DDOS: The Drone Depth and Obstacle Segmentation Dataset

Benedikt Kolbeinsson
Imperial College London
bk915@imperial.ac.uk

Krystian Mikolajczyk
Imperial College London
k.mikolajczyk@imperial.ac.uk

https://huggingface.co/datasets/benediktkol/DDOS

Abstract

The advancement of autonomous drones, essential for sectors such as remote sensing and emergency services, is hindered by the absence of training datasets that fully capture the environmental challenges present in real-world scenarios, particularly operations in non-optimal weather conditions and the detection of thin structures like wires. We present the Drone Depth and Obstacle Segmentation (DDOS) dataset to fill this critical gap with a collection of synthetic aerial images, created to provide comprehensive training samples for semantic segmentation and depth estimation. Specifically designed to enhance the identification of thin structures, DDOS allows drones to navigate a wide range of weather conditions, significantly elevating drone training and operational safety. Additionally, this work introduces innovative drone-specific metrics aimed at refining the evaluation of algorithms in depth estimation, with a focus on thin structure detection. These contributions not only pave the way for substantial improvements in autonomous drone technology but also set a new benchmark for future research, opening avenues for further advancements in drone navigation and safety.

1. Introduction

Fully autonomous drones are poised to revolutionize a multitude of sectors, including remote sensing [3, 16, 17, 26, 31, 35], package delivery [4, 13], emergency services, and disaster response [2, 9–11, 28, 29]. While manually controlled drones have been effectively employed in specific sectors, the advent of fully autonomous drones is poised to unlock an array of novel applications, enhancing efficiency and expanding capabilities. However, realizing this potential is contingent upon the ability of drones to navigate safely and autonomously, which in turn requires a precise understanding of their environment. Current datasets for training drone navigation systems are inadequate, particularly in representing challenging scenarios such as the detection of thin structures like wires and cables, and operation under diverse weather conditions [25]. This deficiency highlights the need for a dataset that provides a comprehensive representation of the environment, enabling accurate semantic segmentation and depth estimation across a wide range of objects and conditions.

To address this gap, we introduce the Drone Depth and Obstacle Segmentation (DDOS) dataset, a novel resource designed to significantly enhance the training of autonomous drones. DDOS stands out for its dual emphasis on depth and semantic segmentation annotations, with a particular focus on the precise identification of thin structures (a critical but often overlooked aspect in existing datasets). By incorporating advanced computer graphics and rendering techniques, DDOS generates synthetic aerial images that mirror the complexity of real-world environments, encompassing a variety of settings and weather conditions ranging from clear skies to adverse weather scenarios such as rain, fog, and snowstorms.

Our objectives with the DDOS dataset are twofold: firstly, to provide a richly annotated resource that reflects the diversity of scenarios encountered by drones, with a particular focus on thin structures and adverse weather conditions. Secondly, to enable the development and evaluation of algorithms that significantly improve the safety, reliability, and operational efficiency of autonomous drones. By achieving these objectives, we aim to bridge the gap in existing datasets and facilitate the advancement of drone technology to meet the demands of real-world applications.

We present a thorough analysis of DDOS which explores key characteristics including class density, flight dynamics, and spatial distribution, providing a granular understanding of its composition and capabilities. Through comparative analysis with existing datasets, we highlight DDOS's contributions such as incorporating numerous thin and ultra-thin structures with accurate depth and segmentation labels, as well as diverse weather conditions. Furthermore, we propose new drone-specific metrics designed to accurately evaluate class-specific depth estimation performance. These metrics are tailored to reflect the operational realities of drone applications, offering a refined lens through which to assess algorithmic performance and contributing to the broader goal of advancing drone technology and safety.

	USF [7]	NE-VBWD [33]	TTPLA [1]	PIM [36]	UAVid [21]	AeroScapes [27]	Ruralscapes [23]	Mid-Air [12]	TartanAir [38]	SynthWires [22]	SynDrone [30]	DDOS (ours)
Data type	Real	Real	Real	Real	Real	Real	Real	Synthetic	Synthetic	Synthetic	Synthetic	Synthetic
Flight Trajectories	86	41	80	NA	30	141	20	54	1037	154	8	340
Frames	6 k	15 k	1 k	159	300	3 k	51 k	119 k†	1 M	68 k	72 k	34 k
Labeled frames	3 k	91	1 k	159	300	3 k	1 k*	119 k†	1 M	68 k	72 k	34 k
Resolution	640×480	6576×4384	3840×2160	1280×960	3840×2160	1280×720	3840×2160	1382×512	640×480	640×480	1920×1080	1280×720
Frame rate	25 Hz	2 Hz	30 Hz	-	0.2 Hz	-	50 Hz	25 Hz	-	-	25 Hz	10 Hz
Environment	Town	Town/Nature	Pylons	Pylons	Town/Nature	Various	Town/Nature	Nature	Various	Various	Town	Town/Nature
Camera motion	Handheld	Helicopter	Drone	Drone	Drone	Drone	Drone	Drone	Random	Drone	Drone	Drone
Altitude	2 m	+300 m	-	-	50 m	5 – 50 m	-	-	-	-	20, 50, 80 m	1 – 25 m
Weather variations	No	No	No	No	No	No	No	Yes	No	No	No	Yes
Camera pose	No	No	No	No	No	No	No	Yes	Yes	No	Yes	Yes
Optical flow	No	No	No	No	No	No	No	No	Yes	No	No	Yes
Depth map	No	Sparse	No	No	No	No	No	Yes	Yes	No	Yes	Yes
Segmentation	Wires only	Wires only	Yes	No	Yes	Yes	Yes	Yes	No‡	Wires only	Yes	Yes
Thin structures	Yes	Yes	Yes	Patches	No	Yes	No	No	No‡	Yes	No	Yes
Mesh structures	No	No	Rough	Patches	No	Large only	No	No	No‡	No	No	Yes

Table 1. **Comparison between our DDOS dataset and related datasets.** *Ruralscapes also includes automatically generated labels for the remaining 98% of the dataset. †Mid-Air includes additional variations for the same trajectory. ‡TartanAir does not include labeled segmentation classes (i.e. each object is assigned to a random unlabeled class, with variations of the same object type in different classes).

