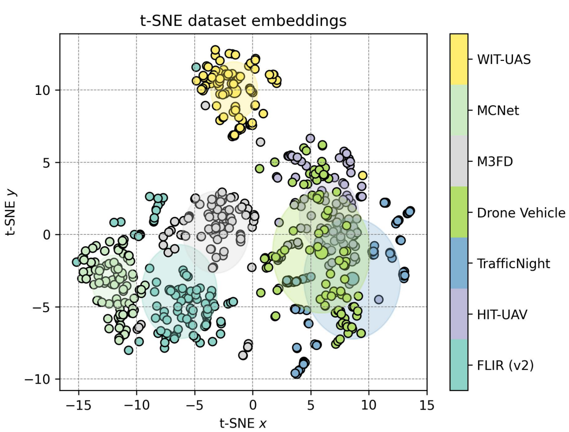


(a) MCNET [12]

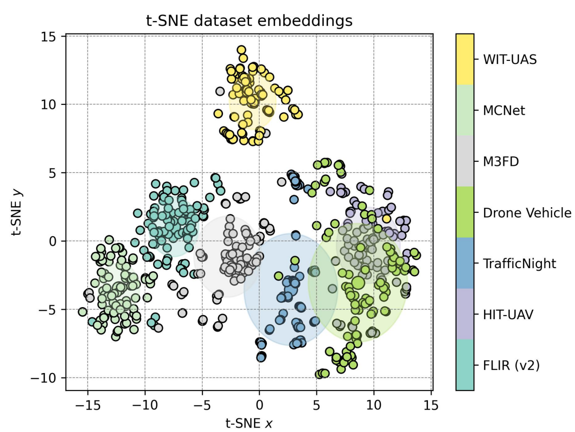


(b) WIT-UAS

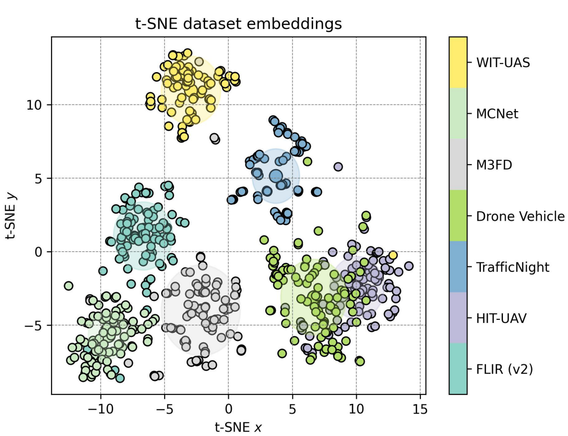
Figure 2. Example results using DAS [1] on the MCNET [12] and WIT-UAS [5] datasets.The color of the bounding boxes encode the assigned origin dataset. (color coding: Flir (v2) , M3FD , DroneVehicle , TrafficNight , HIT-UAV)



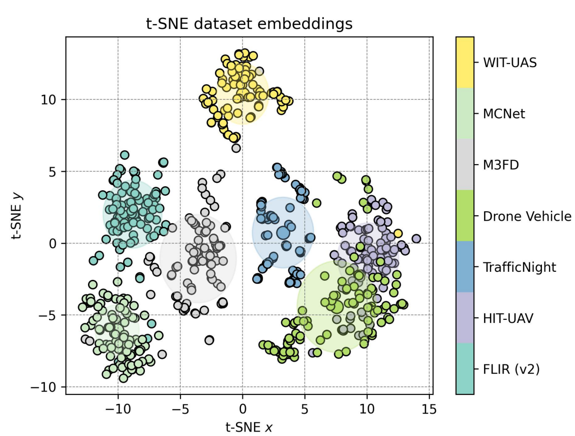
(a) Original images



(b) HE



(c) CLAHE



(d) Fieldscaled

Figure 3. t-SNE plot of CLIP-based feature embedding for all datasets used for the experiments with applying different pre-processing methods (original images, HE, CLAHE, Fieldscale. For every dataset 100 samples are randomly drown. Due to random sampling and applying t-SNE, there is always variation in cluster positions.

