

# SynDroneVision: A Synthetic Dataset for Image-Based Drone Detection

Tamara R. Lenhard<sup>1,2</sup>

Andreas Weinmann<sup>2</sup>

Kai Franke<sup>1</sup>

Tobias Koch<sup>1</sup>

<sup>1</sup>Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR)

<sup>2</sup>Working Group Algorithms for Computer Vision, Imaging and Data Analysis,  
University of Applied Sciences Darmstadt

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

## Abstract

*Developing robust drone detection systems is often constrained by the limited availability of large-scale annotated training data and the high costs associated with real-world data collection. However, leveraging synthetic data generated via game engine-based simulations provides a promising and cost-effective solution to overcome this issue. Therefore, we present SynDroneVision, a synthetic dataset specifically designed for RGB-based drone detection in surveillance applications. Featuring diverse backgrounds, lighting conditions, and drone models, SynDroneVision offers a comprehensive training foundation for deep learning algorithms. To evaluate the dataset's effectiveness, we perform a comparative analysis across a selection of recent YOLO detection models. Our findings demonstrate that SynDroneVision is a valuable resource for real-world data enrichment, achieving notable enhancements in model performance and robustness, while significantly reducing the time and costs of real-world data acquisition. SynDroneVision will be publicly released upon paper acceptance.*

## 1. Introduction

Unmanned aerial vehicles (UAVs), commonly known as drones, have become integral to a variety of sectors, including agriculture, logistics, surveillance, and recreation. However, their rapid proliferation introduces new challenges, particularly in terms of security and privacy protection [14]. Therefore, the implementation of effective drone detection systems is crucial to mitigate the risks associated with unauthorized or malicious drone activities. Combining optical sensors, specifically cameras, with advanced deep learning (DL) techniques represents a highly promising and economically efficient detection strategy [18]. Nevertheless, the effectiveness of DL models is heavily reliant on

extensive and diverse training data [44, 52].

In practical applications, the acquisition of substantial amounts of annotated real-world data is both time-consuming and resource-intensive [28, 45]. Additional constraints, such as non-fly zones and adverse weather conditions, further complicate the data collection process. Leveraging synthetic data presents a viable alternative to circumvent environmental limitations and significantly reduce acquisition costs [16, 44] (not only in drone detection, but also in other domains [4, 29, 34, 47]). In particular, the application of game engine-based data generation techniques enables the efficient and physically precise simulation of diverse real-world conditions [7, 16, 30, 35]. It facilitates the seamless interchange of environmental configurations (e.g., from urban landscapes to rural terrains), offering the potential for comprehensive coverage of diverse scenarios, including those inadequately represented by real-world data. Furthermore, the ability to rapidly alter illumination, time of day, and weather conditions – from clear summer days to overcast skies within seconds – provides a time-efficient solution without compromising data diversity. A broad selection of interchangeable drone models, materials, and textures supports a high variability in drone appearances, in contrast to the often limited drone selection in practical applications [2, 55]. Moreover, a key advantage of synthetic data is the automated generation of pixel-precise annotations [31]. This capability accelerates both training and validation processes, enabling more rapid experimentation and iteration cycles. Furthermore, it significantly decreases resource requirements associated with traditional data collection methods [28], particularly in terms of annotation costs and recording time.

Despite the substantial benefits of synthetic data, there is still a gap between simulated scenarios and real-world conditions [6, 35]. This discrepancy can negatively impact detection quality, especially when transferring drone detection

models trained exclusively on synthetic data to real-world applications. However, leveraging hybrid datasets – incorporating both real and synthetic data, with real data shares < 1% – has proven to be an effective training strategy to overcome this issue [17, 46].

Although the generation of synthetic data is a topic of ongoing research, particularly in the field of drone detection, there is currently only one publicly available synthetic dataset: the S-UAV-T dataset by Barisic *et al.* [6]. This dataset is specifically designed for UAV-to-UAV detection, featuring perspectives that deviate from standard surveillance configurations.

**Contributions.** Addressing the scarcity of publicly available synthetic datasets for image-based drone detection, we introduce *SynDroneVision* – a comprehensive dataset featuring diverse environments, drone models, and lighting conditions. We provide a comparative analysis of state-of-the-art models, demonstrating the effectiveness of SynDroneVision, especially in combination with real-world data. Furthermore, we assess the robustness of this approach using out-of-distribution data.

The remainder of this paper is organized as follows: Section 2 provides an overview of related research on image-based drone detection and publicly available datasets. Section 3 details the generation process, defining features, and the composition of SynDroneVision. The experimental design and results are outlined in Section 4. Conclusions are presented in Section 5.

## 2. Related Work

In the following sections, we briefly summarize recent advances in image-based drone detection and assess the characteristics of publicly available drone detection datasets, emphasizing their similarities and differences.

### 2.1. Drone Detection

State-of-the-art techniques for RGB-based drone detection primarily leverage single-stage DL algorithms, which offer an optimal balance between real-time performance and precision. A majority of methodologies employ variants of the You Only Look Once (YOLO) models, including YOLOv3, YOLOv5, and YOLOv8, either in their original configurations [6, 38] or with custom modifications [24, 31, 33]. Architectural innovations typically seek to resolve particular challenges encountered in drone detection, including the identification of small drones [24, 28, 32, 33], the differentiation between drones and other aerial entities (e.g., birds) [11, 33], and the mitigation of camouflage effects [31]. Transformer-based approaches offer an effective, albeit less frequently employed, alternative to traditional detection techniques [27].

Training drone detection models predominantly relies on (self-collected) application-specific real-world data. How-

ever, significant efforts are also directed towards the creation and utilization of synthetic data (e.g., see [6, 35]). Despite variations in generation techniques, prevailing research highlights the substantial potential of synthetic data, particularly in combination with real-world data. Prevalent training strategies for improving detection quality by integrating real and synthetic data include mixed-data training [9, 46] and fine-tuning models, initially trained on synthetic data, with real-world data [6, 35]. However, the optimal ratio of synthetic to real-world data is a controversial topic of ongoing research [17].

## 2.2. Datasets

Publicly available datasets for drone detection can be divided into two primary categories. The first category includes datasets exclusively designed for detection tasks [2, 5, 6], typically featuring individual images. The second category comprises datasets that support both detection and tracking [9, 11, 51, 54], generally including sequential data or specialized subsets featuring both image and video files. Except for the dataset by Barisic *et al.* [6], the majority of datasets consists of real-world data sourced from Google images [2], YouTube videos [2], or self-recorded footage [9, 51, 53] using static, moving, handheld, or drone-mounted devices. Beyond the variability in data origins and collection techniques, available datasets exhibit further variations in the following attributes:

- *Dataset Size* – The costly nature of real-world data acquisition leads to significant discrepancies in the sizes of publicly available datasets. For instance, the datasets UAV-Eagle [5] and Malicious Drones [27] are relatively small, with 510 and 776 images, respectively (see Table 1). Most datasets comprise 4,000 [2] to 40,232 images [41]. Exceptions include the Halmstad Data [45], featuring 203,328 annotated frames (IR + RGB), and the Drone-vs-Bird Detection Challenge dataset [11], with 85,904 annotated frames. (Notably, the Drone-vs-Bird Detection Challenge dataset [11] constitutes a comprehensive compilation of data, collected over time and continually enhanced by input from various contributors.)
- *Image Resolution* – Publicly available drone detection datasets encompass a wide spectrum of resolutions, ranging from low-resolution (e.g., 224×224 pixels [27] and 608×608 pixels [6]) to high-resolution images (e.g., 5616×3744 pixels [54]). The variability in resolution is observed both across different datasets and within individual datasets. While some datasets maintain a uniform resolution [1, 5, 6, 9, 27, 45, 51, 55], others offer a diverse range of resolutions [2, 11, 54]. For instance, the DUT Anti-UAV dataset provided by

Table 1. Comprehensive information on publicly available datasets for image-based drone detection, encompassing both real and synthetic data. The symbols **X** (does not apply) and **✓** (applies) indicate the presence of image sequences (4th column from the right), the inclusion of diverse drone models (3rd column from the right), and the incorporation of distractor objects (2nd column from the right).