Finally, we present baseline results obtained by applying state-of-the-art algorithms to the DDOS dataset, establishing a benchmark for future research in thin object detection. We examine the strengths and limitations of current methodologies, particularly highlighting their notable failure to accurately predict the depth of thin structures. This analysis emphasizes significant opportunities for refinement and innovation within this domain.

To summarize, our main contributions are:

- **DDOS Dataset**: We present the Drone Depth and Obstacle Segmentation (DDOS) dataset, a comprehensive resource developed to significantly improve the training of autonomous drones through extensive depth and semantic segmentation annotations, with a special focus on accurately identifying thin structures.
- **Statistical Analysis and Dataset Comparison**: We provide a thorough examination of the DDOS dataset, highlighting its unique attributes such as class distributions, spatial distribution, and flight dynamics. Our analysis is enriched by a detailed comparative study, positioning DDOS in the broader context of existing datasets and underscoring its distinctive value in addressing specific challenges in drone navigation.
- **Drone-Specific Metrics**: Novel drone-specific metrics are introduced, tailored to the nuances of drone applications, particularly in the evaluation of depth accuracy. These metrics offer a refined and specialized framework for assessing algorithmic performance.
- **Baseline Results and Discussion**: We present baseline results from applying state-of-the-art algorithms to the DDOS dataset, establishing benchmarks for thin object detection research. Our discussion identifies a critical shortfall in existing depth estimation methods, emphasizing the need for future advancements.

2. Related Work

The scarcity of high-quality drone datasets hampers autonomous drone training. This section reviews relevant datasets, evaluating their strengths and weaknesses in regards to training autonomous drones. These evaluations are summarized in Table 1.

2.1. Driving datasets

The KITTI [15, 24], Cityscapes [8], nuScenes [6], and Waymo [34] datasets, essential in computer vision for autonomous driving, fall short in addressing drone-specific requirements. KITTI's concentration on road scenes lacks the aerial views and diverse thin structures crucial for drone navigation. Similarly, Cityscapes, nuScenes, and Waymo fail to capture the unique aerial perspectives and the slender objects like wires and cables vital for drone safety. The absence of these aerial viewpoints and the limited representation of thin structures mean that models trained on these datasets are not fully equipped to meet the challenges of drone-based navigation.

2.2. Wire detection datasets

Several datasets have been specifically designed to tackle the challenge of wire detection, given its critical importance for ensuring the safety of low-flying drones.

The USF dataset [7] and NE-VBWD [33] are pivotal resources dedicated to wire detection, offering a unique perspective on the challenges of identifying thin structures in aerial imagery. The USF dataset, while extensive, is limited by its image quality and the accuracy of its wire annotations, which are not pixel-accurate and often overlook the real-world curvature of wires, instead defining them as straight lines. This simplification fails to capture the complexity of wire shapes in various environments, undermining the dataset's utility for training models to detect thin structures accurately. NE-VBWD, although a more recent addition, offers pixel-wise annotations and distance information, focusing on long-range wire detection. However, its suitability for drone applications is limited due to its emphasis on wires located at distances more relevant to manned aircraft, thus diminishing its relevance for low-altitude drone operations where proximity to wires is a critical safety concern.

TTPLA [1] and PIM [36] also contribute to the field by focusing on transmission towers and power lines, with TTPLA utilizing drone imagery but lacking depth information, and PIM providing small image patches for wire detection without offering semantic segmentation. These datasets, while enriching the domain with specific insights into wire and tower detection, similarly fall short in addressing the broad needs of autonomous drone navigation, such as a diverse range of thin structures, depth mapping, and environmental conditions beyond the mere presence of wires.

2.3. Drone datasets

UAVid [21], AeroScapes [27], and Ruralscapes [23] serve as general drone datasets. They provide a broader view of urban and rural landscapes from a drone's perspective, including various object classes for semantic segmentation. Despite their wider scope, these datasets still lack sufficient emphasis on thin structures, such as wires, which are crucial for the safe navigation of drones in complex environments.

SynthWires [22] utilizes a different approach by overlaying synthetic wires over real-world images from drones. This method enhances the variety of wire scenarios available for training, although the absence of depth information limits the dataset's applicability for comprehensive 3D navigation and obstacle avoidance training.

In enhancing the dataset landscape for drone navigation research, Mid-Air [12], TartanAir [38], and SynDrone [30] represent significant contributions as synthetic datasets offering voluminous labeled training samples. These datasets play a pivotal role in simulating a diverse array of flight dynamics and environmental conditions, providing essential assets such as precise depth maps and camera poses critical for the advancement of sophisticated drone navigation algorithms. Despite their value, these datasets exhibit certain limitations that restrict their comprehensive utility in fully leveraging the potential of synthetic data generation.

One notable shortfall is their failure to encapsulate a complete spectrum of flight scenarios, particularly those involving close encounters, aggressive maneuvering, and very low-altitude flying. Such scenarios, while perilous to execute in real-world settings, are quintessential for preparing drones to navigate through complex, unpredictable environments. Synthetic datasets, with their capacity for controlled simulation, are uniquely positioned to safely incorporate these high-risk flight patterns, thereby enriching the training regime without endangering equipment or safety.

Moreover, while synthetic datasets offer the advantage of generating pixel-perfect segmentation and precise depth measurements, especially for thin structures – attributes unattainable with conventional data collection methods – they fall short in representing thin structures like wires, cables, and fences. These elements are critical for ensuring the navigational reliability of drones in densely populated or structurally complex areas. The absence of such objects in the datasets underscores a missed opportunity to leverage some of the benefits of synthetic data generation.