Table 2. Results for the FLIR (v2) dataset [2]. Using a pre-pocessing methods is highligtehd with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP ↑ mAP@.5	mAP@.5-.95
FLIR (v2)	✗	❄	✗	✓	✗	✗	✗	✗	0.691	0.462
FLIR (v2)	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.753	0.506
FLIR (v2)	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.763	0.512
FLIR (v2)	HE	🔥	✗	✓	✗	✗	✗	✗	0.749	0.504
FLIR (v2)	✗	🔥	✗	✓	✗	✗	✗	✗	0.771	0.522
FLIR (v2)	✗	❄	✗	✗	✓	✗	✗	✗	0.448	0.290
FLIR (v2)	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.554	0.340
FLIR (v2)	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.559	0.345
FLIR (v2)	HE	🔥	✗	✗	✓	✗	✗	✗	0.514	0.317
FLIR (v2)	✗	🔥	✗	✗	✓	✗	✗	✗	0.538	0.334
FLIR (v2)	✗	❄	✗	✗	✗	✓	✗	✗	0.007	0.003
FLIR (v2)	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.007	0.003
FLIR (v2)	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.002	0.001
FLIR (v2)	HE	🔥	✗	✗	✗	✓	✗	✗	0.008	0.004
FLIR (v2)	✗	🔥	✗	✗	✗	✓	✗	✗	0.009	0.004
FLIR (v2)	✗	❄	✗	✗	✗	✗	✓	✗	0.000	0.000
FLIR (v2)	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.001	0.000
FLIR (v2)	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.001	0.000
FLIR (v2)	HE	🔥	✗	✗	✗	✗	✓	✗	0.001	0.000
FLIR (v2)	✗	🔥	✗	✗	✗	✗	✓	✗	0.001	0.000
FLIR (v2)	✗	❄	✗	✗	✗	✗	✗	✓	0.094	0.049
FLIR (v2)	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.038	0.021
FLIR (v2)	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.048	0.026
FLIR (v2)	HE	🔥	✗	✗	✗	✗	✗	✓	0.069	0.036
FLIR (v2)	✗	🔥	✗	✗	✗	✗	✗	✓	0.104	0.057
FLIR (v2)	✗	❄	✓	✓	✓	✓	✓	✓	0.707	0.465
FLIR (v2)	✗	❄	✗	✓	✓	✓	✓	✓	0.698	0.461
FLIR (v2)	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.742	0.495
FLIR (v2)	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.760	0.507
FLIR (v2)	HE	🔥	✓	✓	✓	✓	✓	✓	0.738	0.494
FLIR (v2)	✗	🔥	✓	✓	✓	✓	✓	✓	0.768	0.515
FLIR (v2)	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.742	0.496
FLIR (v2)	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.759	0.505
FLIR (v2)	HE	🔥	✗	✓	✓	✓	✓	✓	0.742	0.497
FLIR (v2)	✗	🔥	✗	✓	✓	✓	✓	✓	0.759	0.513
FLIR (v2)	✗	❄	✓	✗	✗	✓	✓	✓	0.004	0.001
FLIR (v2)	✗	❄	✗	✗	✗	✓	✓	✓	0.009	0.003
FLIR (v2)	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.013	0.005
FLIR (v2)	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.013	0.010
FLIR (v2)	HE	🔥	✓	✗	✗	✓	✓	✓	0.012	0.004
FLIR (v2)	✗	🔥	✓	✗	✗	✓	✓	✓	0.007	0.002
FLIR (v2)	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.009	0.003
FLIR (v2)	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.017	0.006
FLIR (v2)	HE	🔥	✗	✗	✗	✓	✓	✓	0.014	0.010
FLIR (v2)	✗	🔥	✗	✗	✗	✓	✓	✓	0.014	0.008
FLIR (v2)	✗	❄	✓	✓	✓	✗	✗	✗	0.708	0.464
FLIR (v2)	✗	❄	✗	✓	✓	✗	✗	✗	0.709	0.464
FLIR (v2)	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.762	0.507
FLIR (v2)	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.769	0.515
FLIR (v2)	HE	🔥	✓	✓	✓	✗	✗	✗	0.759	0.504
FLIR (v2)	✗	🔥	✓	✓	✓	✗	✗	✗	0.778	0.523
FLIR (v2)	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.765	0.513
FLIR (v2)	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.773	0.517
FLIR (v2)	HE	🔥	✗	✓	✓	✗	✗	✗	0.753	0.506
FLIR (v2)	✗	🔥	✗	✓	✓	✗	✗	✗	0.783	0.529

