Air-to-Air Simulated Drone Dataset for AI-powered problems

Hamid Azad
School of Electrical Engineering and Computer Science
University of Ottawa, Ottawa, Canada
hamid.azad@uottawa.ca

Varun Mehta
Aerospace Research Centre
National Research Council Canada, Ottawa, Canada
varunkumar.mehta@cnrc-nrc.gc.ca

Fardad Dadboud
School of Electrical Engineering and Computer Science
University of Ottawa, Ottawa, Canada
fardad.dadboud@uottawa.ca

Miodrag Bolic
School of Electrical Engineering and Computer Science
University of Ottawa, Ottawa, Canada
miodrag.bolic@uottawa.ca

Iraj Mantegh
Aerospace Research Centre
National Research Council Canada, Montreal, Canada
iraj.mantegh@cnrc-nrc.gc.ca

2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC) | 979-8-3503-3357-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/DASC58513.2023.10311339

***Abstract*—This paper introduces the multi-view Air-to-Air Simulated Drone Dataset (A2A-SDD), a comprehensive simulated drone dataset captured using AirSim©. The dataset encompasses diverse scenarios where one or two drones are pursued by one to three monitoring drones. It includes five types of drones, such as DJI models and a generic quadrotor model, recorded in various weather conditions and environments. Both loaded and unloaded drones are represented, and the dataset provides extensive annotations, including object detection and XYZ coordinates. The dataset offers potential applications in training deep learning-based models for counter-UAV measures such as localization and payload detection in single- and multi-view cases. Furthermore, preliminary experiments demonstrate the promising performance of trained networks on practical data, affirming the dataset's value in addressing real-world drone challenges using optical sensors. The synthetic dataset is publicly available on GitHub (https://github.com/CARG-uOttawa/Multiview-Air-to-Air-simulated-drone-dataset).**

***Index Terms*—Drone, Uncrewed Aerial Vehicle (UAV), Counter-UAV measures, Simulated dataset**

I. Introduction

Uncrewed Aerial Vehicles (UAVs) have seen a significant rise in their applications across various sectors [1]–[7]. Their versatility and small size have made them popular for commercial and governmental purposes, including agriculture, transportation, surveillance, and photography. However, their widespread use raises concerns about safety and security, such as illegal activities and unauthorized flights near restricted areas and airports. As a result, there is a growing need for countermeasures that involve the detection, localization, classification, and neutralization of UAVs.

Based on the type of available sensors, counter-UAV measures can be broadly classified into two categories: ground-based and aerial. Ground-based solutions encompass a range of sensors and technologies, including radars and pan-tilt-zoom (PTZ) cameras [8], [9]. However, these ground-based platforms have limitations in terms of flexibility and coverage area for long distances, which may hinder their effectiveness in mitigating potential threats posed by unauthorized vehicles. To address these limitations, another solution is to deploy an observer drone in proximity to the target drone for countermeasures. By utilizing the mobility of the observer drone, the defense coverage area can be expanded, thereby enhancing the overall performance against the target drone. Nevertheless, it is not feasible to mount multiple sensors due to constraints such as payload size, weight, and limited power supply on the observer drone. In such cases, a vision camera emerges as a cost-effective solution which results in an air-to-air case. Vision-based sensing offers several advantages, including affordability, compact implementation, and lower power consumption.

The field of computer vision has witnessed significant advancements, particularly in deep learning (DL) algorithms, which have made optical sensing a highly appealing approach for drone detection and classification [10], [11]. Researchers have explored the use of cameras with different fields of view (FOV), combining wide field coverage with narrow FOV to achieve higher resolution and enhance identification performance. This has led to the development of various DL-based visual data analysis methods specifically tailored to address the challenges associated with drone detection and classification. However, DL methods typically require a substantial volume of data for effective training. Gathering a significant amount of training data and annotating it through real-world tests can be both expensive and time-consuming. Moreover, practical tests are often subject to uncontrollable factors such as weather conditions, which can introduce variability in the data. To mitigate these challenges, simulations present a valuable alternative. Simulations provide a controlled environment where various factors, including environmental conditions, can be precisely manipulated. This allows for the generation of large amounts of annotated data without the logistical constraints associated with real-world data collection. Due to the scarcity of publicly available data in the field of counter-UAV measures, this paper presents an air-to-air simulated drone dataset for DL-based solutions which includes various drone types, backgrounds, and weather conditions.