Our proposed dataset, DDOS, is designed to surpass the limitations of existing datasets in wire detection and drone navigation. It provides detailed representations of thin structures and a wide array of other entities, incorporating weather variability and extensive drone motion. Its synthetic foundation enables simulations of close encounters with objects, typically unsafe in reality, enhancing the dataset's utility and realism for drone training.

3. Dataset Features

We introduce the DDOS dataset, specifically designed for the training of autonomous drones, utilizing synthetic data generation to compile 340 unique drone flights. This dataset is characterized by its comprehensive coverage of various weather conditions, from clear skies to snowstorms, and includes high-risk scenarios such as close encounters and minor collisions. These scenarios, crucial for drone training, are typically too hazardous to replicate in real-world settings. The dataset is notable for its provision of pixel-level precision in semantic segmentation and depth information, particularly for challenging objects such as wires, cables, and fences, thus offering a photo-realistic simulation of environments drones are likely to encounter.

Each flight within the DDOS dataset consists of 100 frames, culminating in a total of $34\,000$ frames across the dataset. This substantial volume of data supports detailed analysis and algorithm training. The dataset emphasizes thin structures, which present significant navigational challenges, thereby serving as a critical resource for the development of algorithms that require precise segmentation and depth estimation capabilities in complex aerial scenarios. Accompanying the high-resolution images captured by a monocular front facing camera are depth maps, semantic segmentation masks, optical flow data, and surface normals. These components are provided at a resolution of 1280×720 pixels, with depth maps covering a range from 0 to $100\,\mathrm{m}$. Additionally, the dataset incorporates exact drone pose, velocity, and acceleration data for each frame.

The DDOS dataset is systematically divided into training, validation, and testing subsets, consisting of 300, 20, and 20 flights, respectively. It features pixel-wise segmentation masks for ten distinct classes, enabling in-depth analysis of various obstacles and environmental elements. Fig-

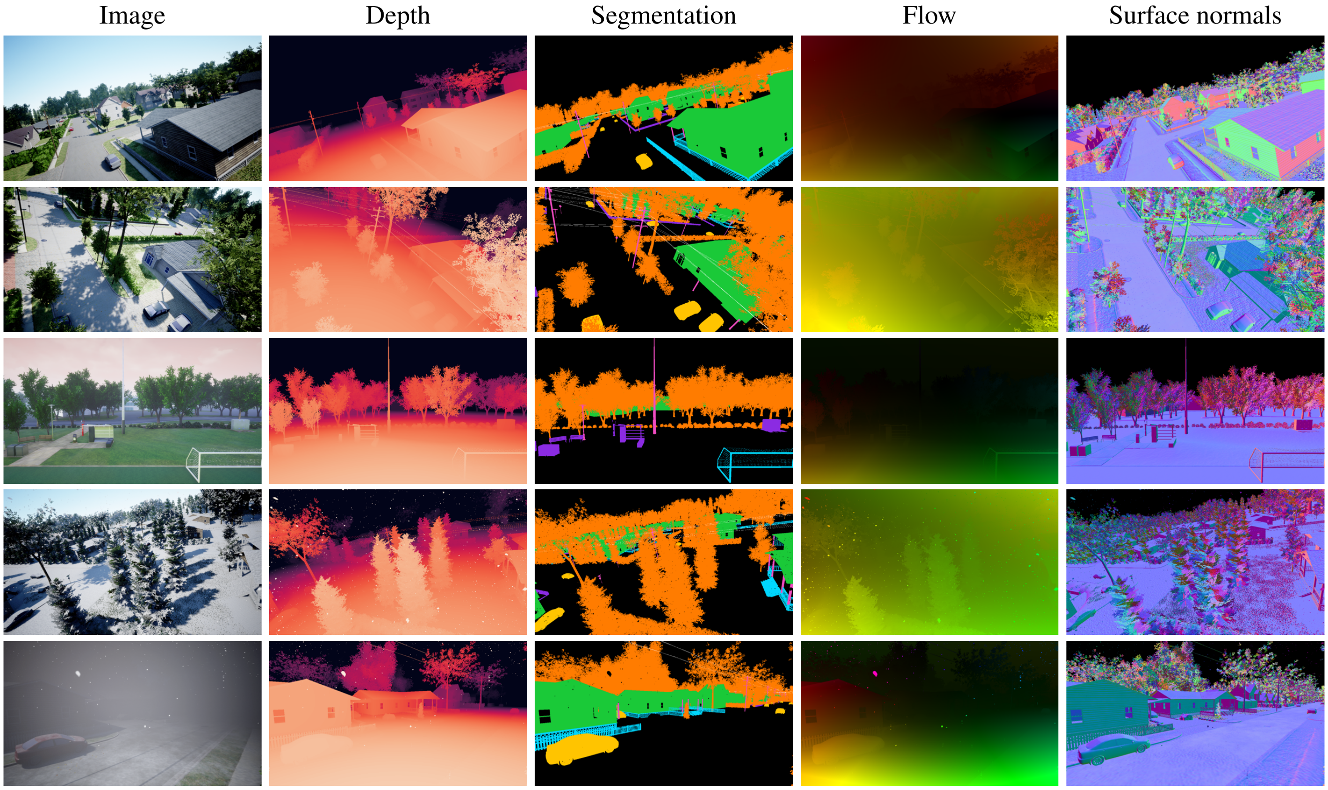


Figure 1. **Examples from our DDOS dataset.** This figure showcases an overview of the DDOS dataset's multifaceted annotations. It includes RGB images from drone flights, depth maps (0–100 m), pixel-wise semantic segmentation, optical flow and surface normals, illustrating the dataset's richness and diversity.

ure 1 displays select examples from the dataset, demonstrating the diversity of classes represented. More examples are available in Appendix B. The methodological approach to dataset generation and the classification scheme are further elaborated in Section 4, providing insight into the dataset's design choices and structure.