Dataset	No. Images				Img. Sequ.	Diff. Drones	Dist. Objs.	Max. Img. Resolution
	train	val	test	total				
USC Drone Detect. & Track. [9, 51]	–	–	–	27,000*	✓	X	X	1920×1080
Drone Dataset [2]	3,611	–	401	4,012	X	X	X	3840×2160
MAV-VID [41]	29,500	10,732	–	40,232	✓	✓	X	–
Det-Fly [55]	–	–	–	13,271*	(✓)	X	X	3840×2160
UAV-Eagle [5]	–	–	–	510*	✓	X	X	1920×1080
UAVData [53]	–	–	–	13,803*	✓	✓	✓	1280×720
Halmstadt Data [45]	–	–	–	203,328▲	✓	✓	✓	640×512
DUT Anti-UAV [54]	5,200	2,600	2,200	10,000	(✓)	✓	X	5616×3744
VisioDECT [1]	–	–	–	20,924*	✓	✓	X	852×480
Malicious Drones [27]	543	–	233	776	X	✓	✓	224×224
S-UAV-T [6] ( <i>synthetic</i> )	–	–	–	52,500*	X	✓	X	608×608
Drone-vs-Bird Detection Ch. [11]	85,904	–	–*	85,904	✓	✓	✓	3840×2160

\* not publicly available    ★ no subdivision into train, val, and test

▲ RGB + IR data

Zhao *et al.* [54] includes images with resolutions varying from 240×160 to 5616×3744 pixels, whereas the Det-Fly dataset by Zheng *et al.* [55] is characterized by a consistent resolution of 3840×2160 pixels.

- *Drone Models, Size & Position* – The representation of drone models across various datasets exhibits considerable heterogeneity. While some datasets are restricted to a single drone model [2, 41, 55], others encompass multiple models [1, 11, 45, 53, 54], typically comprising three [45] to eight [11] distinct types. Exceptions include the DUT Anti-UAV dataset [54], featuring 35 drone models, and the synthetic dataset by Barisic *et al.* [6]. Additionally, variability is observed in the number of drones per image (ranging from single [5, 41] to multiple drones [6, 53, 54] per frame), as well as in their size and position. Nevertheless, a common feature across most datasets is the prevalence of small, centrally positioned drones, typically in conventional colors such as black and / or white.
- *Distractor Objects* – In addition to drones, some datasets encompass other (drone-like) objects [2], including birds [11, 27, 45], airplanes [27, 45], helicopters [27, 45], and balloons [53]. These distractor objects are either explicitly annotated [27, 45] or incorporated in a more implicit manner [2, 11, 53].
- *Backgrounds* – The majority of datasets feature outdoor environments such as urban landscapes, forests, farmland, airports, and coastal areas across different regions around the globe (e.g., Sweden [45] vs. China [54]). The visual compositions comprise an assortment of elements, including skies, buildings, play-

grounds, vegetation, and other landscape features, captured from diverse viewing angles (e.g., top-down and bottom-up, as in [54]). An exception is the dataset by Zeng *et al.* [53] which also features indoor scenes.

- *Illumination & Weather* – Real-world datasets are predominantly recorded in daylight [45, 53, 54] and feature diverse weather conditions [11, 27] (including cloudy [1, 54], sunny [1, 54], and snowy [54]). Some datasets also account for low-light conditions such as night, dawn, and dusk [1, 11, 54]. Reflecting the unpredictable nature of real-world lighting, these datasets often (unintentionally) exhibit rapid illumination changes or direct sun glare [11, 53]. In contrast, the synthetic dataset by Barisic *et al.* [6] is characterized by more controlled lighting conditions, including daylight and twilight.
- *Annotation Process* – Annotations are typically generated manually by individual experts [41] or teams [1]. An exception is the synthetic dataset by Barisic *et al.* [6], which employs an automated annotation pipeline based on Blender [8] and Cycles [13]. Manual annotation techniques often suffer from inconsistencies in terms of quality and precision. Additionally, variations in file formats (e.g., .txt [5] vs. .mat [45]) and bounding box definitions further compromise the datasets’ practical applicability. Methodologies, tools, and quality control measures are often insufficiently documented, resulting in a lack of transparency.

In addition to RGB data, some datasets include other imaging modalities, such as infrared (IR) data [45]. Table 1

provides an overview of existing datasets and their characteristics, with further details in the supplementary material.

### 3. SynDroneVision Dataset

This section provides a comprehensive overview of the proposed SynDroneVision dataset, detailing the employed generation technique, simulation parameter variations, its composition, and inherent characteristic properties.

#### 3.1. Data Generation Process

The synthetic RGB data of the proposed dataset is generated using an advanced iteration of the data generation pipeline introduced by Dieter *et al.* [16]. Unlike Dieter *et al.*'s pipeline — reliant on Microsoft AirSim [36] and Unreal Engine 4.25 [22] — the employed pipeline leverages Colosseum [10] (the successor to Microsoft AirSim) and Unreal Engine 5.0. Unreal Engine 5.0 introduces advanced capabilities for rendering dynamic global illumination and reflections through the implementation of the fully dynamic global illumination and reflections system Lumen [19]. This advancement significantly enhances the realism of depicted scenes (especially in terms of daytime-dependent light and shadow variations), thereby elevating the fidelity of synthetic RGB data. To ensure precise representation of Lumen-based lighting effects and reflections, we refine the data generation process by modifying the pipeline's data generation module (cf. [16]). In Dieter *et al.*'s pipeline, visual sensor data was acquired through Scene Capture 2D Actors. However, when integrated with Unreal Engine 5.0, these components lack the ability to accurately capture the intricate details of Lumen's dynamic global illumination and reflections. To address this limitation, we implement the capture of RGB data via high resolution screenshots. All other pipeline components remain consistent with [16]. For detailed information on individual components, refer to [16].

The data acquisition process itself is predicated on a strategic placement of stationary virtual camera sensors, whose position and orientation are pre-determined with respect to the underlying simulation environment (inspired by a typical surveillance setup, cf. Figure 1). Data collection is systematically performed from each designated camera in a sequential manner, adhering to a pre-defined recording duration. The recording duration is synchronized with the drone's flight time. The drone's flight trajectory is determined by a probabilistic selection of waypoints, randomly positioned within the camera's field of view (up to 30 meters from its vantage point).

#### 3.2. Simulation Parameter and Domain Variations

To foster a high level of diversity in the generated data, we introduce variations across multiple simulation components, including the environment, drone models, and light-

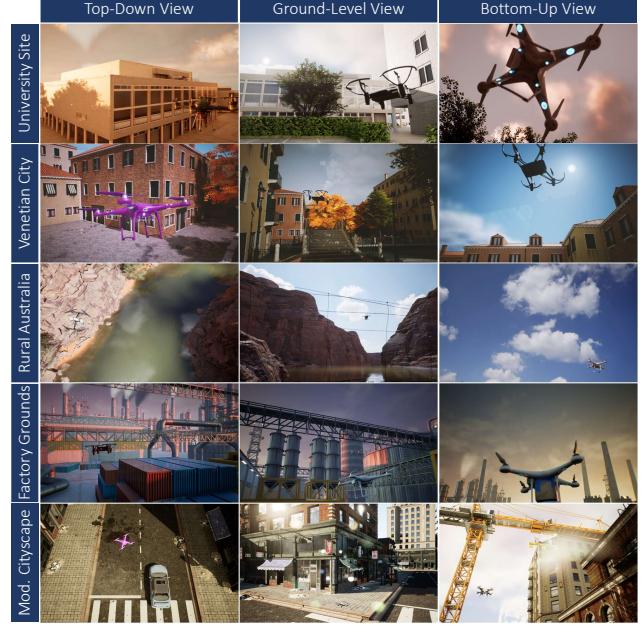


Figure 1. A selection of synthetic images captured from diverse virtual environments – University Site (row 1), Venetian City (row 2), Rural Australia (row 3), Factory Grounds (row 4), and Modular Cityscape (last row) – demonstrating SynDroneVision’s diversity in terms of environmental conditions and camera perspectives (top-down, ground-level, and bottom-up).