Table 3. Results for the M3FD dataset [6]. Using a pre-pocessing methods is highligtehd with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool					mAP ↑	
				FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP@.5	mAP@.5-.95
M3FD	✗	❄	✗	✓	✗	✗	✗	✗	0.676	0.445
M3FD	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.704	0.450
M3FD	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.704	0.450
M3FD	HE	🔥	✗	✓	✗	✗	✗	✗	0.705	0.449
M3FD	✗	🔥	✗	✓	✗	✗	✗	✗	0.706	0.453
M3FD	✗	❄	✗	✗	✓	✗	✗	✗	0.817	0.586
M3FD	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.863	0.632
M3FD	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.862	0.631
M3FD	HE	🔥	✗	✗	✓	✗	✗	✗	0.862	0.632
M3FD	✗	🔥	✗	✗	✓	✗	✗	✗	0.863	0.628
M3FD	✗	❄	✗	✗	✗	✓	✗	✗	0.032	0.011
M3FD	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.038	0.012
M3FD	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.033	0.010
M3FD	HE	🔥	✗	✗	✗	✓	✗	✗	0.044	0.015
M3FD	✗	🔥	✗	✗	✗	✓	✗	✗	0.039	0.014
M3FD	✗	❄	✗	✗	✗	✗	✓	✗	0.002	0.001
M3FD	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.004	0.001
M3FD	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.002	0.001
M3FD	HE	🔥	✗	✗	✗	✗	✓	✗	0.003	0.001
M3FD	✗	🔥	✗	✗	✗	✗	✓	✗	0.002	0.001
M3FD	✗	❄	✗	✗	✗	✗	✗	✓	0.132	0.056
M3FD	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.052	0.023
M3FD	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.050	0.024
M3FD	HE	🔥	✗	✗	✗	✗	✗	✓	0.083	0.038
M3FD	✗	🔥	✗	✗	✗	✗	✗	✓	0.082	0.041
M3FD	✗	❄	✓	✓	✓	✓	✓	✓	0.812	0.578
M3FD	✗	❄	✗	✓	✓	✓	✓	✓	0.798	0.548
M3FD	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.850	0.616
M3FD	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.850	0.614
M3FD	HE	🔥	✓	✓	✓	✓	✓	✓	0.852	0.617
M3FD	✗	🔥	✓	✓	✓	✓	✓	✓	0.851	0.619
M3FD	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.842	0.612
M3FD	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.841	0.610
M3FD	HE	🔥	✗	✓	✓	✓	✓	✓	0.842	0.607
M3FD	✗	🔥	✗	✓	✓	✓	✓	✓	0.850	0.612
M3FD	✗	❄	✓	✗	✗	✓	✓	✓	0.009	0.004
M3FD	✗	❄	✗	✗	✗	✓	✓	✓	0.007	0.003
M3FD	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.019	0.005
M3FD	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.020	0.007
M3FD	HE	🔥	✓	✗	✗	✓	✓	✓	0.011	0.004
M3FD	✗	🔥	✓	✗	✗	✓	✓	✓	0.013	0.005
M3FD	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.019	0.006
M3FD	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.021	0.008
M3FD	HE	🔥	✗	✗	✗	✓	✓	✓	0.026	0.012
M3FD	✗	🔥	✗	✗	✗	✓	✓	✓	0.021	0.008
M3FD	✗	❄	✓	✓	✓	✗	✗	✗	0.841	0.601
M3FD	✗	❄	✗	✓	✓	✗	✗	✗	0.832	0.596
M3FD	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.863	0.631
M3FD	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.862	0.635
M3FD	HE	🔥	✓	✓	✓	✗	✗	✗	0.871	0.633
M3FD	✗	🔥	✓	✓	✓	✗	✗	✗	0.862	0.633
M3FD	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.862	0.632
M3FD	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.862	0.632
M3FD	HE	🔥	✗	✓	✓	✗	✗	✗	0.861	0.628
M3FD	✗	🔥	✗	✓	✓	✗	✗	✗	0.862	0.632