4. Data Generation

DDOS is generated using AirSim [32], an open-source drone simulator. DDOS is composed of two environments that mimic real-world scenarios. The first environment resembles a small suburban town, featuring dense trees and numerous power lines, replicating the challenges faced during drone flights in residential areas. The second environment represents a park setting, incorporating elements such as a football field with floodlights, a beach volleyball court, dense trees as well as office buildings. These environments collectively offer diverse obstacles and structures, allowing researchers to develop and evaluate algorithms capable of addressing the complexities associated with different real-world environments. By encompassing characteristics like dense tree coverage, power lines, and varying weather conditions, the dataset provides a comprehensive platform for advancing obstacle segmentation and depth estimation algorithms for safe and effective drone flights.

Flight trajectories To construct each flight trajectory, a random starting location (x_0, y_0, z_0), within the environment bounds is selected. Subsequently, multiple intermediate target points (x_t, y_t, z_t) are generated within predefined relative bounding boxes, dictating the areas to which the drone navigates. Flight characteristics, are varied across different flights, providing diversity in the dataset. During each flight, observations are recorded at a rate of 10 Hz for a duration of 10 seconds. These observations encompass a rich set of data, including images, depth maps, pixel-wise object segmentation, optical flow, and surface-normals.

Collision avoidance In order to promote relatively safe flight paths, we developed a dynamic obstacle detection algorithm to modify intermediate targets in response to potential collision risks. This algorithm utilizes the most recent ground truth depth map obtained during the recorded flight observations. By empirically determining a threshold, objects that are deemed too close trigger updates to the intermediate targets. The updated targets are strategically adjusted based on the detected obstacle's location, causing the drone to navigate away from the identified collision risk.

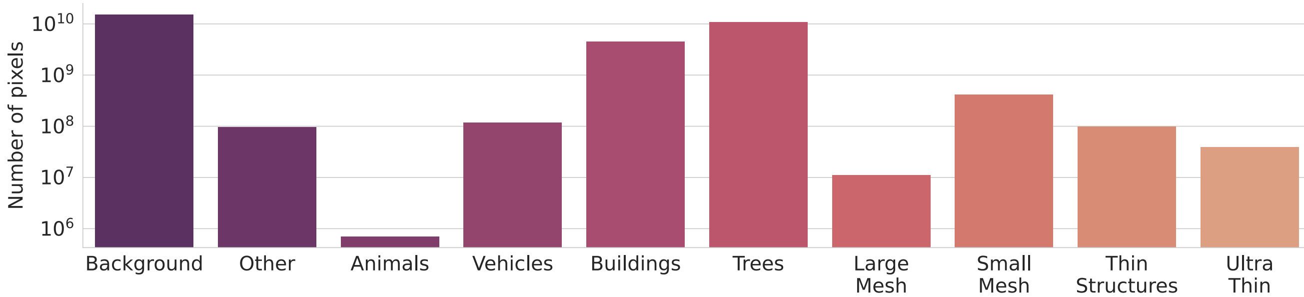


Figure 2. **Distribution of class labels within DDOS.** DDOS effectively captures the presence of various thin object classes, which are characterized by a relatively sparse distribution of pixels within each image. Despite their limited pixel coverage, these thin object classes are well-represented in DDOS, ensuring comprehensive coverage and enabling robust training and evaluation of algorithms specifically designed to address the challenges posed by such objects.

This obstacle avoidance approach is not flawless, especially when dealing with thin structures, occasional collisions resulting in crashes still occur. In such cases, the observations associated with the crash event are discarded and the flight process is restarted to ensure data integrity. It is important to note, the collision avoidance mechanism is purposefully designed to be lax, as near misses and even minor crashes can offer valuable data points for training purposes.

Post-processing To uphold the overall integrity of the dataset and exclude instances of undesired behavior, additional validation criteria are applied after flight generation. These criteria serve to filter out scenarios where the drone becomes stuck or encounters unusual situations, such as becoming entangled in trees. By incorporating these post-flight validation steps, the dataset ensures that the collected observations reflect reliable and meaningful flight behaviors, enabling robust algorithm training, and evaluation.

Data augmentation We do not augment the dataset with additional transformations or modifications, such as chromatic aberration, added lens flares, corruption, or noise, during the data collection process. The decision to exclude these augmentation techniques at the initial phase ensures that the dataset remains in its original state, preserving the inherent characteristics and properties of the collected data. Instead, we provide the flexibility to incorporate these augmentation techniques at a later stage, if deemed necessary, during algorithm development and evaluation.

Weather DDOS encompasses diverse environmental and weather conditions, including sunny, dusk, and brightly lit night scenes, along with rain, fog, snow, and changes due to wet surfaces and snow cover. These conditions challenge vision-based algorithms with reduced visibility and altered surface characteristics, such as increased reflectivity from snow and glare from wet roads, complicating object detection and scene analysis. Including these varied scenarios is essential for developing models that adapt and perform consistently in all real-world settings.

Classes Objects are systematically classified based on their significance for drone navigation. *Ultra Thin* encompasses wires and cables; *Thin Structures* includes poles and signs; *Small Mesh* pertains to fences and nets; and *Large Mesh* covers objects such as transmission towers that permit drone passage. Additionally, *Trees*, *Buildings*, *Vehicles*, and *Animals* are categorized based on straightforward characteristics. The *Other* class encompasses diverse objects like bus stops, post boxes, chairs, and tables. *Background* refers to elements such as the ground and sky, providing context within the scene.

5. Dataset Statistics

In this section, we provide a comprehensive analysis of key properties inherent in the DDOS dataset. Figure 2 illustrates the distribution of annotations across diverse classes within DDOS. Significantly, the dataset adeptly captures and represents various classes of thin structures, even when these objects occupy a relatively small number of pixels in each image. This nuanced representation ensures that DDOS offers a substantial and well-balanced dataset for thin object classes. This richness in diversity is paramount for facilitating thorough analysis, robust algorithm training, and effective evaluation, particularly in addressing the challenges associated with thin structures in real-world scenarios. The carefully crafted distribution of classes within DDOS contributes to its utility as a reliable benchmark for advancing the capabilities of algorithms designed for thin structure detection and segmentation.

