Supplementary Material

Tamara R. Lenhard[1,2] Andreas Weinmann[2] Kai Franke[1] Tobias Koch[1]

[1]Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR), Germany
[2]Working Group Algorithms for Computer Vision, Imaging and Data Analysis,
University of Applied Sciences Darmstadt, Germany

{tamara.lenhard, kai.franke, tobias.koch}@dlr.de, andreas.weinmann@h-da.de

A. Details on Drone Detection Datasets

In the field of image-based drone detection, diverse datasets have been established, each defined by specific attributes and features tailored to distinct application contexts and objectives (see also Table I). Complementing the information in Section 2.2 (main paper), Table I and the subsequent sections offer a more comprehensive exploration of the distinctive properties of each dataset.

USC Drone Detection and Tracking. The USC Drone Detection and Tracking dataset [6,25] consists of 30 videos, each with a resolution of 1920×1080 pixels. Recorded at a frame rate of 15 FPS and an approximate duration of one minute per video, the dataset contains approximately 27,000 images. The videos were captured on the University of Southern California (USC) campus, featuring a wide variety of backgrounds, camera angles, drone appearances, and diverse weather and lighting conditions. Only a single drone model (DJI Phantom) was used for data generation.

Drone Dataset by [2]. The Drone Dataset, provided by Aksoy *et al.* [2], comprises approximately 4,000 annotated RGB images sourced from YouTube drone videos and Google image searches. These images exhibit a resolution range from 300×168 to 3840×2160 pixels (4K) and exclusively feature DJI Phantom drones. The dataset also includes images of various non-drone objects.

MAV-VID. The Multirotor Aerial Vehicle VID (MAV-VID) dataset by Rodriguez-Ramos *et al.* [21] comprises 64 videos (i.e., 40,232 images) of single drones. The videos are captured in various setups using different recording techniques, including handheld mobile devices, ground-based surveillance cameras, and other drones [15]. The average drone size within the dataset is 136×77 pixels.

Det-Fly. The Det-Fly dataset by Zheng *et al.* [28] focuses on air-to-air visual detection of micro UAVs and comprises 13,271 high-resolution images (3840×2160 pixels). The images, captured by another UAV, were sourced from videos at a sampling rate of 5 FPS or taken from selected positions. They feature diverse environmental backgrounds (sky, urban, field, mountain) and perspectives (front, top, bottom) based on relative viewing angles. Despite the considerable variability in drone size, with nearly half of the drones occupying less than 5% of the total image area, the dataset exclusively covers a single drone model (DJI Mavic).

UAV-Eagle. The UAV-Eagle dataset [3] is designed to evaluate the effectiveness of drone detection algorithms under varying conditions, including diverse illumination settings, motion artifacts, and viewpoint alterations. It comprises 510 annotated images featuring complex environments characterized by diverse background objects (e.g., trees, buildings, clouds, vehicles, and people). Employing a UAV-mounted camera for data collection, the dataset includes aerial images of both single- and multi-drone scenarios; however, limited to the Eagle quadcopter model.

UAVData. Zeng *et al.* [26] introduce UAVData, a dataset designed for visual drone detection, consisting of 13,803 manually recorded and annotated RGB images with a resolution of 1280×720 pixels. The UAVData dataset captures a diverse array of real-world environments, encompassing both indoor settings (e.g., workshops and laboratories) and outdoor scenes featuring distinct background compositions (e.g., sky, trees, and buildings). This dataset aims to address the challenges inherent in real-world scenarios by incorporating rapid illumination changes, complex scenarios, and blurring effects caused by high-speed motion. In addition to six common drone models, UAVData includes balloon distractors, thus yielding 7,320 uni-drone images, 4,346 multi-drone images, and 2,137 balloon images. Drone sizes within the images range from 5×23 to 720×303 pixels.

Halmstad Data. The Halmstad Dataset, developed by Svanström *et al.* [24], is a multi-sensor dataset for drone de-

Table I. Overview of additional characteristics of publicly available datasets for image-based drone detection. The symbols ✗ (does not apply) and ✓ (applies) indicate the designated computer vision (CV) task – detection (detect) and / or tracking (track) – the represented drone types (multicopter and / or fixed-wing), and the camera configurations (static and / or moving).

Dataset	CV Task		Objective	Drone Type		Camera Config.	
	detect	track		multicopter	fixed-wing	static	moving
USC Drone Detect. & Track. [6, 25]	✓	✓	drone monitoring	✓	✗	✓	✓
Drone Dataset [2]	✓	✗	drone detection	✓	✗	–	–
MAV-VID [21]	✓	✓	drone detection	✓	✗	✓	✓
Det-Fly [28]	✓	✗	detection of micro-UAVs	✓	✗	–	–
UAV-Eagle [3]	✓	✓	UAV detection in unconstrained environments	✓	✗	✓	✗
UAVData [26]	✓	✓	UAV detection	✓	✗	✓	✗
Halmstadt Data [24]	✓	✓	drone detection at airports	✓	✗	✓	✓
DUT Anti-UAV [27]	✓	✓	anti-UAV detection	✓	✗	✓	✓
Malicious Drones [16]	✓	✗	hazardous payload drone detection	✓	✗	–	–
VisioDECT [1]	✓	✗	detection of unauthorized UAVs	✓	✗	✓	✓
S-UAV-T [4] *(synthetic)*	✓	✗	UAV-to-UAV detection	✓	✗	✗	✓
Drone-vs-Bird Detection Ch. [7]	✓	✓	distinction between drones and birds	✓	✓	✓	✓
Anti-UAV [17]	✗	✓	single UAV tracking	✓	✗	✓	✓
SynDroneVision (Ours, *synthetic*)	✓	(✓)	drone detection	✓	✗	✓	✗