Table 4. Results for the DroneVehicle dataset [9]. Using a pre-pocessing methods is highligtehd with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool					mAP ↑	
				FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP@.5	mAP@.5-.95
DroneVehicle	✗	❄	✗	✓	✗	✗	✗	✗	0.003	0.002
DroneVehicle	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.003	0.001
DroneVehicle	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.010	0.009
DroneVehicle	HE	🔥	✗	✓	✗	✗	✗	✗	0.006	0.004
DroneVehicle	✗	🔥	✗	✓	✗	✗	✗	✗	0.010	0.008
DroneVehicle	✗	❄	✗	✗	✓	✗	✗	✗	0.003	0.001
DroneVehicle	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.009	0.007
DroneVehicle	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.010	0.009
DroneVehicle	HE	🔥	✗	✗	✓	✗	✗	✗	0.002	0.001
DroneVehicle	✗	🔥	✗	✗	✓	✗	✗	✗	0.007	0.004
DroneVehicle	✗	❄	✗	✗	✗	✓	✗	✗	0.911	0.596
DroneVehicle	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.913	0.601
DroneVehicle	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.913	0.605
DroneVehicle	HE	🔥	✗	✗	✗	✓	✗	✗	0.913	0.604
DroneVehicle	✗	🔥	✗	✗	✗	✓	✗	✗	0.912	0.601
DroneVehicle	✗	❄	✗	✗	✗	✗	✓	✗	0.339	0.179
DroneVehicle	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.250	0.141
DroneVehicle	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.259	0.150
DroneVehicle	HE	🔥	✗	✗	✗	✗	✓	✗	0.265	0.148
DroneVehicle	✗	🔥	✗	✗	✗	✗	✓	✗	0.271	0.147
DroneVehicle	✗	❄	✗	✗	✗	✗	✗	✓	0.727	0.443
DroneVehicle	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.678	0.404
DroneVehicle	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.678	0.403
DroneVehicle	HE	🔥	✗	✗	✗	✗	✗	✓	0.678	0.402
DroneVehicle	✗	🔥	✗	✗	✗	✗	✗	✓	0.699	0.421
DroneVehicle	✗	❄	✓	✓	✓	✓	✓	✓	0.900	0.590
DroneVehicle	✗	❄	✗	✓	✓	✓	✓	✓	0.911	0.594
DroneVehicle	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.911	0.598
DroneVehicle	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.910	0.598
DroneVehicle	HE	🔥	✓	✓	✓	✓	✓	✓	0.911	0.597
DroneVehicle	✗	🔥	✓	✓	✓	✓	✓	✓	0.912	0.598
DroneVehicle	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.912	0.601
DroneVehicle	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.912	0.602
DroneVehicle	HE	🔥	✗	✓	✓	✓	✓	✓	0.912	0.601
DroneVehicle	✗	🔥	✗	✓	✓	✓	✓	✓	0.912	0.601
DroneVehicle	✗	❄	✓	✗	✗	✓	✓	✓	0.901	0.592
DroneVehicle	✗	❄	✗	✗	✗	✓	✓	✓	0.902	0.593
DroneVehicle	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.912	0.599
DroneVehicle	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.913	0.601
DroneVehicle	HE	🔥	✓	✗	✗	✓	✓	✓	0.910	0.595
DroneVehicle	✗	🔥	✓	✗	✗	✓	✓	✓	0.911	0.599
DroneVehicle	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.913	0.602
DroneVehicle	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.913	0.602
DroneVehicle	HE	🔥	✗	✗	✗	✓	✓	✓	0.914	0.600
DroneVehicle	✗	🔥	✗	✗	✗	✓	✓	✓	0.913	0.602
DroneVehicle	✗	❄	✓	✓	✓	✗	✗	✗	0.010	0.005
DroneVehicle	✗	❄	✗	✓	✓	✗	✗	✗	0.010	0.006
DroneVehicle	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.005	0.002
DroneVehicle	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.010	0.007
DroneVehicle	HE	🔥	✓	✓	✓	✗	✗	✗	0.004	0.002
DroneVehicle	✗	🔥	✓	✓	✓	✗	✗	✗	0.004	0.002
DroneVehicle	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.005	0.002
DroneVehicle	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.010	0.009
DroneVehicle	HE	🔥	✗	✓	✓	✗	✗	✗	0.005	0.003
DroneVehicle	✗	🔥	✗	✓	✓	✗	✗	✗	0.004	0.002