Figure 2. Drone models from [3, 25, 37] employed in the generation of the SynDroneVision dataset.

ing conditions. Detailed information on these aspects is provided in the following sections.

##### 3.2.1 Environments

To establish a foundation for simulating physically-realistic drone flights, we leverage a variety of three-dimensional environments, encompassing both commercially licensed and freely available options. The selection of environments is guided by the defining attributes of the real-world settings utilized in creating the DUT Anti-UAV dataset [54]. Particular emphasis is placed on incorporating environments with diverse complexity levels and substantial variations (cf. Figure 1), ensuring thorough data diversification. The

employed environments and settings include: University Site, Venetian City [15], Farming Grounds [12], Rural Australia [23], City Park [43], Factory Grounds [42], Urban Downtown [39], and Modular Cityscape [40]. In publicly accessible environments (commercial or free), we utilized pre-existing demo maps (with minor modifications, cf. Section 3.2.3) for data generation. Note that the environments predominantly consist of static geometry, with two notable exceptions: the shader-based animation of foliage and the dynamic motion of the drone. A detailed overview of the environments and their inherent characteristics can be found in the supplementary material (Table II and Figure II).

### 3.2.2 Drone Models

To ensure a high degree of diversity in drone representation, we employ a selection of drone models from the commercially available Quadcopter Pack [37], Drone Pack [3], and Military Drone Pack [25] for synthetic data generation (see Figure 2). Our selection features a variety of widely-deployed drone models, including the DJI Phantom (Figure 2, second row, second model from the left) and the DJI Tello Ryze (Figure 2, last row, third model from the left). Each drone model is rendered with realistic textures sourced from the respective asset packages. This contrasts with the approach of Barisic *et al.* [6], characterized by texture randomization featuring unconventional drone textures.

### 3.2.3 Illumination

To incorporate a variety of realistic illumination conditions, we utilize the dynamic illumination and reflection system, Lumen [19], in conjunction with the Sun and Sky Actor [21] and the Post Process Volume [20] provided by Unreal Engine. The Sun and Sky Actor offers precise control over the sun’s positioning based on geographic location, date, and time, while the Post Process Volume provides a comprehensive toolkit for regulating visual aesthetics and atmospheric properties (e.g., color grading, contrast, or bloom). By leveraging the dynamic nature of directional lights within the Sun and Sky Actor, lightmap baking becomes obsolete, thus eliminating the need for pre-computed lighting. Consequently, this configuration enables the creation of authentic renderings, accurately portraying the interplay of sunlight and shadow.

During the data generation process, we systematically introduce variations in the intensity of the Directional Light Actor and the Rayleigh scattering properties of the Sky Atmosphere – both fundamental components of the Sun and Sky Actor. Additionally, adjustments are made to the color temperature parameter within the Post Process Volume to further refine atmospheric properties. Thus, SynDroneVision offers a broad spectrum of illumination conditions, ranging from dawn (Figure 1, first row, first image) to dusk

(Figure 1, first row, third image), from clear blue skies (Figure 1, second row, third image) to overcast conditions (Figure 1, first row, second image). Note that the combination of Lumen, Sun and Sky Actor, and Post Process Volume is employed exclusively for data generation within the following environments: University Site, Venetian City [15], Farming Grounds [12], and Modular Cityscape [40]. The default illumination setup included in the environments Rural Australia [23], City Park [43], Factory Grounds [42], and Urban Downtown [39] remains unchanged, as it already features a sophisticated (Lumen-based) implementation of lighting and reflections. Further details on illumination parameters are provided in the supplementary material.

### 3.3 Post-Processing

To further increase the data diversity, a subset of randomly selected images undergoes post-capture blurring. Considering a synthetically generated image  $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$ , where  $W \in \mathbb{N}$  denotes the width and  $H \in \mathbb{N}$  denotes the height, the blurring procedure is defined by the following convolution operation:

$$\mathbf{I}'(x, y) = \sum_{i=0}^{2m} \sum_{j=0}^{2n} \mathbf{I}(x + i - m, y + j - n) \cdot \mathbf{K}(i, j) \quad (1)$$

where  $x \in \{0, \dots, W - 1\}$  and  $y \in \{0, \dots, H - 1\}$ . The kernel  $\mathbf{K} \in \mathbb{R}^{M \times N}$  is characterized by the dimensions  $M = 2m - 1$  and  $N = 2n - 1$  with  $m, n \in \mathbb{N}$ . In our application, the kernel is specified as either an average kernel or a Gaussian kernel. The average kernel is given by  $\mathbf{K}(i, j) = \frac{1}{(2m-1)(2n-1)}$ , with indices  $i \in \{0, \dots, 2m\}$  and  $j \in \{0, \dots, 2n\}$  determining the kernel’s spatial extent. Conversely, the Gaussian kernel is described by  $\mathbf{K}(i, j) = \frac{1}{(2\pi\sigma^2)} \exp\left(-\frac{(i-m)^2 + (j-n)^2}{2\sigma^2}\right)$ , where  $\sigma$  represents the standard deviation of the Gaussian distribution.

For the creation of SynDroneVision, square kernels with dimensions  $M, N \in \{13, 15, 17, 19, 21\}$ ,  $M = N$  are employed. The choice of kernel type and size is randomized independently for each image. Note that for kernel extensions beyond image boundaries (i.e., when  $x \notin \{0, \dots, W - 1\}$  or  $y \notin \{0, \dots, H - 1\}$ ), explicit boundary conditions are imposed. Specifically, replication padding is employed to extend the image boundaries by duplicating edge pixel values.

### 3.4 Composition

For the generation of SynDroneVision, four to thirteen distinct camera positions are established per environment (cf. Table 2). This environment-specific camera placement results in the capture of 72 annotated image sequences. Each sequence is characterized by a unique combination of drone model, lighting configurations, and background composition, providing a high degree of realism and variation.

Table 2. Comprehensive composition overview of the SynDroneVision dataset.

Environment	Image Count				Camera Positions	Drone Models
	train	val	test	total		
University Site	29,111	1,000	500	30,611	13	8
Venetian City	19,566	1,000	500	21,066	8	6
Farming Grounds	9,053	1,000	500	10,553	6	5
Rural Australia	11,589	1,000	500	13,089	6	5
City Park	7,310	1,000	500	8,810	4	4
Factory Grounds	13,759	1,000	500	15,259	7	6
Urban Downtown	14,404	1,000	500	15,904	9	6
Modular Cityscape	14,515	1,000	500	16,015	5	6
<b>Total</b>	<b>119,307 / 131,238*</b>	<b>8,000 / 8,800*</b>	<b>4,000</b>	<b>131,307 / 140,038*</b>	<b>58</b>	<b>13</b>

\* incl. blurring

Table 3. Comprehensive details regarding the area and aspect ratios of objects featured in SynDroneVision – across training, validation, and test splits – in comparison to DUT Anti-UAV [54].