In our continued investigation, we analyze the pitch and roll angles observed during flight sessions. As depicted in Figure 3, there is a wide range of pitch and roll angles, indicating significant variations in the drone's orientation across the dataset. Despite the drone's primary forward mo-

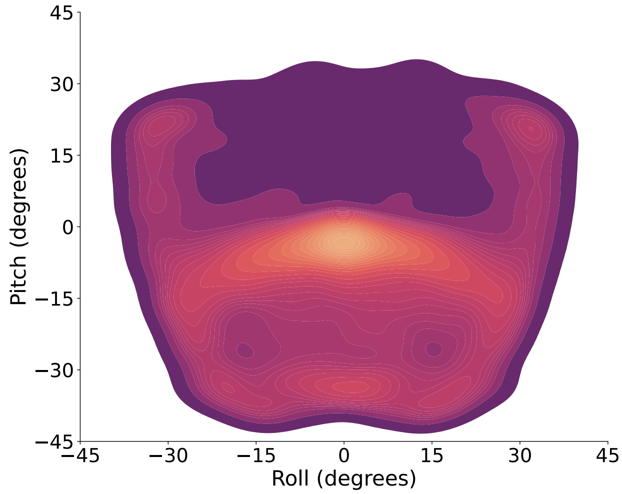


Figure 3. **Distribution of pitch and roll angles.** The colors represent the intensity levels, with warmer colors indicating higher occurrences. Flight characteristics vary between each flight, as highlighted by the diverse pitch and roll degrees. The pitch is negative when the drone is accelerating forward and positive when braking or to go backwards. Emergency braking is often accompanied with a sharp turn, either to the left or to the right.

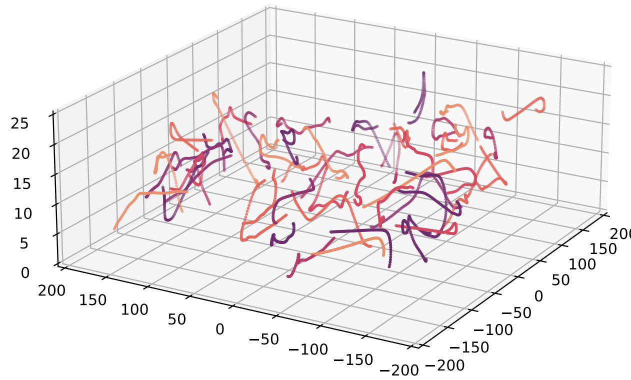


Figure 4. **Illustrated flight paths.** The figure presents a collection of 50 randomly selected flight paths conducted within the same environment. The paths exhibit significant variations in trajectory, highlighting the diverse nature of drone flights.

tion, the angles demonstrate a notable diversity. This variety in orientation provides valuable perspectives for evaluating algorithms under different flight conditions. The broad distribution of pitch and roll angles emphasizes the DDOS dataset's ability to mimic real-world flying scenarios, where drones encounter various orientations. This characteristic enhances the dataset's utility for training and evaluating algorithms to ensure consistent performance amidst the orientation challenges that drones face in actual flights.

To gain an intuitive understanding of the spatial distribution of flight paths within an environment, we visually present a subset of the recorded trajectories in Figure 4. The depicted flight paths showcase a diverse array of patterns, ranging from sharp turns and straight lines to curved trajectories. These variations authentically capture the complexity and dynamic nature of the simulated environments. Furthermore, an overhead view of the relative flight paths, presented in Figure 5, offers a normalized perspective with a common starting point and direction. This visualization emphasizes the diverse flight trajectories and patterns observed across individual flights, providing a comprehensive overview of the spatial dynamics inherent in DDOS. Such a representation is instrumental in offering insights into the intricate navigation challenges that algorithms must address, reinforcing the dataset's efficacy in training and evaluating models under diverse and realistic conditions.

Expanding our analysis, we explore the distributions of altitude and speed during the flights, along with the distribution of depth recorded in the depth maps, as illustrated collectively in Figure 6. Examining the altitude distribution reveals that the drone operates at varying heights, encompassing low-level flights near the ground to higher altitudes. The distribution of speed elucidates a spectrum of velocities encountered during the flights, showcasing diverse flight behaviors and maneuvering speeds. Moreover, the depth distribution offers insights into the range and distribution of depth values recorded in the depth maps, shedding light on the variations in perceived depth across the dataset.

6. Depth Metrics

We propose a novel set of depth metrics specifically tailored for drone applications, namely the absolute relative depth estimation error for each distinct class. To illustrate, we introduce the absolute relative depth error metric for the *Ultra Thin* class within the DDOS dataset. This metric quantifies the accuracy of depth estimation specifically for objects classified as *Ultra Thin* in the DDOS dataset.

$$\text{AbsRel}_{\text{ultra thin}} = \frac{1}{N_{\text{ultra thin}}} \sum_{i=1}^{N_{\text{ultra thin}}} \left| \frac{d_i - \hat{d}_i}{d_i} \right| \tag{1}$$

Here, $\text{AbsRel}_{\text{ultra thin}}$ represents the absolute relative depth estimation error for the *Ultra Thin* class. $N_{\text{ultra thin}}$ denotes the total number of samples (pixels) in the *Ultra Thin* class, while d_i and $\hat{d}_i$ represent the ground truth depth and estimated depth for the i-th pixel sample, respectively. The formula calculates the average absolute relative difference between the ground truth and estimated depths for all samples in the *Ultra Thin* class. Trivially, extending this ap-