Table 5. Results for the TrafficNight dataset [13]. Using a pre-pocessing methods is highligtehd with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool					mAP ↑	
				FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP@.5	mAP@.5-.95
TrafficNight	✗	❄	✗	✓	✗	✗	✗	✗	0.016	0.010
TrafficNight	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.008	0.004
TrafficNight	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.010	0.005
TrafficNight	HE	🔥	✗	✓	✗	✗	✗	✗	0.002	0.001
TrafficNight	✗	🔥	✗	✓	✗	✗	✗	✗	0.001	0.001
TrafficNight	✗	❄	✗	✗	✓	✗	✗	✗	0.017	0.011
TrafficNight	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.010	0.006
TrafficNight	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.003	0.002
TrafficNight	HE	🔥	✗	✗	✓	✗	✗	✗	0.000	0.000
TrafficNight	✗	🔥	✗	✗	✓	✗	✗	✗	0.010	0.007
TrafficNight	✗	❄	✗	✗	✗	✓	✗	✗	0.614	0.299
TrafficNight	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.501	0.259
TrafficNight	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.550	0.287
TrafficNight	HE	🔥	✗	✗	✗	✓	✗	✗	0.674	0.345
TrafficNight	✗	🔥	✗	✗	✗	✓	✗	✗	0.526	0.282
TrafficNight	✗	❄	✗	✗	✗	✗	✓	✗	0.865	0.549
TrafficNight	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.859	0.528
TrafficNight	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.861	0.550
TrafficNight	HE	🔥	✗	✗	✗	✗	✓	✗	0.897	0.545
TrafficNight	✗	🔥	✗	✗	✗	✗	✓	✗	0.897	0.564
TrafficNight	✗	❄	✗	✗	✗	✗	✗	✓	0.494	0.276
TrafficNight	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.519	0.290
TrafficNight	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.455	0.254
TrafficNight	HE	🔥	✗	✗	✗	✗	✗	✓	0.494	0.275
TrafficNight	✗	🔥	✗	✗	✗	✗	✗	✓	0.479	0.261
TrafficNight	✗	❄	✓	✓	✓	✓	✓	✓	0.816	0.518
TrafficNight	✗	❄	✗	✓	✓	✓	✓	✓	0.843	0.517
TrafficNight	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.869	0.533
TrafficNight	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.885	0.550
TrafficNight	HE	🔥	✓	✓	✓	✓	✓	✓	0.883	0.542
TrafficNight	✗	🔥	✓	✓	✓	✓	✓	✓	0.890	0.557
TrafficNight	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.885	0.556
TrafficNight	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.874	0.531
TrafficNight	HE	🔥	✗	✓	✓	✓	✓	✓	0.876	0.545
TrafficNight	✗	🔥	✗	✓	✓	✓	✓	✓	0.903	0.563
TrafficNight	✗	❄	✓	✗	✗	✓	✓	✓	0.881	0.530
TrafficNight	✗	❄	✗	✗	✗	✓	✓	✓	0.895	0.534
TrafficNight	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.865	0.546
TrafficNight	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.872	0.554
TrafficNight	HE	🔥	✓	✗	✗	✓	✓	✓	0.853	0.532
TrafficNight	✗	🔥	✓	✗	✗	✓	✓	✓	0.873	0.544
TrafficNight	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.860	0.548
TrafficNight	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.883	0.552
TrafficNight	HE	🔥	✗	✗	✗	✓	✓	✓	0.874	0.540
TrafficNight	✗	🔥	✗	✗	✗	✓	✓	✓	0.886	0.566
TrafficNight	✗	❄	✓	✓	✓	✗	✗	✗	0.010	0.005
TrafficNight	✗	❄	✗	✓	✓	✗	✗	✗	0.010	0.005
TrafficNight	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.010	0.004
TrafficNight	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.003	0.002
TrafficNight	HE	🔥	✓	✓	✓	✗	✗	✗	0.002	0.001
TrafficNight	✗	🔥	✓	✓	✓	✗	✗	✗	0.002	0.001
TrafficNight	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.008	0.004
TrafficNight	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.027	0.011
TrafficNight	HE	🔥	✗	✓	✓	✗	✗	✗	0.026	0.013
TrafficNight	✗	🔥	✗	✓	✓	✗	✗	✗	0.007	0.004

Table 6. Results for the HIT-UAV dataset [10]. Using a pre-pocessing methods is highlighted with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP ↑ mAP@.5	mAP@.5-.95
HIT-UAV	✗	❄	✗	✓	✗	✗	✗	✗	0.134	0.044
HIT-UAV	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.013	0.010
HIT-UAV	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.007	0.005
HIT-UAV	HE	🔥	✗	✓	✗	✗	✗	✗	0.019	0.010
HIT-UAV	✗	🔥	✗	✓	✗	✗	✗	✗	0.018	0.009
HIT-UAV	✗	❄	✗	✗	✓	✗	✗	✗	0.032	0.016
HIT-UAV	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.026	0.012
HIT-UAV	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.014	0.007
HIT-UAV	HE	🔥	✗	✗	✓	✗	✗	✗	0.017	0.012
HIT-UAV	✗	🔥	✗	✗	✓	✗	✗	✗	0.027	0.014
HIT-UAV	✗	❄	✗	✗	✗	✗	✓	✗	0.096	0.048
HIT-UAV	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.217	0.128
HIT-UAV	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.201	0.125
HIT-UAV	HE	🔥	✗	✗	✗	✗	✓	✗	0.246	0.149
HIT-UAV	✗	🔥	✗	✗	✗	✗	✓	✗	0.241	0.144
HIT-UAV	✗	❄	✗	✗	✗	✓	✗	✗	0.783	0.486
HIT-UAV	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.850	0.534
HIT-UAV	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.840	0.542
HIT-UAV	HE	🔥	✗	✗	✗	✓	✗	✗	0.836	0.545
HIT-UAV	✗	🔥	✗	✗	✗	✓	✗	✗	0.850	0.554
HIT-UAV	✗	❄	✗	✗	✗	✗	✗	✓	0.953	0.649
HIT-UAV	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.934	0.680
HIT-UAV	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.925	0.679
HIT-UAV	HE	🔥	✗	✗	✗	✗	✗	✓	0.944	0.678
HIT-UAV	✗	🔥	✗	✗	✗	✗	✗	✓	0.924	0.669
HIT-UAV	✗	❄	✓	✓	✓	✓	✓	✓	0.952	0.683
HIT-UAV	✗	❄	✗	✓	✓	✓	✓	✓	0.953	0.682
HIT-UAV	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.963	0.700
HIT-UAV	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.963	0.702
HIT-UAV	HE	🔥	✓	✓	✓	✓	✓	✓	0.951	0.699
HIT-UAV	✗	🔥	✓	✓	✓	✓	✓	✓	0.953	0.699
HIT-UAV	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.962	0.694
HIT-UAV	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.963	0.699
HIT-UAV	HE	🔥	✗	✓	✓	✓	✓	✓	0.963	0.699
HIT-UAV	✗	🔥	✗	✓	✓	✓	✓	✓	0.962	0.703
HIT-UAV	✗	❄	✓	✗	✗	✓	✓	✓	0.952	0.691
HIT-UAV	✗	❄	✗	✗	✗	✓	✓	✓	0.954	0.687
HIT-UAV	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.964	0.703
HIT-UAV	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.961	0.705
HIT-UAV	HE	🔥	✓	✗	✗	✓	✓	✓	0.963	0.702
HIT-UAV	✗	🔥	✓	✗	✗	✓	✓	✓	0.953	0.699
HIT-UAV	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.964	0.695
HIT-UAV	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.964	0.698
HIT-UAV	HE	🔥	✗	✗	✗	✓	✓	✓	0.963	0.702
HIT-UAV	✗	🔥	✗	✗	✗	✓	✓	✓	0.963	0.696
HIT-UAV	✗	❄	✓	✓	✓	✗	✗	✗	0.027	0.011
HIT-UAV	✗	❄	✗	✓	✓	✗	✗	✗	0.029	0.010
HIT-UAV	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.018	0.009
HIT-UAV	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.010	0.005
HIT-UAV	HE	🔥	✓	✓	✓	✗	✗	✗	0.020	0.007
HIT-UAV	✗	🔥	✓	✓	✓	✗	✗	✗	0.027	0.013
HIT-UAV	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.010	0.008
HIT-UAV	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.010	0.008
HIT-UAV	HE	🔥	✗	✓	✓	✗	✗	✗	0.005	0.003
HIT-UAV	✗	🔥	✗	✓	✓	✗	✗	✗	0.011	0.004