tection, with a specific focus on detecting small UAVs. The dataset comprises 365 infrared (IR) and 285 visible light (RGB) videos, each lasting 10 seconds, alongside audio files. These recordings were primarily captured at airports in Sweden (e.g., Halmstad Airport) under daylight conditions. The dataset encompasses a variety of drone models (including the Hubsan H107D, DJI Flame Wheel, and DJI Phantom 4), as well as potential drone-like objects such as birds and airplanes. In total, the dataset comprises 203,328 annotated frames (across both IR and RGB), categorizing objects into the classes drone, bird, airplane, and helicopter. However, the .mat format annotations are not directly compatible with most DL frameworks.

DUT Anti-UAV. The Dalian University of Technology (DUT) Anti-UAV dataset [27] consists of two subsets: one for detection and one for tracking. The detection dataset includes 10,000 images, partially recorded in a sequential manner. The image resolutions vary significantly, ranging from 240×160 to 5616×3744 pixels. Object sizes within the images also exhibit substantial variation, with an average object area ratio of 0.013, indicating a high proportion of small objects. DUT Anti-UAV features 35 different UAV types for data generation and is characterized by a high diversity of scene information. It includes various outdoor environments such as the sky, dark clouds, jungles, high-rise buildings, residential buildings, farmland, and playgrounds. Additionally, it encompasses diverse lighting settings (day, night, dawn, and dusk) and weather conditions (sunny, cloudy, and snowy days). In terms of object positioning, the majority of drones are located in the central area of the image.

Malicious Drones. Jamil *et al.* [16] introduce the Malicious Drones dataset, specifically designed for detecting harmful drones (e.g., carrying hazardous payloads) and differentiating them from other aerial entities. The dataset comprises 776 images categorized into five classes: aeroplane, bird, drone, helicopter, and malicious drone, with drones (normal and malicious) accounting for approximately half of the dataset (∼ 399 images). All images are standardized to a resolution of 224×224 pixels. The dataset aims to address the complexity of real-world scenarios by including scenarios characterized by low illumination, reduced object visibility, occlusions, and adverse weather conditions.

VisioDECT. The VisioDECT dataset [1] is a specialized aerial dataset designed for scenario-based detection of unauthorized drones. It comprises 20,924 annotated RGB images (852×480 pixels) recorded across three distinct scenarios: cloudy, sunny, and evening. The images were captured at varying altitudes and locations, at different times, and under diverse climatic conditions, using six distinct drone models: Anafi Extended, DJI FPV, DJI Phantom, EFT-E410S, Mavic2-Air, and Mavic2-Enterprise Zoom. The collected data was manually cleaned (excluding images without drones) and annotated by domain experts.

S-UAV-T. The S-UAV-T dataset by Barisic et al. [4] is the only publicly available synthetic dataset for drone – more precisely UAV-to-UAV – detection. The dataset is generated via Blender [5] and the rendering engine Cycles [9], with a particular emphasis on texture randomization. To reflect the diversity of real-world environments, the dataset includes variations in drone models, the quantity of drones per image, lighting conditions (daylight, partly cloudy, twilight), object scales, camera positions and angles, as well as a range of unconventional textures. The dataset comprises 52,500 drone images with a resolution of 608×608 pixels.

Drone-vs-Bird Detection Challenge. The Drone-vs-Bird Detection Challenge dataset [7] is a comprehensive, manually annotated collection designed to assist in accurately distinguishing drones from birds across a wide range of conditions. It comprises 77 video sequences for training, each averaging 1,384 frames, captured with both static and moving cameras at resolutions from 720×576 to 3840×2160 pixels. The dataset includes eight types of commercial drones - three fixed-wing and five rotary-wing models - recorded in diverse environments such as urban areas, woodlands, agricultural fields, and maritime regions across Central Europe and the Mediterranean. These videos feature different weather conditions and times of day, introducing challenges like direct sun glare and varying camera characteristics. While drones are annotated, birds, which frequently appear as primary disturbance, are not. Drone sizes range from 15 pixels to over 1,000,000 pixels, with most annotated drones being smaller than 32×32 pixels. The test set, comprises 30 additional video sequences without annotations, featuring new backgrounds, additional drone types, and other disturbing objects like planes.

Anti-UAV. The Anti-UAV dataset, created by Jiang *et al.* [17], comprises 318 pairs of real RGB-T video sequences tailored for UAV tracking. Each pair features both RGB and thermal IR modalities, capturing a broad spectrum of lighting conditions (day and night) and diverse background compositions (e.g., buildings, clouds, or trees). Furthermore, the dataset includes prominent UAV models – specifically the DJI Inspire, DJI Phantom 4, DJI Marvic Air, DJI Marvic Pro, DJI Spark, and Parrot. Similar to the DUT Anti-UAV dataset, the majority of drones in the Anti-UAV dataset are positioned centrally within the image frames. A comprehensive three-stage annotation process was used to generate precise annotations. The dataset does not specify a version explicitly dedicated to object detection.