Split	SynDroneVision (Ours)						DUT Anti-UAV					
	Object Area Ratio			Object Aspect Ratio			Object Area Ratio			Object Aspect Ratio		
	min	avg.	max	min	avg.	max	min	avg.	max	min	avg.	max
train	0.001	0.322	1.0	0.021	1.291	9.993	0.000026	0.013	0.700	1.0	1.910	5.420
val	0.006	0.323	1.0	0.020	1.330	9.855	0.000002	0.013	0.690	1.0	1.910	6.670
test	0.009	0.323	1.0	0.041	1.302	8.383	0.000041	0.014	0.470	1.0	1.920	5.090

(An overview of selected camera field of views and annotation details are included in the supplementary material.)

SynDroneVision encompasses all 72 recorded image sequences, thus yielding a total of 131,307 annotated images for image-based drone detection (cf. Table 2). Apart from drone images, the dataset also includes a small share of background images ( $\sim 7\%$ ). The dataset is partitioned into a training, validation, and test set. This allocation yields 119,307 images for training, 8,000 for validation, and 4,000 for testing purposes. The training and validation datasets are further augmented through the application of the aforementioned blurring technique (cf. Section 3.3). Specifically,  $\sim 10\%$  of the images from each set are randomly selected, blurred, and included on top of the original data. This yields a final dataset size of 131,238 images for training and 8,800 images for validation (cf. Table 2). The number of test images remains unchanged.

### 3.5. Characteristics

The proposed SynDroneVision dataset exhibits the following characteristic properties:

- *Image Resolution* – An identical, consistently high resolution of  $2560 \times 1489$  pixels is maintained for all image sequences across all environments.
- *Object Position* – The spatial distribution of objects within the image frame, depicted in Figure 3, reveals

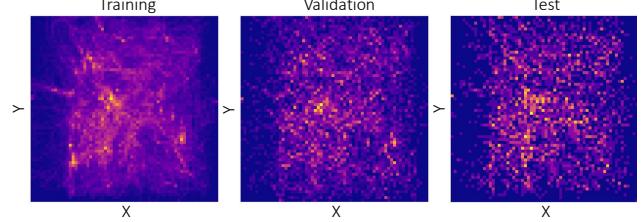


Figure 3. Position distribution of drones within the SynDroneVision dataset. Regions of high frequency are shown in yellow, while areas with no data points are indicated in blue.

a (mostly) uniform dispersion of objects across the entire image area, encompassing both central and peripheral image regions. This contrasts with the distribution patterns observed in datasets like DUT Anti-UAV [28] or MAV-VID [41], where objects are predominantly concentrated in the central region of the image.

- *Object Aspect Ratio* – SynDroneVision exhibits considerable diversity in object aspect ratios. Minimum values range from 0.021 (train) to 0.041 (test), while maximum values range from 8.383 (test) to 9.993 (train). Similar to DUT Anti-UAV, most object aspect ratios fall between 1.0 and 3.0, with average values between 1.291 and 1.330 (cf. Table 3). However, SynDroneVision covers a broader spectrum of aspect ratios compared to DUT Anti-UAV, which features as-

Table 4. Performance and technical details of the YOLOv8m, YOLOv8l, YOLOv9c, and YOLOv9e models evaluated on the DUT Anti-UAV dataset [54] (in-distribution data) across various training data configurations. The SynDroneVision dataset is abbreviated as SDV.

YOLO	Layers	Param. (M)	GFLOPs (B)	Training Data		Evaluation on DUT Anti-UAV			FNR ↓	FDR ↓
				SDV (Ours) (synthetic)	DUT Anti-UAV (real)	mAP ↑ @0.25	@0.5	@0.5-0.95		
v8m	295	25.85	79.1	✓	—	0.677	0.639	0.422	0.525	0.103
				—	✓	0.956	0.933	0.669	0.118	<b>0.021</b>
				✓	✓	<b>0.960</b>	<b>0.938</b>	<b>0.686</b>	<b>0.117</b>	0.025
v8l	287	43.61	164.8	✓	—	0.746	0.716	0.468	0.438	0.079
				—	✓	0.922	0.896	0.628	0.149	0.067
				✓	✓	<b>0.963</b>	<b>0.944</b>	<b>0.696</b>	<b>0.105</b>	<b>0.015</b>
v9c	384	25.32	102.3	✓	—	0.700	0.666	0.429	0.474	0.093
				—	✓	0.959	0.935	0.668	0.123	0.023
				✓	✓	<b>0.961</b>	<b>0.941</b>	<b>0.695</b>	<b>0.107</b>	<b>0.022</b>
v9e	1225	58.15	192.7	✓	—	0.767	0.733	0.460	0.405	0.055
				—	✓	0.944	0.915	0.643	0.149	0.042
				✓	✓	<b>0.969</b>	<b>0.955</b>	<b>0.723</b>	<b>0.095</b>	<b>0.009</b>

pect ratios ranging from 1 to 6.67 with an average of 1.91 (see Table 3).

- *Object Scale* – The object scale, or object area ratio, quantifies the proportion of the drone area captured within the image frame relative to the total image area. Object scales within SynDroneVision exhibit a broad distribution, ranging from minimal values between 0.001 (train) and 0.009 (test) to maximum values of 1 (cf. Table 3). However, the general trend leans towards smaller objects, with average values around 0.32 (cf. Figure III, supplementary material). This aspect is particularly crucial, given the prevalence of small drones in practical applications and the inherent complexity associated with their detection. It also aligns with the characteristic features observed in other drone detection datasets. For instance, the DUT Anti-UAV dataset also comprises predominantly small drone object, albeit with an average object scale of 0.013.

## 4. Analysis

This section outlines the evaluation setup and presents key findings regarding the efficiency of SynDroneVision.

### 4.1. Experimental Setup

To assess the effectiveness of SynDroneVision, we conduct a comparative analysis of various YOLO models and associated training configurations. Building upon the latest developments in YOLO architectures, our analysis focuses on YOLOv8 [49] and YOLOv9 [50], with a detailed examination of their respective variants: YOLOv8m, YOLOv8l, YOLOv9c, and YOLOv9e. Considering the significance

of real-world data in validating SynDroneVision’s practical value, we incorporate the DUT Anti-UAV dataset [54] (selected for its characteristic resemblance to SynDroneVision). To ensure a comprehensive evaluation, especially in terms of model robustness, we also include the UAV-Eagle dataset [5] and the Drone Dataset by [2] as out-of-distribution data.

The training procedure for each model involves three distinct strategies: (i) training exclusively on SynDroneVision, (ii) training solely on DUT Anti-UAV, and (iii) a hybrid approach combining both datasets. Each model is trained for 100 epochs with a batch size of 64. For YOLOv9e, a reduced batch size of 32 is employed due to memory limitations. Other hyperparameters follow their default configurations specified in [48]. Model performance is evaluated using the DUT Anti-UAV test set [54], UAV-Eagle [5], and the Drone Dataset by [2]. Key performance indicators include standard object detection metrics such as mean average precision (mAP) at an intersection over union (IoU) threshold of 0.5, and computed over a range of IoU thresholds from 0.5 to 0.95. To account for precision variations in manually generated annotations (cf. [31]), we also consider mAP values at an IoU threshold of 0.25, along with false negative rates (FNRs) and false discovery rates (FDRs). All experiments are performed using the Ultralytics repository [48] on a single NVIDIA Quadro RTX-8000 GPU.

### 4.2. Results

For evaluations on the SynDroneVision test set, refer to the supplementary material. In the following, we provide an assessment of all trained models on real-world data, highlighting SynDroneVision’s practical applicability.

Table 5. Performance of YOLOv9e on the UAV-Eagle dataset [5] and the Drone Dataset by [2] (out-of-distribution data) across different training data configurations. The SynDroneVision dataset is abbreviated as SDV.