A. Literature review

By leveraging simulations, researchers can expedite the training process and explore a wider range of scenarios, enhancing the robustness and generalization capabilities of DL models for diverse real-world applications. For instance, to address challenges related to the high cost and privacy concerns of collecting real-world data, the authors in [12] proposed a substantial synthetic dataset tailored for Smart City applications. This dataset focuses on multiple vehicle tracking and segmentation across diverse camera views, providing a valuable resource for research in the field.

For the special case of air-to-air counter-drone problems, only two publicly available datasets exist [13], [14]. The authors in [13] presented an approach to detect and track small UAVs using a single camera mounted on another UAV. The proposed algorithm estimates background motion through a perspective transformation model and identifies salient points in the background subtracted image. It then applies optical flow matching to determine the spatio-temporal features of each moving object and classifies them based on their motion patterns compared to the background. The UAV-to-UAV detection and tracking dataset (U2U-D&TD dataset) utilized in this study comprises 50 video sequences with up to 8 UAVs captured in each frame. The videos were recorded outdoors using a GoPro camera mounted on a customized delta-wing airframe. Each video has a duration of approximately one minute, a frame rate of 30 fps, and a resolution of 1080×1920 or 960×1280. Despite the potential of this dataset for multi-target detection and tracking applications [15], there are certain limitations. Firstly, it lacks precise position information for both the observer and target drones, which is essential for addressing localization challenges. Additionally, with only one camera available, the dataset does not provide sufficient information to tackle multi-view problems effectively.

The authors in [16] proposed an approach for detecting small UAVs and aircraft filmed against complex backgrounds. By combining appearance and motion cues, the object-centric motion stabilization technique achieves effective classification of spatio-temporal image cubes. To evaluate the effectiveness of their approach, they constructed two datasets. The UAV dataset comprises 20 video sequences and 4000 frames capturing drones (up to two objects) in diverse lighting and weather conditions, both indoors and outdoors. On the other hand, the aircraft dataset consists of 20 publicly available videos featuring radio-controlled planes. These videos were filmed from different angles and included variations in plane poses. The videos vary in length and resolution, offering a diverse range of scenarios for evaluation. The recorded datasets lack the location information and the multi-view case is the same as [13].

B. Introduced dataset

This paper presents a multi-view air-to-air simulated drone dataset (A2A-SDD) using AirSim. In all cases, one or two coming drones are chased by one to three monitoring drones. The dataset is recorded by importing five types of drones into AirSim©, including various DJI models (Phantom, Mavic, FPV, and Inspire) and a generic quadrotor. The videos were recorded in several weather conditions, including sunny, rainy, and snowy weather. Drones were controlled to fly at different distances from the camera and occupied a range of pixels from 20 to higher values, corresponding to distances between more than 100 meters and less than 10 meters, respectively. Three different environments were selected to have both natural scenes, such as trees, rocks, and lakes, and some man-made objects, such as buildings and bridges. This feature helps to have various types of backgrounds in the collected dataset. All frames were annotated using the ”object detection” feature in AirSim©. Loaded drones were also modeled in AirSim©, and the dataset contains both loaded and unloaded drones. The XYZ information, as well as the quaternion values for all drones and frames, are also provided. The dataset was captured using multiple cameras, resulting in a multi-view recorded dataset. The last three mentioned features are the ones that are not available in the other datasets. The diverse dataset enables the training of deep learning methods for counter-drone problems such as classification, localization, and payload detection.