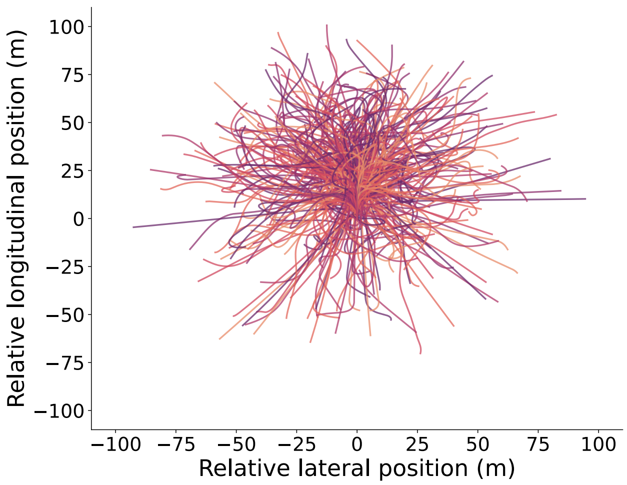
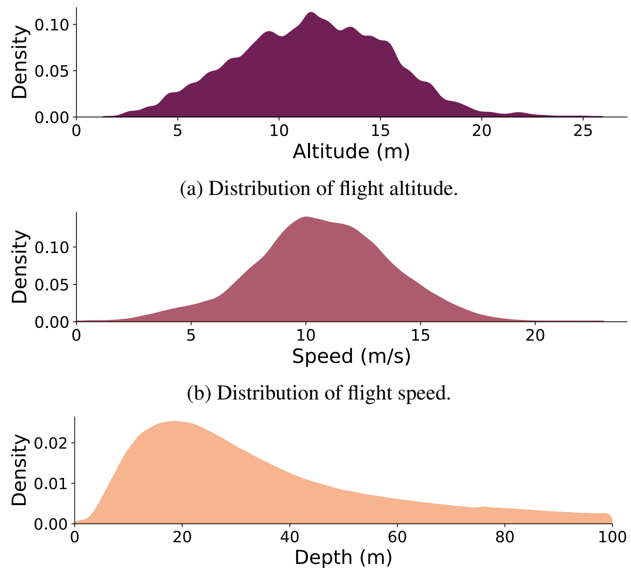


Figure 5. **Overhead view of relative flight paths with a normalized starting point.** In this visualization the starting location and direction have been normalized to highlight the various relative shapes of the flight paths. The actual starting locations are randomly initialized, as shown in Figure 4.



(a) Distribution of flight altitude.

(b) Distribution of flight speed.

(c) Distribution of depth.

Figure 6. **Distributions of altitude, speed, and depth.** The distributions show variation across flights. Depth over 100 m is ignored.

proach to all classes, the general formula for class-specific depth metrics becomes:

$$\text{AbsRel}_{\text{class}} = \frac{1}{N_{\text{class}}} \sum_{i=1}^{N_{\text{class}}} \left| \frac{d_i - \hat{d}_i}{d_i} \right| \tag{2}$$

Assessing class-specific absolute relative depth errors reveals how well depth estimation algorithms perform, especially for intricate structures like wires and cables. This method offers a detailed evaluation, highlighting how algorithms manage the challenges unique to various structures seen from drone viewpoints. The motivation for this nuanced approach stems from the recognition that traditional metrics fail to adequately represent difficult-to-detect obstacles, such as wires, due to their low pixel count. A thorough investigation into these aspects is essential to accurately gauge the efficacy and robustness of vision systems.

7. Baselines

We use a set of commonly-used depth metrics to evaluate the effectiveness of the baselines. These metrics include fundamental measures such as accuracy under the threshold ($\delta_i < 1.25^i$, $i = 1, 2, 3$), which assesses the model's performance within proximity thresholds. Additionally, we use mean absolute relative error (AbsRel), mean squared relative error (SqRel), root mean squared error (RMSE), root mean squared log error (RMSElog), mean log10 error (log10) and scale-invariant logarithmic error (SILog).

Moreover, in pursuit of a more nuanced evaluation, we leverage our newly proposed suite of metrics known as mean absolute relative class error metrics ($AbsRel_{\text{class}}$). This suite is tailored to assess the performance of our methods at a finer class level, offering a more detailed understanding of their capabilities.

We utilize three different baselines, BinsFormer [19], SimIPU [18] and DepthFormer [20]. BinsFormer proposes a novel framework for monocular depth estimation by formulating it as a classification-regression task, employing a transformer [37] decoder to generate adaptive bins [5]. SimIPU introduces a pre-training strategy for spatial-aware visual representation, utilizing point clouds for improved spatial information in contrastive learning. DepthFormer addresses supervised monocular depth estimation by leveraging a transformer for global context modeling, incorporating an additional convolution branch, and introducing a hierarchical aggregation module.

When evaluated using standard depth metrics, the baselines exhibit satisfactory performance, as shown in Table 2. However, using our class-specific depth metrics, shown in Table 3 and depicted in Figure 7, unveils substantial challenges in achieving accurate depth estimations for certain object classes. Specifically, the *Ultra Thin* category is exceptionally challenging, with all tested methods failing to provide accurate depth estimations.

These findings highlight the importance of developing methodologies that are specifically tailored to enhance depth estimation accuracy for ultra-thin structures, particularly in drone-based applications. Future research should focus on addressing these challenges, aiming to enhance the precision and reliability of depth estimations for these challenging scenarios.

Model	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	AbsRel $\downarrow$	RMSE $\downarrow$	log10 $\downarrow$	RMSElog $\downarrow$	SILog $\downarrow$	SqRel $\downarrow$
BinsFormer [19]	0.632	0.792	0.845	0.265	16.211	0.139	0.466	38.009	6.387
SimIPU [18]	0.760	0.918	0.964	0.225	7.095	0.070	0.245	22.715	3.302
DepthFormer [20]	**0.860**	**0.958**	**0.981**	**0.136**	**5.831**	**0.050**	**0.190**	**18.101**	**1.614**

Table 2. **Monocular depth estimation performance.** The table compares BinsFormer, SimIPU, and DepthFormer across various traditional performance metrics. Notably, DepthFormer outperforms the other baselines across all metrics, showcasing seemingly great performance in accurately estimating depth. The arrows indicate desired outcome.