Evaluation Data	Training Data		mAP ↑			FNR ↓	FDR ↓
	SDV (Ours) (synthetic)	DUT Anti-UAV (real)	@0.25	@0.5	@0.5-0.95		
UAV-Eagle [5] (real)	✓	—	0.946	0.810	0.340	0.193	0.125
	—	✓	0.947	0.819	0.290	0.165	0.088
	✓	✓	<b>0.980</b>	<b>0.879</b>	<b>0.351</b>	<b>0.136</b>	<b>0.058</b>
Drone Dataset by [2] (real)	✓	—	0.741	0.474	0.179	0.379	0.109
	—	✓	0.812	0.569	0.202	0.251	0.185
	✓	✓	<b>0.822</b>	<b>0.595</b>	<b>0.222</b>	<b>0.213</b>	<b>0.098</b>

**Performance on DUT Anti-UAV.** A comprehensive analysis of YOLO models trained on the hybrid dataset combining SynDroneVision (synthetic data) and DUT Anti-UAV (real-world data) reveals substantial improvements across all performance indicators, particularly when compared to models trained solely on real-world data. Specifically, at an IoU threshold of 0.5, the hybrid training strategy yields an mAP increase of up to 4.8 percentage points over models trained exclusively on DUT Anti-UAV (cf. YOLOv8l, Table 4). This positive effect is even more pronounced for mAP values averaged across multiple IoU thresholds, resulting in improvements up to eight percentage points (cf. YOLOv9e, Table 4). Moreover, a considerable decline in FNRs and FDRs is observed across all models, except for YOLOv8m. The performance enhancements obtained by integrating SynDroneVision with real-world data are particularly amplified in more complex architectures characterized by an increased number of trainable parameters (e.g., cf. YOLOv8m vs. YOLOv8l, Table 4). The most effective performance is achieved with YOLOv9e.

A comparison between models trained exclusively on either SynDroneVision or DUT Anti-UAV also demonstrates distinct performance variations on the DUT Anti-UAV test set. Models trained on DUT Anti-UAV exhibit superior performance compared to those solely trained on SynDroneVision (cf. Table 4). In particular, models trained on SynDroneVision feature increased FNRs, ranging from 0.405 to 0.525 (cf. Table 4), while models trained on DUT Anti-UAV maintain FNRs below 0.149. This deviation is not surprising, considering the persistent challenge associated with the simulation-reality gap. Nevertheless, training exclusively on SynDroneVision still yields promising outcomes, with mAP values exceeding 0.639 at an IoU threshold of 0.5 for all YOLO architectures.

Visual examples of the detection results are provided in the supplementary material (Figure IV).

**Performance on Out-of-Distribution Data.** The evaluation of YOLO models trained on SynDroneVision, both independently and in combination with real-world data, re-

veals notable robustness when exposed to varying data distributions and sources (cf. Table 5). For instance, on the UAV-Eagle dataset, YOLOv9e trained on SynDroneVision achieves nearly equivalent performance to the model trained exclusively on DUT Anti-UAV, with mAP deviations of less than 0.01. It even exceeds the latter by five percentage points in mAP for an IoU threshold range of 0.5 to 0.95. Compared to corresponding validation outcomes on DUT Anti-UAV (see Table 4), the model trained solely on SynDroneVision shows significant improvements across all metrics, while the model trained exclusively on DUT Anti-UAV experiences notable performance declines (except for mAP at 0.25). Integrating SynDroneVision with DUT Anti-UAV during training further enhances performance, leading to even more pronounced improvements across key indicators. This trend is consistent across other YOLO variants, as detailed in the supplementary material (Table VI).

On the Drone Dataset by [2], the model solely trained on SynDroneVision exhibits lower performance in mAP and FNR relative to the DUT Anti-UAV-trained model. However, the performance gap is less pronounced than the one observed in Table 4. Combining both datasets during training also yields significant improvements on the Drone Dataset by [2] across all metrics. An exception is the FDR, where the combination of SynDroneVision and DUT Anti-UAV is not always as effective for other YOLO variants (cf. Table VI, supplementary material).

**Discussion.** Our findings indicate that SynDroneVision is a highly promising dataset with the potential to substantially enhance the performance of DL models for drone detection (in surveillance applications), particularly when combined with real-world data. Furthermore, SynDroneVision seems to contribute to an improved model robustness, offering a clear benefit over exclusive real-world data training. The practical value of SynDroneVision is further highlighted by the promising performance of models trained exclusively on this dataset, considering that these models have not encountered target domain data or any real-world information. With its pixel-precise, automatically generated annotations,

SynDroneVision significantly enhances bounding box localization, while simultaneously reducing data acquisition costs (without compromising performance).

## 5. Conclusion

In this work, we introduced SynDroneVision, a novel and comprehensive synthetic RGB dataset designed to support the development of robust drone detection systems. By providing a detailed analysis of recent YOLO models trained on both SynDroneVision and real-world data, we demonstrated the effectiveness of SynDroneVision in enhancing model accuracy and robustness.

## 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, 13](#)
- [2] Mehmet Çağrı Aksoy, Alp Sezer Orak, Hasan Mertcan Özkan, and Bilgin Selimoğlu. Drone Dataset: Amateur Unmanned Air Vehicle Detection. *Mendeley Data*, V4, 2019. [1, 2, 3, 7, 8, 12, 13, 18, 19](#)
- [3] Alianza Store. Drone Pack. <https://www.unrealengine.com/marketplace/en-US/product/drone-pack>. accessed: 2024-08-26. [4, 5](#)
- [4] Dafni Anagnostopoulou, George Retsinas, Niki Efthymiou, Panagiotis Filntisis, and Petros Maragos. A Realistic Synthetic Mushroom Scenes Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 6282–6289, 2023. [1](#)
- [5] 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. [2, 3, 7, 8, 12, 13, 18, 19](#)
- [6] Antonella Barisic, Frano Petric, and Stjepan Bogdan. Sim2Air - Synthetic Aerial Dataset for UAV Monitoring. *IEEE Robotics and Automation Letters*, 7(2):3757–3764, 2022. [1, 2, 3, 5, 13](#)
- [7] Michael J. Black, Priyanka Patel, Joachim Tesch, and Jin-long Yang. BEDLAM: A Synthetic Dataset of Bodies Exhibiting Detailed Lifelike Animated Motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8726–8737, 2023. [1](#)
- [8] Blender Foundation. Blender. <https://www.blender.org/>, accessed: 2024-08-26. [3, 13](#)
- [9] 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. [2, 3, 12, 13](#)
- [10] Codex Laboratories LLC. Welcome to Colosseum, a Successor of AirSim. <https://github.com/CodexLabsLLC/Colosseum>, accessed: 2024-08-26. [4](#)
- [11] 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, 13, 14](#)
- [12] CropCraft Studios. Ultimate Farming. <https://www.unrealengine.com/marketplace/en-US/product/ultimate-farming>. accessed: 2024-08-26. [5, 15, 17](#)
- [13] Cycles Developers. Cycles. <https://www.cycles-renderer.org/>, accessed: 2024-08-26. [3, 13](#)
- [14] Safaa Dafrallah and Moulay Akhloufi. Malicious UAV Detection Using Various Modalities. *Drone Systems and Applications*, 12:1–18, 2024. [1](#)
- [15] Deelus. Venice - Fast Building. <https://www.unrealengine.com/marketplace/en-US/product/venice-fast-building>. accessed: 2024-08-26. [5, 15, 17](#)
- [16] Tamara R. Dieter, Andreas Weinmann, and Eva Brucherseifer. Generating Synthetic Data for Deep Learning-Based Drone Detection. *AIP Conference Proceedings*, 2939(1):030007, 2023. [1, 4](#)
- [17] Tamara R. Dieter, Andreas Weinmann, Stefan Jäger, and Eva Brucherseifer. Quantifying the Simulation–Reality Gap for Deep Learning-Based Drone Detection. *Electronics*, 12(10), 2023. [2](#)
- [18] Mohamed Elsayed, Mohamed Reda, Ahmed S. Mashaly, and Ahmed S. Amein. Review on Real-Time Drone Detection Based on Visual Band Electro-Optical (EO) Sensor. In *10th International Conference on Intelligent Computing and Information Systems*, pages 57–65, 2021. [1](#)
- [19] Epic Games. Lumen Global Illumination and Reflections. <https://docs.unrealengine.com/5.0/en-US/lumen-global-illumination-and-reflections-in-unreal-engine/>, accessed: 2024-08-26. [4, 5](#)
- [20] Epic Games. Post Process Effects. <https://docs.unrealengine.com/5.3/en-US/post-process-effects-in-unreal-engine/>, accessed: 2024-08-26. [5, 17](#)
- [21] Epic Games. Sun and Sky Actor. <https://docs.unrealengine.com/4.27/en-US/BuildingWorlds/LightingAndShadows/SunSky/>, accessed: 2024-08-26. [5, 17](#)
- [22] Epic Games. Unreal Engine. <https://www.unrealengine.com/en-US/>, accessed: 2024-08-26. [4, 17](#)
- [23] Andrew Svanberg Hamilton. Rural Australia. <https://www.unrealengine.com/marketplace/en-US/product/rural-australia>. accessed: 2024-08-26. [5, 15](#)
- [24] Min Huang, Wenkai Mi, and Yuming Wang. EDGS-YOLOv8: An Improved YOLOv8 Lightweight UAV Detection Model. *Drones*, 8(7), 2024. [2](#)
- [25] I.G.F Iron Gear Factory. Military Drone Pack. <https://www.unrealengine.com/marketplace/en-US/product/military-drone-pack>. accessed: 2024-08-26. [4, 5](#)