B. Dataset Structure

The dataset is structured into two main folders: *images* and *labels*. Each folders is further divided into training, test,

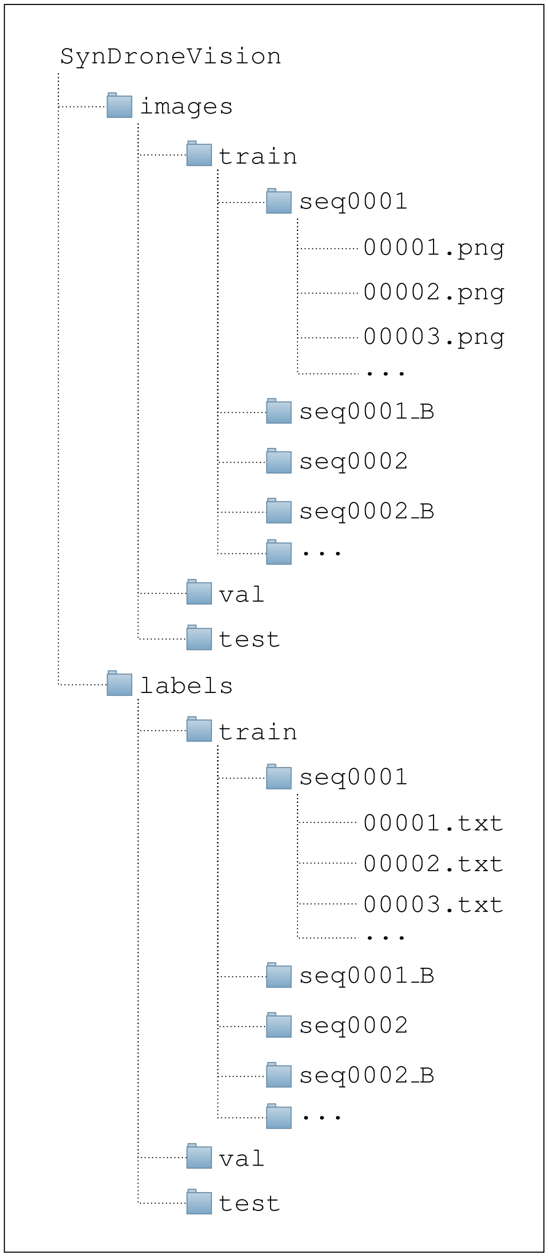

Figure I. Folder configuration of the SynDroneVision dataset.

Table II. Parameter specifications for the Sun and Sky Actor to create environment-dependent illumination variations.

Environment	Solar Time		Direct. Light Intensity (lux)		Rayleigh Scattering (Channel Values)					
					Red		Green		Blue	
	from	to	min	max	min	max	min	max	min	max
University Site	6.00	21.00	1	126	0.014	0.708	0.148	0.900	0.361	1
Venetian City [10]	9.40	17.00	6	15	0.100	0.565	0.199	0.739	0.410	0.990
Farming Grounds [8]	7.24	14.25	5	13	0.119	0.599	0.188	0.836	0.361	1
Modular Cityscape [20]	6.87	17.30	15	70	0.125	0.686	0.225	0.687	0.719	1

and validation sets. Within these subdivisions, there are distinct folders for each image sequence, along with a subset of randomly blurred images (denoted by the suffix ’_B’). Annotations in the labels folder are provided as text files according to the YOLO standard format:

```
<object-class> <x> <y> <width> <height>.
```

Note that <x> and <y> correspond to the normalized coordinates of the bounding box center. The normalization extends to both the bounding box coordinates and dimensions. Figure I shows SynDroneVision’s structural organization.

C. Further Details on SynDroneVision

C.1. Environments

The creation of SynDroneVision, as outlined in Section 3.1 (main paper), involved the application of diverse virtual environments. The majority of these environments are publicly available (some free of charge and some requiring payment). The only exception is the University Site environment, which was specifically designed to replicate a real-world scenario. Table III provides a comprehensive overview of the environments and their respective characteristics. Figure II illustrates camera perspectives and lighting configurations determined for each environment.

C.2. Illumination Parameters

To enhance the range of illumination within the SynDroneVision dataset, we primarily modified the settings of the Sun and Sky Actor [12] and the Post Process Volume [11], essential tools within the Unreal Engine [13].

Sun and Sky Actor. For the Sun and Sky Actor, the following parameters were systematically modified:

- *Solar Time* – The solar time parameter of the Sun and Sky Actor controls the position of the sun with respect to a pre-defined geographical location, simulating the natural progression of time during the day. Adjusting the solar time changes the sun’s position relative to the horizon, creating different lighting conditions and shadows.
- *Directional Light Intensity* – The intensity parameter of the Directional Light Actor controls the brightness of the light. Adjusting this parameter alters the overall illumination and shadow strength in the scene. Higher values increase brightness, while lower values decrease it.
- *Rayleigh Scattering* – The Rayleigh scattering parameter in Unreal’s Sky Atmosphere contains both an RGB value and a scale. While the RGB value specifies the color tint of the scattering effect, the scale controls the overall intensity. This affects the sky’s color and appearance, simulating natural atmospheric phenomena such as blue skies during the day and red hues at sunrise or sunset.

Table II summarizes the (environment-dependent) parameter value ranges employed in generating SynDroneVision.