Table 7. Results for the MSNet dataset [12]. Using a pre-pocessing methods is highlighted with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool					mAP ↑	
				FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP@.5	mAP@.5-.95
MCNet	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.786	0.649
MCNet	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.792	0.654
MCNet	HE	🔥	✗	✓	✗	✗	✗	✗	0.785	0.645
MCNet	✗	🔥	✗	✓	✗	✗	✗	✗	0.786	0.656
MCNet	✗	❄	✗	✓	✗	✗	✗	✗	0.732	0.599
MCNet	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.856	0.639
MCNet	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.858	0.648
MCNet	HE	🔥	✗	✗	✓	✗	✗	✗	0.844	0.621
MCNet	✗	🔥	✗	✗	✓	✗	✗	✗	0.845	0.626
MCNet	✗	❄	✗	✗	✓	✗	✗	✗	0.693	0.497
MCNet	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.017	0.006
MCNet	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.018	0.007
MCNet	HE	🔥	✗	✗	✗	✗	✓	✗	0.008	0.003
MCNet	✗	🔥	✗	✗	✗	✗	✓	✗	0.012	0.005
MCNet	✗	❄	✗	✗	✗	✗	✓	✗	0.011	0.004
MCNet	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.039	0.011
MCNet	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.056	0.015
MCNet	HE	🔥	✗	✗	✗	✓	✗	✗	0.033	0.011
MCNet	✗	🔥	✗	✗	✗	✓	✗	✗	0.030	0.011
MCNet	✗	❄	✗	✗	✗	✓	✗	✗	0.002	0.001
MCNet	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.149	0.077
MCNet	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.150	0.079
MCNet	HE	🔥	✗	✗	✗	✗	✗	✓	0.149	0.081
MCNet	✗	🔥	✗	✗	✗	✗	✗	✓	0.180	0.112
MCNet	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.851	0.663
MCNet	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.835	0.664
MCNet	HE	🔥	✓	✓	✓	✓	✓	✓	0.815	0.640
MCNet	✗	🔥	✓	✓	✓	✓	✓	✓	0.838	0.663
MCNet	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.830	0.666
MCNet	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.828	0.675
MCNet	HE	🔥	✗	✓	✓	✓	✓	✓	0.813	0.657
MCNet	✗	🔥	✗	✓	✓	✓	✓	✓	0.831	0.681
MCNet	✗	❄	✓	✓	✓	✓	✓	✓	0.763	0.615
MCNet	✗	❄	✗	✓	✓	✓	✓	✓	0.747	0.605
MCNet	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.018	0.008
MCNet	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.019	0.008
MCNet	HE	🔥	✓	✗	✗	✓	✓	✓	0.014	0.006
MCNet	✗	🔥	✓	✗	✗	✓	✓	✓	0.017	0.007
MCNet	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.036	0.012
MCNet	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.029	0.011
MCNet	HE	🔥	✗	✗	✗	✓	✓	✓	0.022	0.007
MCNet	✗	🔥	✗	✗	✗	✓	✓	✓	0.046	0.020
MCNet	✗	❄	✓	✗	✗	✓	✓	✓	0.003	0.001
MCNet	✗	❄	✗	✗	✗	✓	✓	✓	0.002	0.001
MCNet	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.847	0.649
MCNet	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.844	0.658
MCNet	HE	🔥	✓	✓	✓	✗	✗	✗	0.816	0.644
MCNet	✗	🔥	✓	✓	✓	✗	✗	✗	0.841	0.673
MCNet	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.822	0.662
MCNet	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.829	0.668
MCNet	HE	🔥	✗	✓	✓	✗	✗	✗	0.810	0.649
MCNet	✗	🔥	✗	✓	✓	✗	✗	✗	0.830	0.672
MCNet	✗	❄	✓	✓	✓	✗	✗	✗	0.744	0.604
MCNet	✗	❄	✗	✓	✓	✗	✗	✗	0.760	0.616