Model	Ultra Thin	Thin Structures	Small Mesh	Large Mesh	Trees	Buildings	Vehicles	Animals	Other	Background
BinsFormer [19]	**0.945**	**0.216**	0.129	0.209	0.248	0.137	0.141	0.150	0.141	0.257
SimIPU [18]	1.036	0.317	0.178	0.233	0.380	0.198	0.204	0.176	0.184	0.122
DepthFormer [20]	0.998	0.229	**0.115**	**0.177**	**0.206**	**0.121**	**0.120**	**0.121**	**0.128**	**0.082**

Table 3. **Class-wise absolute relative depth errors.** Each baseline's performance is evaluated per class, with lower values indicating better performance. DepthFormer achieves the lowest errors for the larger classes but completely fails to estimate depth for Ultra Thin. All methods severely struggle for the Ultra Thin class.

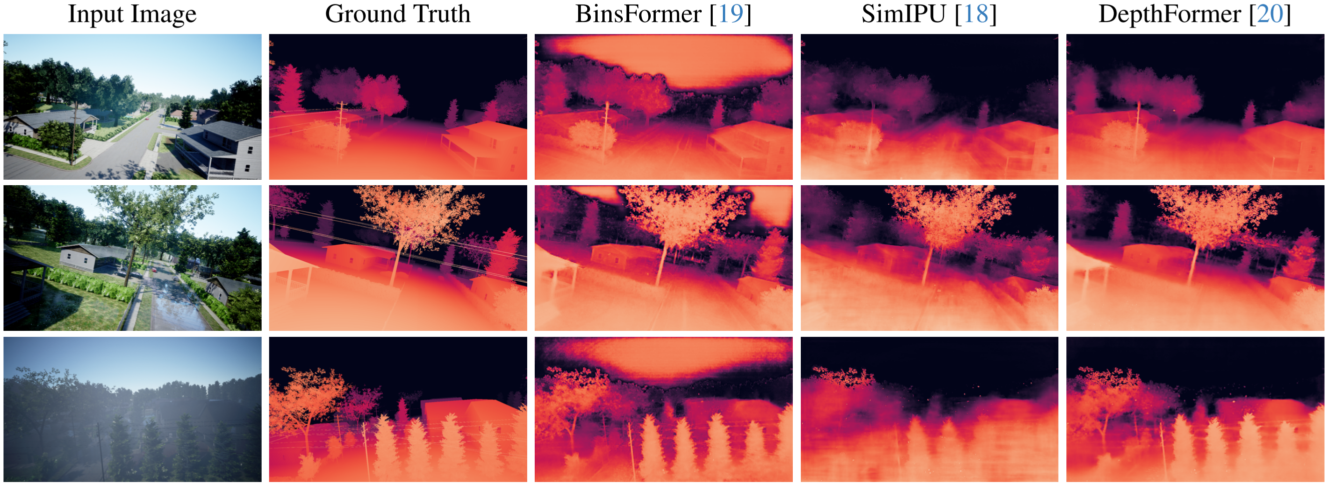


Figure 7. **Depth estimation performance of baselines.** This qualitative assessment underscores the challenges faced by state-of-the-art methods in accurately estimating depth, particularly for the *Ultra Thin* class. The results showcases the shared difficulty encountered by all methods in capturing the *Ultra Thin* class. This emphasizes the intricate nature of accurately discerning depth for such instances.

8. Conclusion

In summary, we introduce the DDOS dataset along with novel drone-specific depth metrics, marking a pivotal advancement in the field of autonomous drone navigation. The DDOS dataset addresses the critical challenges of detecting thin structures and operating under varied weather conditions, thereby filling an essential gap in the current scope of drone research. Through a detailed analysis of the dataset and the deployment of tailored evaluation metrics, we provide a nuanced methodology for systematically assessing the performance of depth estimation algorithms in drone-specific scenarios.

These efforts establish a new standard for future investigations aimed at enhancing the safety and efficiency of drone navigation through superior depth estimation and semantic segmentation techniques. The introduction of the DDOS dataset and corresponding metrics not only propels forward the development of drone technology but also extends the potential for computer vision applications within aerial environments. Our work lays a crucial groundwork for future innovations, steering the creation of algorithms that adeptly navigate the complexities of real-world settings, thus amplifying the functional prowess of drones across a multitude of industries.

References

[1] Rabab Abdelfattah, Xiaofeng Wang, and Song Wang. Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2, 3

[2] Stuart M Adams and Carol J Friedland. A survey of unmanned aerial vehicle (uav) usage for imagery collection in disaster research and management. In *9th international workshop on remote sensing for disaster response*, pages 1–8, 2011. 1

[3] Babankumar Bansod, Rangoli Singh, Ritula Thakur, and Gaurav Singhal. A comparision between satellite based and drone based remote sensing technology to achieve sustainable development: A review. *Journal of Agriculture and Environment for International Development (JAEID)*, 111(2): 383–407, 2017. 1

[4] Taha Benarbia and Kyandoghere Kyamakya. A literature review of drone-based package delivery logistics systems and their implementation feasibility. *Sustainability*, 14(1):360, 2021. 1

[5] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. Adabins: Depth estimation using adaptive bins. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4009–4018, 2021. 7

[6] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020. 2

[7] Joshua Candamo, Rangachar Kasturi, Dmitry Goldgof, and Sudeep Sarkar. Detection of Thin Lines using Low-Quality Video from Low-Altitude Aircraft in Urban Settings. *IEEE Transactions on Aerospace and Electronic Systems*, 45(3): 937–949, 2009. 2

[8] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[9] Sharifah Mastura Syed Mohd Daud, Mohd Yusmiaidil Putera Mohd Yusof, Chong Chin Heo, Lay See Khoo, Mansharan Kaur Chainchel Singh, Mohd Shah Mahmood, and Hapizah Nawawi. Applications of drone in disaster management: A scoping review. *Science & Justice*, 62(1):30–42, 2022. 1

[10] Milan Erdelj, Enrico Natalizio, Kaushik R. Chowdhury, and Ian F. Akyildiz. Help from the Sky: Leveraging UAVs for Disaster Management. *IEEE Pervasive Computing*, 16(1): 24–32, 2017.