Post Process Volume. To create variations in the scene’s color grading, we refined the following color temperature-related parameters within the Post Process Volume:

- *Temperature Type* – The Temperature Type parameter specifies the method for adjusting the color temperature of a scene. Available options are White Balance (default) and Color Temperature. White Balance leverages the Temp value to calibrate the virtual camera, maintaining accurate white tones. Color Temperature utilizes the Temp value to directly adjust the scene’s overall color hue. Both methods were employed in the generation process of SynDroneVision.
- *Temp* – The Temp parameter regulates the white balance relative to the scene’s light temperature. While higher values introduce a warm (yellow) coloration, lower values generate a cool (blue) tint. Matching temperature values ensure a neutral white light.
- *Tint* – The Tint parameter refines the white balance tint of a scene, correcting color imbalances to attain a more natural color representation across different light temperatures.

The parameter value ranges are detailed in Table IV.

Table III. Overview of environments used for synthetic data generation. The symbols ✓ (applies) and ✗ (does not apply) indicate an environment's public availability (2nd column from the left) and mandatory cost (3rd column from the left).

Environment	Publ. Avail.	Chargeable	Description
University Site	✗	–	A custom-designed environment replicating a German university campus situated within a wooded landscape. This urban setting features mid-rise building structures and a moderate vegetation density. **Figure II:** images 1–6 (rows 1–3).
Venetian City [10]	✓	✓	A commercially available environment offering a realistic representation of Venice. The included demo map showcases Mediterranean-style buildings, autumnal trees, canals, stone bridges, and additional exterior elements such as benches and street lamps. **Figure II:** images 1–6 (rows 4–5).
Farming Grounds [8]	✓	✓	A small agricultural environment featuring grain fields, fruit-bearing trees, and a variety of vegetable plants. In addition, the scene includes a small greenhouse, multiple raised garden beds, fencing, and other typical agricultural elements such as wooden barrels and crates. **Figure II:** images 1–6 (row 6).
Rural Australia [14]	✓	✗	A publicly accessible environment capturing the expansive fields and open spaces characteristic of the Australian countryside. It includes detailed representations of natural elements, such as rivers, creeks, and rock formations, as well as native vegetation (e.g., shrubs and grasses) and local fauna (e.g., different bird species in flight). **Figure II:** images 1–6 (row 7).
City Park [23]	✓	✗	An urban park environment characterized by a rich diversity of lush vegetation, including trees, shrubs, flowers, and grass. The park features winding pathways and serene water features such as ponds and fountains. In addition to a few small buildings, the environment includes playgrounds, picnic areas, and sports grounds, as well as urban furniture such as benches, lampposts, and trash cans. **Figure II:** images 1–5 (row 8).
Factory Grounds [22]	✓	✗	An open-access environment showcasing a factory site. It exhibits various aspects of industrial architecture, including structures such as warehouses, production facilities, assembly lines, and storage installations, along with an extensive network of pipes, ducts, and other infrastructure. The environment also features a variety of machinery and equipment commonly found in factories or industrial settings. **Figure II:** image 6 (row 8), images 1–6 (row 9).
Urban Downtown [19]	✓	✗	A freely accessible environment featuring a Midwestern outdoor mall. Thus, the buildings are predominantly commercial, including shops, cafes, and restaurants. The urban design incorporates outdoor seating areas, green spaces, and playgrounds, set against a backdrop of mountains. The represented vegetation comprises flowers, small shrubs, and trees, evoking a summer-like setting. **Figure II:** images 1–6 (row 10), images 1–4 (row 11).
Modular Cityscape [20]	✓	✗	An urban scene characterized predominantly by buildings (both commercial and residential) with diverse architectural styles. The environment integrates urban infrastructure, including streets and sidewalks, and is equipped with urban furniture such as benches, bus stops, streetlights, and trash receptacles. **Figure II:** images 5–6 (row 11), images 1–6 (last row)