Table 8. Results for the WIT-UAS dataset [5]. Using a pre-pocessing methods is highlighted with the corresponding abbreviation (HE, CLAHE, Fieldscale) or using the original images is marked with ✗. Datasets used in the pool are marked with ✓ and excluded datasets are marked with ✗. Fully fine-tuned is highlighted with a fire icon 🔥 and a frozen backbone is highlighted with a snowflake icon ❄. Using DAS is highlighted with an ✓ icon in the corresponding column and standard models are marked with ✗.

Test Dataset	Pre-processing {HE, CLAHE, FieldScale }	Backbone	DAS	Training Dataset Pool					mAP ↑	
				FLIR (v2)	M3FD	DroneVehicle	TrafficNight	HIT-UAV	mAP@.5	mAP@.5-.95
WIT-UAS	CLAHE	🔥	✗	✓	✗	✗	✗	✗	0.016	0.010
WIT-UAS	FieldScale	🔥	✗	✓	✗	✗	✗	✗	0.029	0.017
WIT-UAS	HE	🔥	✗	✓	✗	✗	✗	✗	0.041	0.015
WIT-UAS	✗	🔥	✗	✓	✗	✗	✗	✗	0.005	0.002
WIT-UAS	✗	❄	✗	✓	✗	✗	✗	✗	0.002	0.001
WIT-UAS	CLAHE	🔥	✗	✗	✓	✗	✗	✗	0.002	0.001
WIT-UAS	FieldScale	🔥	✗	✗	✓	✗	✗	✗	0.005	0.002
WIT-UAS	HE	🔥	✗	✗	✓	✗	✗	✗	0.000	0.000
WIT-UAS	✗	🔥	✗	✗	✓	✗	✗	✗	0.000	0.000
WIT-UAS	✗	❄	✗	✗	✓	✗	✗	✗	0.000	0.000
WIT-UAS	CLAHE	🔥	✗	✗	✗	✗	✓	✗	0.368	0.133
WIT-UAS	FieldScale	🔥	✗	✗	✗	✗	✓	✗	0.392	0.150
WIT-UAS	HE	🔥	✗	✗	✗	✗	✓	✗	0.388	0.140
WIT-UAS	✗	🔥	✗	✗	✗	✗	✓	✗	0.388	0.143
WIT-UAS	✗	❄	✗	✗	✗	✗	✓	✗	0.375	0.114
WIT-UAS	CLAHE	🔥	✗	✗	✗	✓	✗	✗	0.509	0.250
WIT-UAS	FieldScale	🔥	✗	✗	✗	✓	✗	✗	0.527	0.252
WIT-UAS	HE	🔥	✗	✗	✗	✓	✗	✗	0.534	0.231
WIT-UAS	✗	🔥	✗	✗	✗	✓	✗	✗	0.445	0.219
WIT-UAS	✗	❄	✗	✗	✗	✓	✗	✗	0.448	0.191
WIT-UAS	CLAHE	🔥	✗	✗	✗	✗	✗	✓	0.383	0.175
WIT-UAS	FieldScale	🔥	✗	✗	✗	✗	✗	✓	0.411	0.170
WIT-UAS	HE	🔥	✗	✗	✗	✗	✗	✓	0.448	0.187
WIT-UAS	✗	🔥	✗	✗	✗	✗	✗	✓	0.461	0.202
WIT-UAS	CLAHE	🔥	✓	✓	✓	✓	✓	✓	0.470	0.199
WIT-UAS	FieldScale	🔥	✓	✓	✓	✓	✓	✓	0.393	0.168
WIT-UAS	HE	🔥	✓	✓	✓	✓	✓	✓	0.503	0.196
WIT-UAS	✗	🔥	✓	✓	✓	✓	✓	✓	0.483	0.175
WIT-UAS	CLAHE	🔥	✗	✓	✓	✓	✓	✓	0.498	0.202
WIT-UAS	FieldScale	🔥	✗	✓	✓	✓	✓	✓	0.467	0.212
WIT-UAS	HE	🔥	✗	✓	✓	✓	✓	✓	0.535	0.223
WIT-UAS	✗	🔥	✗	✓	✓	✓	✓	✓	0.547	0.217
WIT-UAS	✗	❄	✓	✓	✓	✓	✓	✓	0.440	0.166
WIT-UAS	✗	❄	✗	✓	✓	✓	✓	✓	0.502	0.188
WIT-UAS	CLAHE	🔥	✓	✗	✗	✓	✓	✓	0.501	0.205
WIT-UAS	FieldScale	🔥	✓	✗	✗	✓	✓	✓	0.464	0.193
WIT-UAS	HE	🔥	✓	✗	✗	✓	✓	✓	0.528	0.205
WIT-UAS	✗	🔥	✓	✗	✗	✓	✓	✓	0.467	0.173
WIT-UAS	CLAHE	🔥	✗	✗	✗	✓	✓	✓	0.576	0.227
WIT-UAS	FieldScale	🔥	✗	✗	✗	✓	✓	✓	0.529	0.226
WIT-UAS	HE	🔥	✗	✗	✗	✓	✓	✓	0.580	0.236
WIT-UAS	✗	🔥	✗	✗	✗	✓	✓	✓	0.526	0.214
WIT-UAS	✗	❄	✓	✗	✗	✓	✓	✓	0.445	0.155
WIT-UAS	✗	❄	✗	✗	✗	✓	✓	✓	0.503	0.179
WIT-UAS	CLAHE	🔥	✓	✓	✓	✗	✗	✗	0.010	0.006
WIT-UAS	FieldScale	🔥	✓	✓	✓	✗	✗	✗	0.010	0.007
WIT-UAS	HE	🔥	✓	✓	✓	✗	✗	✗	0.010	0.007
WIT-UAS	✗	🔥	✓	✓	✓	✗	✗	✗	0.013	0.005
WIT-UAS	CLAHE	🔥	✗	✓	✓	✗	✗	✗	0.010	0.005
WIT-UAS	FieldScale	🔥	✗	✓	✓	✗	✗	✗	0.010	0.007
WIT-UAS	HE	🔥	✗	✓	✓	✗	✗	✗	0.007	0.002
WIT-UAS	✗	🔥	✗	✓	✓	✗	✗	✗	0.001	0.001
WIT-UAS	✗	❄	✓	✓	✓	✗	✗	✗	0.002	0.002
WIT-UAS	✗	❄	✗	✓	✓	✗	✗	✗	0.003	0.001