The paper is structured as follows: The next section details the procedure for creating the simulated dataset and provides specifications. Section III provides an overview of the problems addressed using deep networks. Section IV showcases samples from the recorded dataset and presents the results for solving these problems using the dataset. It includes the outcomes from both simulated and practical data. The paper concludes with a summary of the findings.

II. Simulated dataset using AirSim

A brief explanation of the creation of the Air-to-Air Simulated Drone Dataset (A2A-SDD) using AirSim© is given in this section. The dataset aims to address the lack of publicly available datasets for air-to-air counter-UAV problems. Various drone models, including DJI Phantom, Mavic, FPV, Inspire, and a generic quadrotor, were imported into AirSim to simulate different flying trajectories and collect the data.

The procedure for importing custom drone models into AirSim© involves several steps. Firstly, the 3D models of the drones are prepared by extracting the body and one propeller using 3D design tools such as Autodesk 3ds Max [17]. These components are then combined into a single object, which

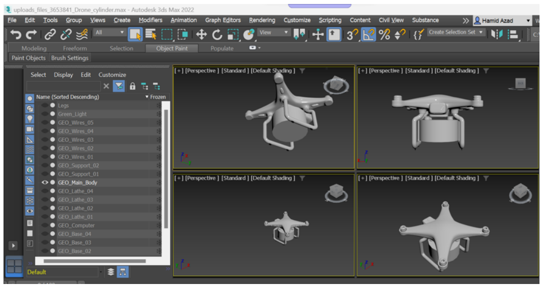


Fig. 1: Preparing the 3D model of loaded DJI Phantom with cylinder shape payload

is exported as an *fbx* file for use in AirSim© simulations. Afterwards, the imported drone models are constructed by replacing the corresponding parts of the default drone model in AirSim©. This is done by duplicating the default model and substituting the body and propeller with the imported objects. The blueprint editing window in AirSim© is used for this purpose. To enhance the diversity of the dataset, we simulate various forms of payloads attached to the drone models. The process involves separately creating the payload using 3D design tools and then merging it with the drone body to create a complete 3D model of the loaded drone (Fig. 1). In addition to importing custom drone models and simulating payloads, custom environments in AirSim© are utilized for the simulations. These environments include the Blocks, Landscape Mountains, and City Park environments. These environments provide a range of natural and man-made elements, such as trees, rocks, lakes, buildings, and bridges as the backgrounds for drone flights.

By following this procedure, we successfully import different drone models, simulate payloads, and generate diverse backgrounds, resulting in a comprehensive simulated air-to-air dataset for counter-UAV problems. As some examples of the developed dataset, Fig. 2 shows two models of simulated drones.

A. Specifications of the simulated dataset

This section offers specifications about the dataset, highlighting its characteristics. To provide an overview of the dataset, it is important to highlight the following details:

- The dataset includes five types of drones: DJI Phantom, Inspire, Mavic, FPV, and a generic drone model. We will refer to this model as the Quadrotor in this paper, assuming an approximate width of 70 cm, accounting for the widely spread propellers.
- In all scenarios, one or two coming drones are chased by one to three monitoring drones.
- The simulations are conducted in various environments such as Blocks, Landscape Mountains, and City Park.
- Different weather conditions are simulated, including sunny, rainy, and snowy cases.
- The dataset also incorporates various payload conditions, including unloaded drones, drones with attached loads, drones with hanging loads, and payloads with different colors and shapes.

(a) DJI Mavic drone model in the City Park environment

(b) DJI Inspire drone model in the snowy Blocks environment

Fig. 2: Sample frames of the developed dataset

For each case, around 1000 frames with a resolution of 1080×1920 pixels were recorded (for each camera) and annotated, and available through the provided link. The annotation files include the following information:

- The XYZ (i.e. Cartesian) coordinates of all the observer drones work as the platform for mounting the cameras
- The XYZ (i.e. Cartesian) coordinates of all the target drones in the captured image
- Bounding box information for each drone in every individual frame
- Timestamps to synchronize recorded videos by different cameras
- Flags indicating the detection status of a specific drone in each frame, as well as whether the drone is loaded or unloaded