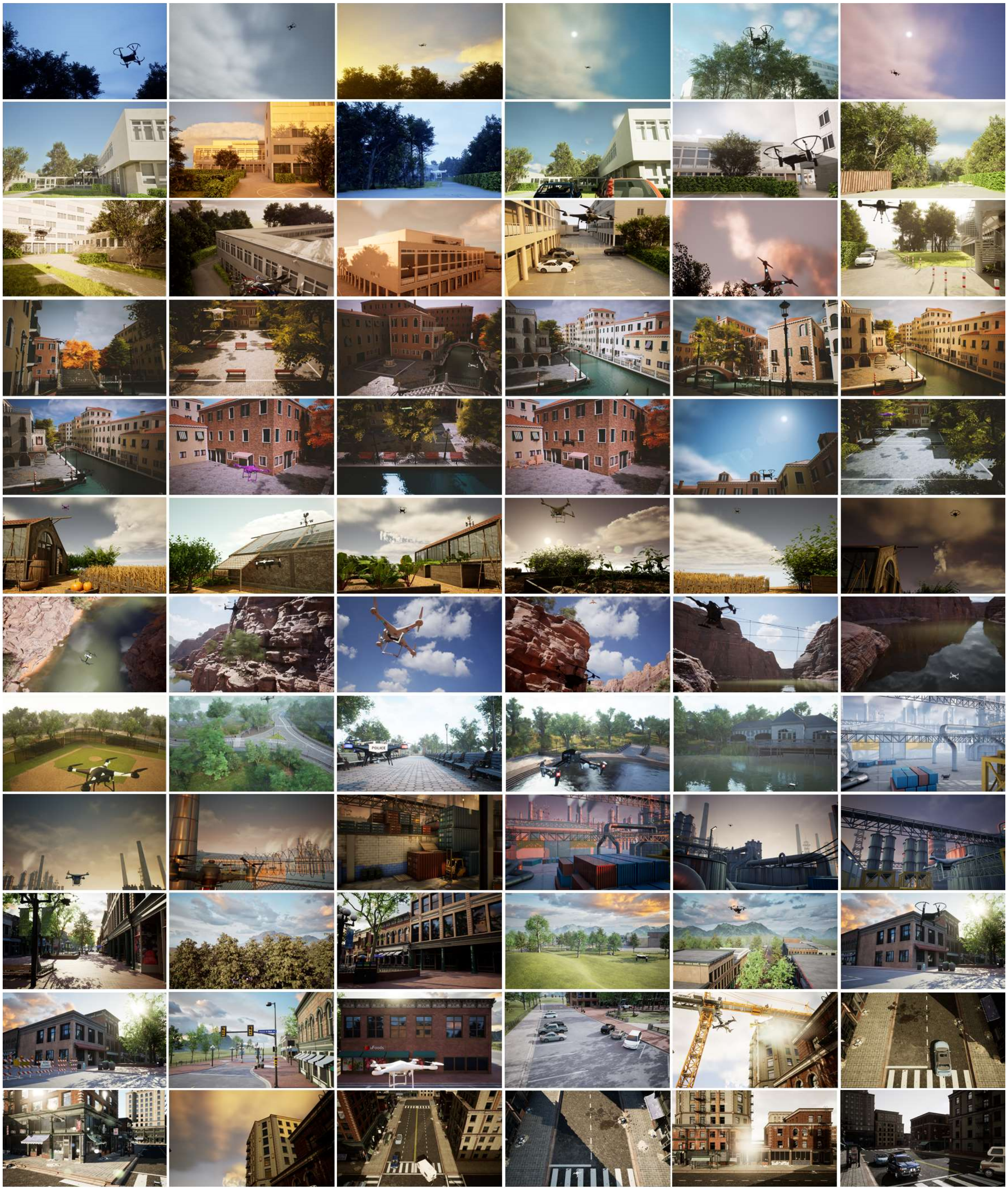

Figure II. Customized camera perspectives and lighting configurations tailored to each environment. The camera fields of view (FOVs) correspond to the following environments (arranged from left to right, top to bottom): University Site (rows 1-3), Venetian City (rows 4-5), Farming Grounds (row 6), Rural Australia (row 7), City Park (images 1-5, row 8), Factory Grounds (image 6, row 8; images 1-6, row 9), Urban Downtown (images 1-6, row 10; images 1-4, row 11), and Modular Cityscape (images 5-6, row 11; images 1-6, last row).

Table IV. Post Process Volume settings.

Environment	Temp		Tint	
	min	max	min	max
University Site	4,400	12,000	0	0.30
Venetian City [10]	3,840	15,000	0	0.25
Farming Grounds [8]	4,588	4,588	0.05	0.05
Modular Cityscape [20]	4,770	9,500	-0.02	0.03

Table V. Technical configuration details for Unreal projects.

Global Illumination	
Dynamic Global Illumination Methods	Lumen
Reflection	
Reflection Method	Lumen
Reflection Capture Resolution	128
Reduce Lightmap Mixing on Smooth Surfaces	✓
Support Global Clip Plane for Planar Reflections	✓ ★
Lumen	
Use Hardware Ray Tracing	✓
Ray Lighting Mode	Surface Cache
Software Ray Tracing Mode	Detail Tracing
Hardware Ray Tracing	
Support Hardware Ray Tracing	✓
Path Tracing	✓
Software Ray Tracing	
Generate Mesh Distance Fields	✓
Distance Field Voxel Density	0.2

★ not enabled for University Site

Rendering Settings. The rendering settings of an Unreal project have a profound impact on both visual quality and system performance. In the generation process of SynDroneVision, we employed the rendering configurations specified in Table V for the majority of environments. Exceptions include the environments Factory Grounds [22] and City Park [23], which retained the default settings.

C.3. Object Area Ratio and Object Aspect Ratio

Supplementing the characteristics presented in Section 3.5 (main paper), Figure III illustrates the distributions of object area (top) and object aspect ratios (bottom) for drones in the SynDroneVision dataset. Across all dataset partitions – training, validation, and test – the distribution of object area ratios exhibit a pronounced rightward skew. A comparable trend is observed in the distribution of aspect ratios.

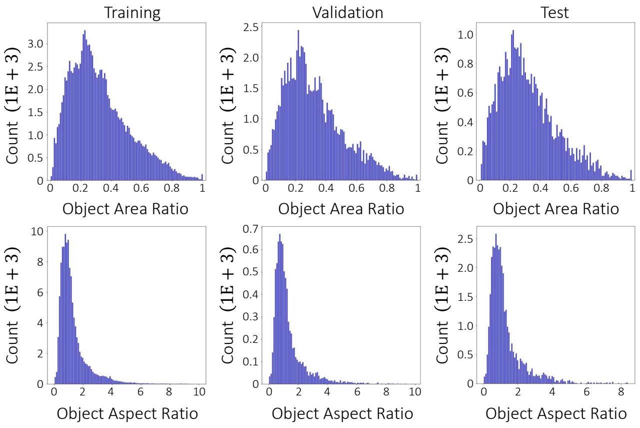

Figure III. Object area and object aspect ratio distribution in the SynDroneVision dataset across training, validation, and test splits.