References

[1] Stefan Becker, Jens Bayer, Ronny Hug, Wolfgang Huebner, and Michael Arens. Utilizing dataset affinity prediction in object detection to assess training data. In *Robotics, Computer Vision and Intelligent Systems*, pages 260–277, Cham, 2024. Springer Nature Switzerland. 1, 3

[2] FLIR (V2). Free flir thermal dataset for algorithm training. 2022. *Available at* `https://www.flir.com/oem/adas/dataset/european-regional-thermal-dataset/`. 1, 2, 4

[3] Hyeonjae Gil, Myung-Hwan Jeon, and Ayoung Kim. Fieldscale: Locality-aware field-based adaptive rescaling for thermal infrared image. *IEEE Robotics and Automation Letters*, 2024. 1

[4] Geoffrey Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15:833–840, 2003. 2

[5] Andrew Jong, Mukai Yu, Devansh Dhrafani, Siva Kailas, Brady Moon, Katia Sycara, and Sebastian Scherer. Wituas: A wildland-fire infrared thermal dataset to detect crew assets from aerial views. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11464–11471, 2023. 1, 2, 3, 10

[6] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5792–5801, 2022. 1, 2, 5

[7] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987. 1

[8] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 2

[9] Yiming Sun, Bing Cao, Pengfei Zhu, and Qinghua Hu. Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2022. 1, 2, 6

[10] Jiashun Suo, Tianyi Wang, Xingzhou Zhang, Haiyang Chen, Wei Zhou, and Weisong Shi. HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection. *Scientific Data*, 10:227, 2023. 1, 2, 8

[11] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. In *European Confernce on Computer Vision (ECCV) 2024*, pages 1–21, Cham, 2025. Springer Nature Switzerland. 1

[12] Haitao Xiong, Wenjie Cai, and Qiong Liu. Mcnet: Multi-level correction network for thermal image semantic segmentation of nighttime driving scene. *Infrared Physics and Technology*, page 103628, 2021. 1, 2, 3, 9

[13] Guoxing Zhang, Yiming Liu, Xiaoyu Yang, Chao Huang, and Hailong Huang. TrafficNight: An aerial multimodal benchmark for nighttime vehicle surveillance. 2024. 1, 2, 7