- [26] Brian K. S. Isaac-Medina, Matt Poyer, 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. 12
- [27] Sonain Jamil, Muhammad Sohail Abbas, and Arunabha M. Roy. Distinguishing Malicious Drones Using Vision Transformer. *AI*, 3(2):260–273, 2022. 2, 3, 13
- [28] 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. 1, 2, 6, 13, 14
- [29] Andreas Kloukiniotis, Andreas Papandreou, Christos Anagnostopoulos, Aris Lalos, Petros Kapsallas, D.-V. Nguyen, and Konstantinos Moustakas. CarlaScenes: A Synthetic Dataset for Odometry in Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 4519–4527, 2022. 1
- [30] Benedikt Kolbeinsson and Krystian Mikolajczyk. DDOS: The Drone Depth and Obstacle Segmentation Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024. 1
- [31] 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. 1, 2, 7
- [32] Zhaoyang Liu, Limin Wang, Wayne Wu, Chen Qian, and Tong Lu. TAM: Temporal Adaptive Module for Video Recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13688–13698, 2021. 2
- [33] Yaowen Lv, Zhiqing Ai, Manfei Chen, Xuanrui Gong, Yuxuan Wang, and Zhenghai Lu. High-Resolution Drone Detection Based on Background Difference and SAG-YOLOv5s. *Sensors*, 22(15), 2022. 2
- [34] Aboli Marathe, Deva Ramanan, Rahee Walambe, and Ketan V. Kotecha. WEDGE: A Multi-Weather Autonomous Driving Dataset Built from Generative Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 3318–3327, 2023. 1
- [35] Diego Marez, Samuel Borden, and Lena Nans. UAV Detection with a Dataset Augmented by Domain Randomization. In *Geospatial Informatics X*, volume 11398, page 1139807. SPIE, 2020. 1, 2
- [36] Microsoft Research. Welcome to AirSim. <https://microsoft.github.io/AirSim/>, accessed: 2024-08-26. 4
- [37] Mixall Studio. Quadcopter Pack - Drones. <https://www.unrealengine.com/marketplace/en-US/product/quadcopter-pack-drones>. accessed: 2024-08-26. 4, 5
- [38] Adnan Munir, Abdul Jabbar Siddiqui, and Saeed Anwar. Investigation of UAV Detection in Images with Complex Backgrounds and Rainy Artifacts. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 232–241, 2023. 2
- [39] PurePolygons. Downtown West Modular Pack. <https://www.unrealengine.com/marketplace/en-US/product/6bb93c7515e148a1a0a0ec263db67d5b>. accessed: 2024-08-26. 5, 15
- [40] PurePolygons. Modular Building Set. <https://www.unrealengine.com/marketplace/en-US/product/modular-building-set>. accessed: 2024-08-26. 5, 15, 17
- [41] 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. 2, 3, 6, 12, 13
- [42] Denys Rutkovskyi. Factory Environment Collection. <https://www.unrealengine.com/marketplace/en-US/product/factory-environment-collection>. accessed: 2024-08-26. 5, 15, 17
- [43] SilverTm. City Park Environment Collection. <https://www.unrealengine.com/marketplace/en-US/product/city-park-environment-collection>. accessed: 2024-08-26. 5, 15, 17
- [44] Krishnakant Singh, Thanush Navaratnam, Jannik Holmer, Simone Schaub-Meyer, and Stefan Roth Roth. Is Synthetic Data all We Need? Benchmarking the Robustness of Models Trained with Synthetic Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2505–2515, 2024. 1
- [45] Fredrik Svanström, Fernando Alonso-Fernandez, and Cristofer Englund. A Dataset for Multi-Sensor Drone Detection. *Data in Brief*, 39:107521, 2021. 1, 2, 3, 12, 13
- [46] Charalampos Symeonidis, Charalampos Anastasiadis, and Nikos Nikolaidis. A UAV Video Data Generation Framework for Improved Robustness of UAV Detection Methods. In *IEEE 24th International Workshop on Multimedia Signal Processing*, pages 1–5, 2022. 2
- [47] Martin Tran, Jordan Shipard, Hermawan Mulyono, Arnold Willem, and Clinton Fookes. SafeSea: Synthetic Data Generation for Adverse & Low Probability Maritime Conditions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 821–829, 2024. 1
- [48] Ultralytics. YOLOv8.2: Unleashing Next-Gen AI Capabilities. <https://github.com/ultralytics/ultralytics>, accessed: 2024-08-26. 7
- [49] Rejin Varghese and Sambath M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pages 1–6, 2024. 7
- [50] Chien-Yao Wang, I-Hau Yeh, and Hongpeng Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *ArXiv*, abs/2402.13616, 2024. 7

- [51] 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. [2](#), [3](#), [12](#), [13](#)
- [52] Changfeng Yu, Shiming Chen, Yi Chang, Yibing Song, and Luxin Yan. Both Diverse and Realism Matter: Physical Attribute and Style Alignment for Rainy Image Generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12353–12363, 2023. [1](#)
- [53] 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. [2](#), [3](#), [12](#), [13](#)
- [54] 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](#), [3](#), [4](#), [6](#), [7](#), [12](#), [13](#), [18](#)
- [55] 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](#), [3](#), [12](#), [13](#)

## Supplementary Material

### 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 [9,51] consists of 30 videos, each with a resolution of  $1920 \times 1080$  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  $300 \times 168$  to  $3840 \times 2160$  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.* [41] 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 [26]. The average drone size within the dataset is  $136 \times 77$  pixels.

**Det-Fly.** The Det-Fly dataset by Zheng *et al.* [55] focuses on air-to-air visual detection of micro UAVs and comprises 13,271 high-resolution images ( $3840 \times 2160$  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 [5] 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.* [53] introduce UAVData, a dataset designed for visual drone detection, consisting of 13,803 manually recorded and annotated RGB images with a resolution of  $1280 \times 720$  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  $5 \times 23$  to  $720 \times 303$  pixels.