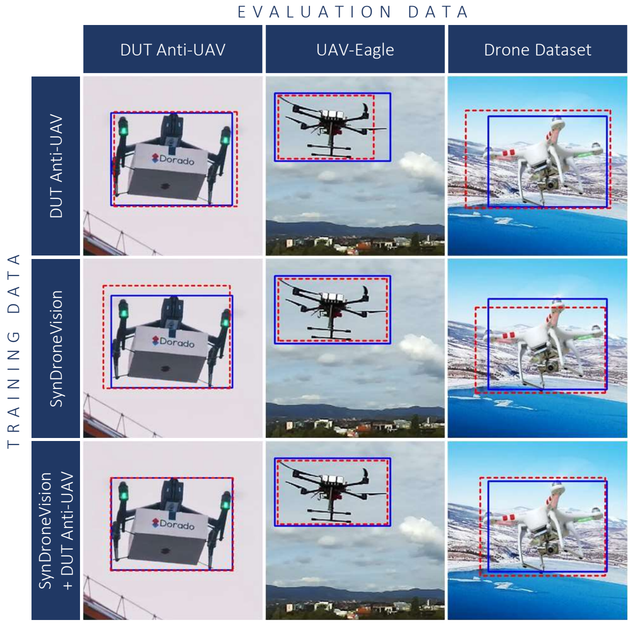

Figure IV. YOLOv9e detections on the DUT Anti-UAV test set [27] (1st column), the UAV-Eagle dataset [3] (2nd column), and the Drone Dataset by [2] (last column) demonstrating improved bounding box precision for models trained on both SynDroneVision and DUT Anti-UAV data (last row). Predictions (red dashed line) are marked alongside ground truth (solid blue line).

D. Analysis Details

D.1. Detection Examples

Figure IV presents selected examples from the DUT Anti-UAV [27], the UAV-Eagle [3], and the Drone Dataset by [2], along with their corresponding detection outcomes obtained using YOLOv9e. The effectiveness of the detection results is compared across all three training strategies,

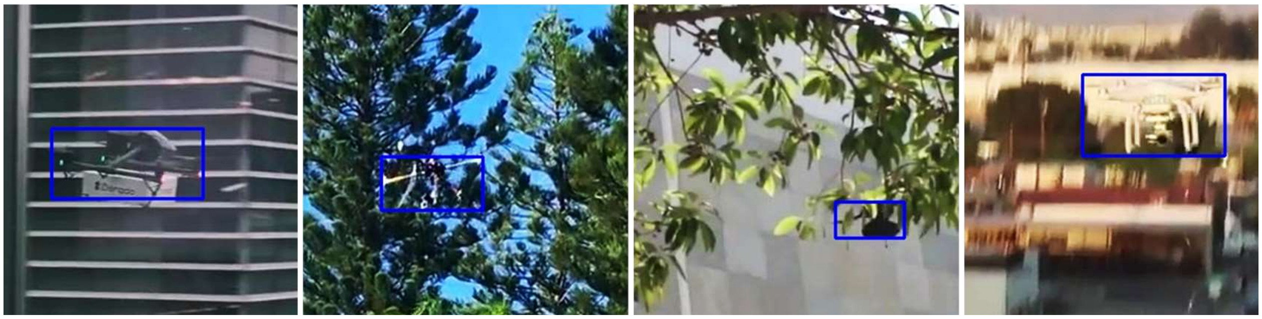

Figure V. Small cut-outs of selected samples from the DUT Anti-UAV test set that failed to be detected by YOLOv9e, irrespective of the employed training data. Ground truth bounding boxes are marked in blue.

Table VI. Performance of YOLOv8m, YOLOv8l, and YOLOv9c on the UAV-Eagle dataset [3] and the Drone Dataset by [2] (out-of-distribution data) across different training data configurations. The SynDroneVision dataset is abbreviated as SDV.

Evaluation Data	YOLO	Training Data		mAP ↑			FNR ↓	FDR ↓
		SDV (Ours)	DUT Anti-UAV	@0.25	@0.5	@0.5-0.95		
		(synthetic)	(real)					
UAV-Eagle [3] (real)	v8m	✓	–	0.944	0.771	0.293	0.201	0.169
		–	✓	0.935	0.823	0.302	0.199	**0.063**
		✓	✓	**0.961**	**0.849**	**0.350**	**0.136**	0.089
	v8l	✓	–	0.951	0.786	0.304	0.217	0.125
		–	✓	0.920	0.725	0.217	0.180	0.224
		✓	✓	**0.979**	**0.869**	**0.368**	**0.126**	**0.074**
	v9c	✓	–	0.926	0.770	0.289	0.216	0.163
		–	✓	0.922	0.799	0.275	0.219	**0.077**
		✓	✓	**0.975**	**0.859**	**0.353**	**0.141**	0.092
Drone Dataset by [2] (real)	v8m	✓	–	0.758	0.527	0.188	0.310	0.138
		–	✓	0.801	0.560	0.208	0.278	**0.113**
		✓	✓	**0.824**	**0.613**	**0.232**	**0.196**	0.114
	v8l	✓	–	0.768	0.515	0.193	0.389	**0.076**
		–	✓	**0.800**	0.552	0.199	0.263	0.227
		✓	✓	0.799	**0.603**	**0.227**	**0.216**	0.116
	v9c	✓	–	0.737	0.530	0.199	0.401	**0.073**
		–	✓	0.806	0.556	0.206	0.292	0.134
		✓	✓	**0.825**	**0.606**	**0.224**	**0.198**	0.126