[11] Mario Arturo Ruiz Estrada and Abrahim Ndoma. The uses of unmanned aerial vehicles–uav's-(or drones) in social logistic: Natural disasters response and humanitarian relief aid. *Procedia Computer Science*, 149:375–383, 2019. 1

[12] Michael Fonder and Marc Van Droogenbroeck. Mid-air: A multi-modal dataset for extremely low altitude drone flights. In *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2019. 2, 3

[13] Vipul Garg, Suman Niranjan, Victor Prybutok, Terrance Pohlen, and David Gligor. Drones in last-mile delivery: A systematic review on efficiency, accessibility, and sustainability. *Transportation Research Part D: Transport and Environment*, 123:103831, 2023. 1

[14] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. 11

[15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 2

[16] Yoshio Inoue. Satellite-and drone-based remote sensing of crops and soils for smart farming–a review. *Soil Science and Plant Nutrition*, 66(6):798–810, 2020. 1

[17] James R. Kellner, John Armston, Markus Birrer, K. C. Cushman, Laura Duncanson, Christoph Eck, Christoph Falleger, Benedikt Imbach, Kamil Král, Martin Krůček, Jan Trochta, Tomáš Vrška, and Carlo Zgraggen. New opportunities for forest remote sensing through ultra-high-density drone lidar. *Surveys in Geophysics*, 40:959–977, 2019. 1

[18] Zhenyu Li, Zehui Chen, Ang Li, Liangji Fang, Qinhong Jiang, Xianming Liu, Junjun Jiang, Bolei Zhou, and Hang Zhao. Simipu: Simple 2d image and 3d point cloud unsupervised pre-training for spatial-aware visual representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1500–1508, 2022. 7, 8

[19] Zhenyu Li, Xuyang Wang, Xianming Liu, and Junjun Jiang. Binsformer: Revisiting adaptive bins for monocular depth estimation. *arXiv preprint arXiv:2204.00987*, 2022. 7, 8

[20] Zhenyu Li, Zehui Chen, Xianming Liu, and Junjun Jiang. Depthformer: Exploiting long-range correlation and local information for accurate monocular depth estimation. *Machine Intelligence Research*, pages 1–18, 2023. 7, 8

[21] Ye Lyu, George Vosselman, Gui-Song Xia, Alper Yilmaz, and Michael Ying Yang. Uavid: A semantic segmentation dataset for uav imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 165:108–119, 2020. 2, 3

[22] Ratnesh Madaan, Daniel Maturana, and Sebastian Scherer. Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3487–3494. IEEE, 2017. 2, 3

[23] Alina Marcu, Vlad Licaret, Dragos Costea, and Marius Leordeanu. Semantics through time: Semi-supervised segmentation of aerial videos with iterative label propagation. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2, 3

[24] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3061–3070, 2015. 2

[25] Payal Mittal, Raman Singh, and Akashdeep Sharma. Deep learning-based object detection in low-altitude uav datasets: A survey. *Image and Vision computing*, 104:104046, 2020. 1

[26] Norzailawati Mohd Noor, Alias Abdullah, and Mazlan Hashim. Remote sensing uav/drones and its applications for urban areas: A review. In *IOP conference series: Earth and environmental science*, page 012003. IOP Publishing, 2018. 1

[27] Ishan Nigam, Chen Huang, and Deva Ramanan. Ensemble knowledge transfer for semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1499–1508. IEEE, 2018. 2, 3

[28] Yalong Pi, Nipun D Nath, and Amir H Behzadan. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics*, 43:101009, 2020. 1

[29] Chengyi Qu, Francesco Betti Sorbelli, Rounak Singh, Prasad Calyam, and Sajal K Das. Environmentally-aware and energy-efficient multi-drone coordination and networking for disaster response. *IEEE Transactions on Network and Service Management*, 2023. 1

[30] Giulia Rizzoli, Francesco Barbato, Matteo Caligiuri, and Pietro Zanuttigh. Syndrone-multi-modal uav dataset for urban scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2210–2220, 2023. 2, 3

[31] Aamina Shah, Komali Kantamaneni, Shirish Ravan, and Luiza C Campos. A systematic review investigating the use of earth observation for the assistance of water, sanitation and hygiene in disaster response and recovery. *Sustainability*, 15(4):3290, 2023. 1

[32] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics*, 2017. 4, 12

[33] Adam Stambler, Gary Sherwin, and Patrick Rowe. Detection and Reconstruction of Wires Using Cameras for Aircraft Safety Systems. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 697–703, 2019. ISSN: 1050-4729. 2

[34] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 2

[35] Lina Tang and Guofan Shao. Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, 26(4):791–797, 2015. 1

[36] Ashley Varghese, Jayavardhana Gubbi, Hrishikesh Sharma, and P Balamuralidhar. Power infrastructure monitoring and damage detection using drone captured images. In *2017 international joint conference on neural networks (IJCNN)*, pages 1681–1687. IEEE, 2017. 2, 3

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 7

[38] Wenshan Wang, Delong Zhu, Xiangwei Wang, Yaoyu Hu, Yuheng Qiu, Chen Wang, Yafei Hu, Ashish Kapoor, and Sebastian Scherer. Tartanair: A dataset to push the limits of visual slam. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4909–4916. IEEE, 2020. 2, 3

[39] World Economic Forum Global Future Council on Human Rights 2016–2018. How to prevent discriminatory outcomes in machine learning. https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning, 2018. 11