**Halmstad Data.** The Halmstad Dataset, developed by Svanström *et al.* [45], is a multi-sensor dataset for drone detection, 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 [54] 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  $240 \times 160$  to  $5616 \times 3744$  pixels. Object sizes within

Table I. Overview of additional characteristics of publicly available datasets for image-based drone detection. The symbols **X** (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. [9, 51]	✓	✓	drone monitoring	✓	X	✓	✓
Drone Dataset [2]	✓	X	drone detection	✓	X	–	–
MAV-VID [41]	✓	✓	drone detection	✓	X	✓	✓
Det-Fly [55]	✓	X	detection of micro-UAVs	✓	X	–	–
UAV-Eagle [5]	✓	✓	UAV detection in unconstrained environments	✓	X	✓	X
UAVData [53]	✓	✓	UAV detection	✓	X	✓	X
Halmstadt Data [45]	✓	✓	drone detection at airports	✓	X	✓	✓
DUT Anti-UAV [54]	✓	✓	anti-UAV detection	✓	X	✓	✓
VisioDECT [1]	✓	X	detection of unauthorized UAVs	✓	X	✓	✓
Malicious Drones [27]	✓	X	hazardous payload drone detection	✓	X	–	–
S-UAV-T [6] ( <i>synthetic</i> )	✓	X	UAV-to-UAV detection	✓	X	X	✓
Drone-vs-Bird Detection Ch. [11]	✓	✓	distinction between drones and birds	✓	✓	✓	✓
Anti-UAV [28]	X	✓	single UAV tracking	✓	X	✓	✓

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.

**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.

**Malicious Drones.** Jamil *et al.* [27] 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.

**S-UAV-T.** The S-UAV-T dataset by Barisic *et al.* [6] is the only publicly available synthetic dataset for drone – more precisely UAV-to-UAV – detection. The dataset is generated via Blender [8] and the rendering engine Cycles [13], 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  $608 \times 608$  pixels.

**Drone-vs-Bird Detection Challenge.** The Drone-vs-Bird Detection Challenge dataset [11] 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  $720 \times 576$  to  $3840 \times 2160$  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  $32 \times 32$  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.* [28], 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 folder is further divided into training, test, 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 II provides a comprehensive

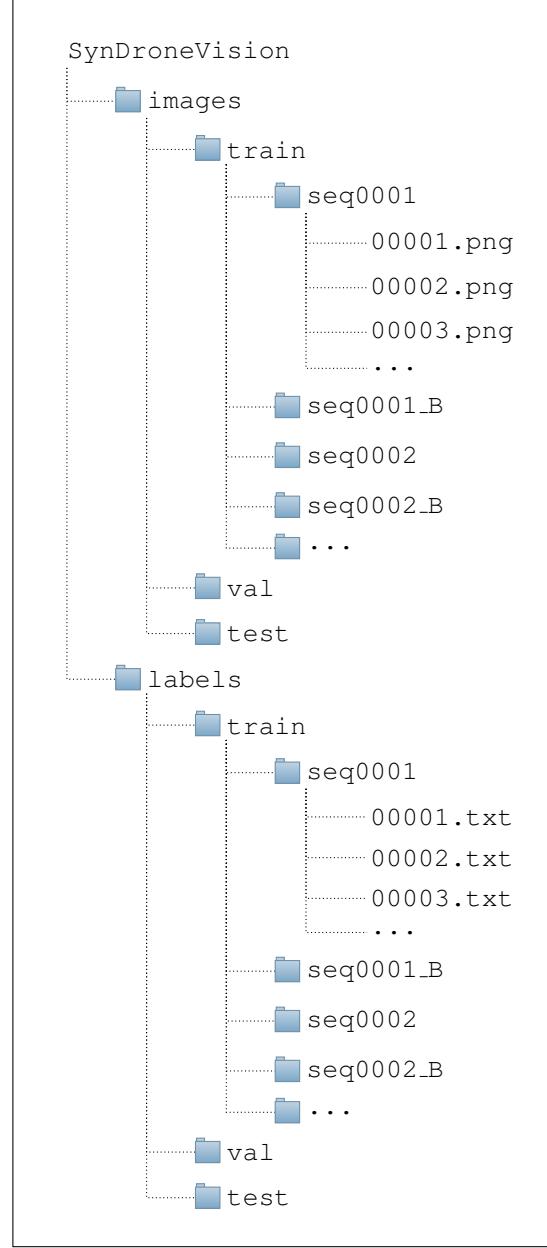


Figure I. Folder configuration of the SynDroneVision dataset.

Table II. 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. <b>Figure II:</b> images 1–6 (rows 1–3).
Venetian City [15]	✓	✓	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. <b>Figure II:</b> images 1–6 (rows 4–5).
Farming Grounds [12]	✓	✓	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. <b>Figure II:</b> images 1–6 (row 6).
Rural Australia [23]	✓	✗	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). <b>Figure II:</b> images 1–6 (row 7).
City Park [43]	✓	✗	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. <b>Figure II:</b> images 1–5 (row 8).
Factory Grounds [42]	✓	✗	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. <b>Figure II:</b> image 6 (row 8), images 1–6 (row 9).
Urban Downtown [39]	✓	✗	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. <b>Figure II:</b> images 1–6 (row 10), images 1–4 (row 11).
Modular Cityscape [40]	✓	✗	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. <b>Figure II:</b> images 5–6 (row 11), images 1–6 (last row)