i.e., YOLOv9e trained (i) exclusively on SynDroneVision (first row), (ii) solely on DUT Anti-UAV (second row), and (iii) on a combination of both datasets (last row). The figure illustrates the superior bounding box localization achieved by the strategic combination of both datasets during training, supporting the significant performance enhancements in mAP values discussed in Section 4.2 (main paper). Conversely, Figure V displays selected examples from DUT Anti-UAV where YOLOv9e fails to detect existing drones, irrespective of the training data. Detection failures are most commonly observed in scenarios where the drone's visibility is significantly compromised, e.g., due to camouflage effects [18] (see Figure V, second image from the left) or occlusions (see Figure V, third image from the left).

D.2. Performance on Out-of-Distribution Data

Section 4.2 (main paper) highlights that the performance and robustness enhancements achieved with SynDroneVision on out-of-distribution data are not limited to YOLOv9e, but extend to other YOLO variants as well. Evaluating YOLOv8m, YOLOv8l, and YOLOv9c on the UAV-Eagle dataset also demonstrates that training exclusively with either SynDroneVision or DUT Anti-UAV yields comparably strong results across all performance indicators (see Table VI). In some cases, models trained solely on SynDroneVision perform even better than those trained on real-world data, particularly in terms of mAP values at an IoU threshold of 0.25. In analogy to YOLOv9e, the best performance is achieved when combining both datasets dur-

Table VII. Performance of YOLOv8m, YOLOv8l, YOLOv9c, and YOLO9e on the SynDroneVision test set across different training data configurations. The SynDroneVision dataset is abbreviated as SDV.

YOLO	Training Data		Evaluation on SynDroneVision				
	SDV (Ours)	DUT Anti-UAV	mAP ↑			FNR ↓	FDR ↓
	(synthetic)	(real)	@0.25	@0.5	@0.5-0.95		
v8m	✓	–	0.995	0.995	0.944	0.013	0
	✓	✓	0.995	0.995	0.942	0.014	0
v8l	✓	–	0.995	0.995	0.955	0.014	0.001
	✓	✓	0.995	0.995	0.956	0.013	0
v9c	✓	–	0.995	0.995	0.952	0.014	0.001
	✓	✓	0.995	0.995	0.954	0.014	0
v9e	✓	–	0.995	0.995	0.967	0.014	0.001
	✓	✓	0.995	0.995	0.967	0.014	0

ing training. Here, YOLOv8l exhibits the most significant improvement over exclusive real-world data training, featuring a 14.4 percentage point increase in mAP at an IoU threshold of 0.5 and a 10.51 percentage point improvement across a range of IoU thresholds from 0.5 to 0.95 (cf. Table VI). Furthermore, integrating synthetic and real-world data effectively lowers the FNR, whereas variations in the FDR remain inconsistent.

For the Drone Dataset by [2], models trained exclusively on SynDroneVision exhibit slightly lower mAP values compared to those trained solely on DUT Anti-UAV. Nevertheless, the integration of both datasets yields overall performance enhancements, as detailed in Table VI. The only exception seems to be YOLOv8l, where the mAP value at an IoU threshold of 0.25 is marginally higher for the model trained exclusively on DUT Anti-UAV. However, the discrepancy is negligible, with a difference of only 0.001.

D.3. Performance on SynDroneVision

To provide a comprehensive understanding of model performance, we also incorporate the SynDroneVision test set into our evaluation. Specifically, we focus on models trained either exclusively on SynDroneVision or on a combination of SynDroneVision and DUT Anti-UAV. Table VII highlights the consistently high performance of the models across all performance indicators.

References

[1] Simeon Okechukwu Ajakwe, Vivian Ukamaka Ihekoronye, Golam Mohtasin, Rubina Akter, Ali Aouto, Dong Seong Kim, and Jae Min Lee. *VisioDECT Dataset: An Aerial Dataset for Scenario-Based Multi-Drone Detection and Identification*. IEEE Dataport, 2022. 2, 3

[2] Mehmet Çağri Aksoy, Alp Sezer Orak, Hasan Mertcan Özkan, and Bilgin Selimoğlu. Drone Dataset: Amateur Unmanned Air Vehicle Detection. *Mendeley Data*, V4, 2019. 1, 2, 7, 8, 9

[3] Antonella Barisic, Frano Petric, and Stjepan Bogdan. Brain over Brawn: Using a Stereo Camera to Detect, Track, and Intercept a Faster UAV by Reconstructing the Intruder's Trajectory. *Field Robotics*, 2:222–240, 2021. 1, 2, 7, 8

[4] Antonella Barisic, Frano Petric, and Stjepan Bogdan. Sim2Air - Synthetic Aerial Dataset for UAV Monitoring. *IEEE Robotics and Automation Letters*, 7(2):3757–3764, 2022. 2, 3

[5] Blender Foundation. Blender. https://www.blender.org/, accessed: 2024-10-15. 3

[6] Yueru Chen, Pranav Aggarwal, Jongmoo Choi, and C.-C. Jay Kuo. A Deep Learning Approach to Drone Monitoring. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 686–691, 2017. 1, 2

[7] Angelo Coluccia, Alessio Fascista, Lars Sommer, Arne Schumann, Anastasios Dimou, and Dimitrios Zarpalas. The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results. *IEEE Open Journal of Signal Processing*, 5:766–779, 2024. 2, 3

[8] CropCraft Studios. Ultimate Farming. https://www.unrealengine.com/marketplace/en-US/product/ultimate-farming. accessed: 2024-10-15. 4, 5, 7

[9] Cycles Developers. Cycles. https://www.cycles-renderer.org/", accessed: 2024-10-15. 3

[10] Deelus. Venice - Fast Building. https://www.unrealengine.com/marketplace/en-US/product/venice-fast-building. accessed: 2024-10-15. 4, 5, 7

[11] Epic Games. Post Process Effects. https://docs.unrealengine.com/5.3/en-US/post-process-effects-in-unreal-engine/, accessed: 2024-10-15. 4

[12] Epic Games. Sun and Sky Actor. https://docs.unrealengine.com/4.27/en-US/BuildingWorlds/LightingAndShadows/SunSky/, accessed: 2024-10-15. 4

[13] Epic Games. Unreal Engine. https://www.unrealengine.com/en-US/, accessed: 2024-10-15. 4

[14] Andrew Svanberg Hamilton. Rural Australia. https://www.unrealengine.com/marketplace/en-US/product/rural-australia. accessed: 2024-10-15. 5

[15] Brian K. S. Isaac-Medina, Matt Poyser, Daniel Organisciak, Chris G. Willcocks, T. Breckon, and Hubert P. H. Shum. Unmanned Aerial Vehicle Visual Detection and Tracking using Deep Neural Networks: A Performance Benchmark. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 1223–1232, 2021. 1

[16] Sonain Jamil, Muhammad Sohail Abbas, and Arunabha M. Roy. Distinguishing Malicious Drones Using Vision Transformer. *AI*, 3(2):260–273, 2022. 2

[17] Nan Jiang, Kuiran Wang, Xiaoke Peng, Xuehui Yu, Qiang Wang, Junliang Xing, Guorong Li, Guodong Guo, Qixiang Ye, Jianbin Jiao, Jian Zhao, and Zhenjun Han. Anti-UAV: A Large-Scale Benchmark for Vision-Based UAV Tracking. *IEEE Transactions on Multimedia*, 25:486–500, 2023. 2, 3

[18] Tamara R. Lenhard, Andreas Weinmann, Stefan Jäger, and Tobias Koch. YOLO-FEDER FusionNet: A Novel Deep Learning Architecture for Drone Detection. In *IEEE International Conference on Image Processing*, pages 2299–2305, 2024. 8

[19] PurePolygons. Downtown West Modular Pack. https://www.unrealengine.com/marketplace/en-US/product/6bb93c7515e148a1a0a0ec263db67d5b. accessed: 2024-10-15. 5

[20] PurePolygons. Modular Building Set. https://www.unrealengine.com/marketplace/en-US/product/modular-building-set. accessed: 2024-10-15. 4, 5, 7

[21] Alejandro Rodriguez-Ramos, Javier Rodriguez-Vazquez, Carlos Sampedro, and Pascual Campoy. Adaptive Inattentional Framework for Video Object Detection With Reward-Conditional Training. *IEEE Access*, 8:124451–124466, 2020. 1, 2

[22] Denys Rutkovskyi. Factory Environment Collection. https://www.unrealengine.com/marketplace/en-US/product/factory-environment-collection. accessed: 2024-10-15. 5, 7

[23] SilverTm. City Park Environment Collection. https://www.unrealengine.com/marketplace/en-US/product/city-park-environment-collection. accessed: 2024-10-15. 5, 7

[24] Fredrik Svanström, Fernando Alonso-Fernandez, and Cristofer Englund. A Dataset for Multi-Sensor Drone Detection. *Data in Brief*, 39:107521, 2021. 1, 2

[25] Ye Wang, Yueru Chen, Jongmoo Choi, and C.-C. Jay Kuo. Towards Visible and Thermal Drone Monitoring with Convolutional Neural Networks. *Asia-Pacific Signal and Information Processing Association Transactions on Signal and Information Processing*, 8, 2018. 1, 2

[26] Yuni Zeng, Qianwen Duan, Xiangru Chen, Dezhong Peng, Yao Mao, and Ke Yang. UAVData: A Dataset for Unmanned Aerial Vehicle Detection. *Soft Comput.*, 25(7):5385—5393, 2021. 1, 2

[27] Jie Zhao, Jingshu Zhang, Dongdong Li, and Dong Wang. Vision-Based Anti-UAV Detection and Tracking. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):25323–25334, 2022. 2, 7

[28] Ye Zheng, Zhang Chen, Dailin Lv, Zhixing Li, Zhenzhong Lan, and Shiyu Zhao. Air-to-Air Visual Detection of Micro-UAVs: An Experimental Evaluation of Deep Learning. *IEEE Robotics and Automation Letters*, 6(2):1020–1027, 2021. 1, 2