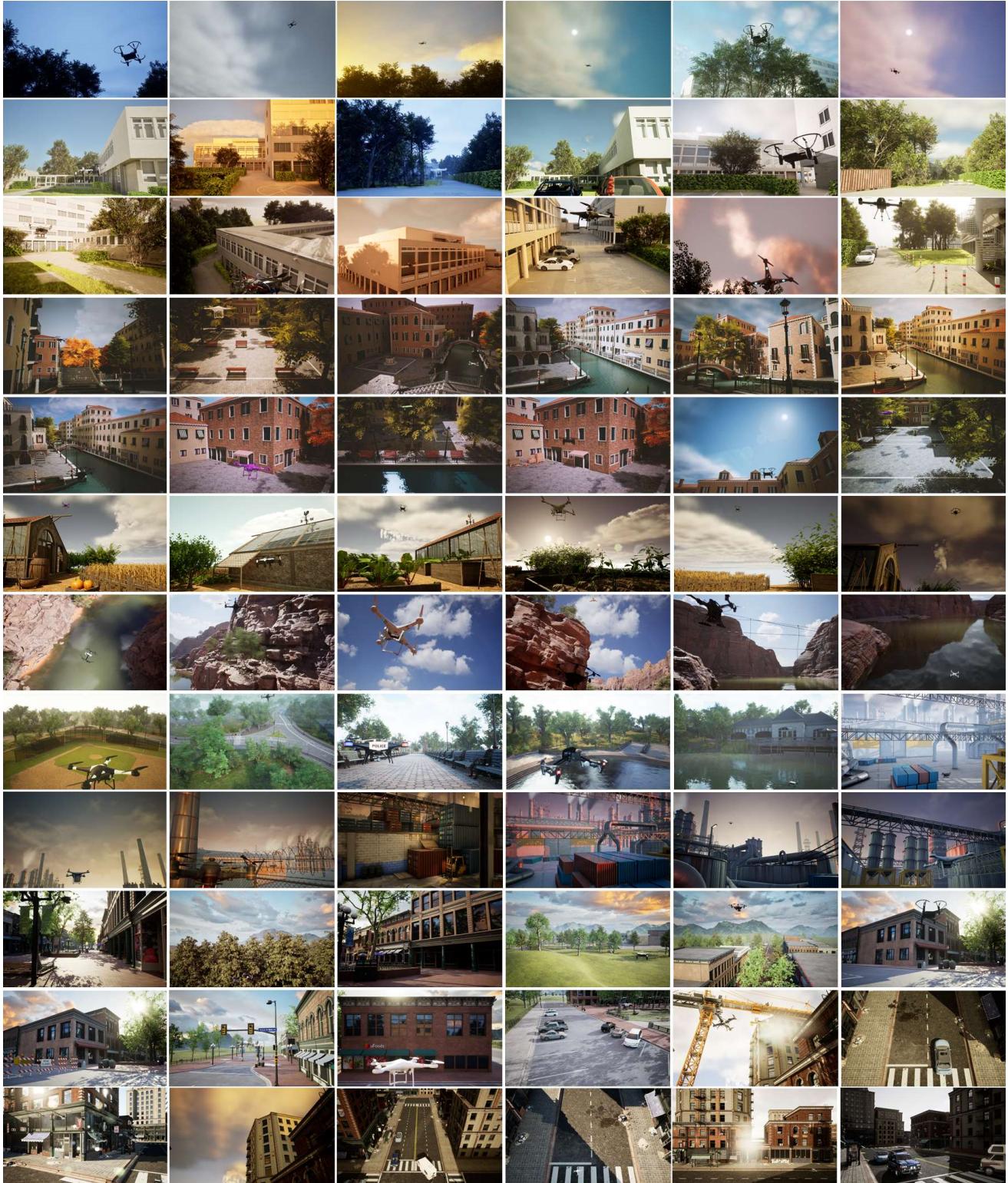


Figure II. Customized camera perspectives and lighting configurations tailored to each environment. The camera fields of view (FOVs) correspond to the following environments (arranged from left to right, top to bottom): University Site (rows 1-3), Venetian City (rows 4-5), Farming Grounds (row 6), Rural 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 III. 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)					
	from	to	min	max	Red		Green		Blue	
					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 [15]	9.40	17.00	6	15	0.100	0.565	0.199	0.739	0.410	0.990
Farming Grounds [12]	7.24	14.25	5	13	0.119	0.599	0.188	0.836	0.361	1
Modular Cityscape [40]	6.87	17.30	15	70	0.125	0.686	0.225	0.687	0.719	1

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 [21] and the Post Process Volume [20], essential tools within the Unreal Engine [22].

**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 III 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:

Table IV. Post Process Volume settings.

Environment	Temp		Tint	
	min	max	min	max
University Site	4,400	12,000	0	0.30
Venetian City [15]	3,840	15,000	0	0.25
Farming Grounds [12]	4,588	4,588	0.05	0.05
Modular Cityscape [40]	4,770	9,500	-0.02	0.03

- *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.

**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 [42] and City Park [43], 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

Table V. Technical configuration details for Unreal projects.

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

\* not enabled for University Site

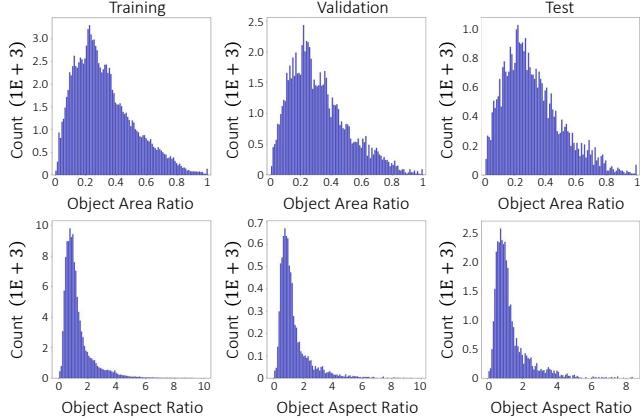


Figure III. Object area and object aspect ratio distribution in the SynDroneVision dataset across training, validation, and test splits.

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.

## D. Analysis Details

### D.1. Detection Examples

Figure IV presents selected examples from the DUT Anti-UAV [54], the UAV-Eagle [5], and the Drone Dataset

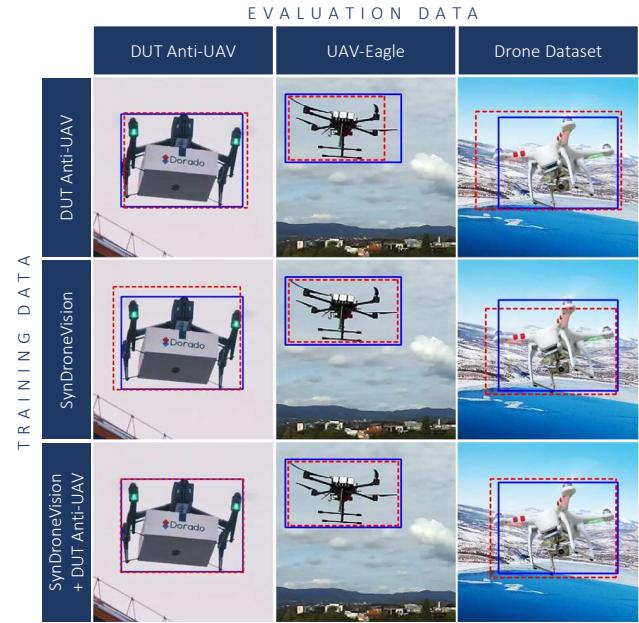


Figure IV. YOLOv9e detections on the DUT Anti-UAV test set [54] (1st column), the UAV-Eagle dataset [5] (2nd column), and the Drone Dataset by [2] (last column) demonstrating improved bounding box precision for models trained on both SynDroneVision and DUT Anti-UAV data (last row). Predictions (red dashed line) are marked alongside ground truth (solid blue line).

by [2], along with their corresponding detection outcomes obtained using YOLOv9e. The effectiveness of the detection results is compared across all three training strategies, 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.

### 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



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 [5] 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) (synthetic)	DUT Anti-UAV (real)	@0.25	@0.5	@0.5-0.95		
UAV-Eagle [5] (real)	v8m	✓	—	0.944	0.771	0.293	0.201	0.169
		—	✓	0.935	0.823	0.302	0.199	<b>0.063</b>
		✓	✓	<b>0.961</b>	<b>0.849</b>	<b>0.350</b>	<b>0.136</b>	0.089
	v8l	✓	—	0.951	0.786	0.304	0.217	0.125
		—	✓	0.920	0.725	0.217	0.180	0.224
		✓	✓	<b>0.979</b>	<b>0.869</b>	<b>0.368</b>	<b>0.126</b>	<b>0.074</b>
	v9c	✓	—	0.926	0.770	0.289	0.216	0.163
		—	✓	0.922	0.799	0.275	0.219	<b>0.077</b>
		✓	✓	<b>0.975</b>	<b>0.859</b>	<b>0.353</b>	<b>0.141</b>	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	<b>0.113</b>
		✓	✓	<b>0.824</b>	<b>0.613</b>	<b>0.232</b>	<b>0.196</b>	0.114
	v8l	✓	—	0.768	0.515	0.193	0.389	<b>0.076</b>
		—	✓	<b>0.800</b>	0.552	0.199	0.263	0.227
		✓	✓	0.799	<b>0.603</b>	<b>0.227</b>	<b>0.216</b>	0.116
	v9c	✓	—	0.737	0.530	0.199	0.401	<b>0.073</b>
		—	✓	0.806	0.556	0.206	0.292	0.134
		✓	✓	<b>0.825</b>	<b>0.606</b>	<b>0.224</b>	<b>0.198</b>	0.126

an IoU threshold of 0.25. In analogy to YOLOv9e, the best performance is achieved when combining both datasets during 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 perfor-

mance 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.

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) (synthetic)	DUT Anti-UAV (real)	mAP ↑ @0.25	@0.5	@0.5-0.95	FNR ↓	FDR ↓
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