The A2A-SDD addresses the limitations of existing datasets by providing a comprehensive and diverse collection of simulated air-to-air drone data. The dataset's characteristics, including diverse drone models, simulated environments, weather conditions, and payload variations, make it suitable for training and evaluating deep learning models using optical sensors. In addition, it introduced two distinct new features. Firstly, the simulation incorporated a multi-view case by utilizing multiple cameras, providing valuable additional information that can significantly benefit the development of localization and payload detection algorithms. This multi-view perspec-

(a) 1st camera (b) 2nd camera

Fig. 3: A sample of multi-view recorded scene with two cameras and two target drones in the Landscape Mountains environment (images are cropped for better illustration)

tive enhances the understanding of the scene and improves the accuracy of the algorithms. Secondly, detailed location information for all cameras and drones was included in the dataset, enabling the resolution of both single-view and multi-view localization problems. Importantly, these features are applicable and accessible in both single-target and multi-target scenarios, contributing to the dataset's versatility and utility in various applications. As an example, Fig. 3 illustrates a multi-view test case featuring two observer drones equipped with cameras and two target drones. In Fig. 3a, the image captured by camera 1 shows both the second observer drone and two target drones, with the second observer drone positioned between the targets. Conversely, Fig. 3b presents the scene captured by the second camera. It is evident that the second observer drone can only observe one of the target drones. Such cases allow for the exploration of counter-UAV challenges, such as localization, by leveraging sensor fusion between the cameras.

III. Considered problems

Based on the above-mentioned specifications, the developed A2A-SDD dataset is diverse enough to be used in training DL-based methods. In this paper, two of the potential problems have been addressed using deep networks such as a convolutional neural network (as a distance regressor), and some deep learning-based classifiers for distinguishing between loaded and unloaded drones.

A. Localization

Localization refers to the process of determining the precise position of a drone in a given environment. Accurate drone localization is essential for enabling autonomous navigation, precise control, and coordination of multiple drones in complex scenarios. Vision-based localization techniques leverage visual data, such as images or videos captured by onboard cameras or external sensors, to estimate the drone's position [18], [19]. Given that the developed dataset includes XYZ coordinates of both the observer and target drones, it provides ground truth location information, making it suitable for addressing localization challenges. As a result, localization is one of the selected problems considered in this study. The most challenging part of localization using 2D images is distance estimation. One fundamental challenge arises from the inherent loss of depth information in a two-dimensional representation. In this paper, we propose to leverage convolutional neural networks (CNNs) for regression tasks, estimating the distance of a drone from the captured image. The considered CNN network consists of 4 convolutional layers (with 8, 16, 32, and 32 filters, respectively) plus a fully-connected one as the last layer. Average pooling layers as well as batch normalization ones are also used in the architecture to improve training speed and dimension reduction. The CNN takes a cropped image within the drone's bounding box as input and outputs the estimated distance. As part of the preprocessing step, the cropped image is resized to a fixed size to ensure compatibility with the network. It is important to mention that the localization block is applied after the drone detection step. This means that an object detector, such as YOLO[1] [20], is used initially to detect the drone and extract its bounding box. However, in the case of simulation data, this information is obtained directly from AirSim©. Additionally, the drone's azimuth and elevation angles have been estimated based on its position in the captured image using a simple pinhole model for the camera.

B. Payload detection

Drone payload detection is an important research problem that concerns with classification of drones into two distinct categories: loaded and unloaded [21]–[23]. The ability to accurately identify whether a drone is carrying a payload or not is of great significance in various applications, including security, surveillance, law enforcement, and border control. In this study, the ResNet-50 and ResNet-101 architectures [24] are used as the classifier for addressing the loaded vs. unloaded drone classification problem. These networks belong to the ResNet family, and have proven to be highly effective in various computer vision tasks, including image classification. With 50 and 101 layers, respectively, these architectures can capture intricate features and learn discriminative representations, making them well-suited for complex classification problems.

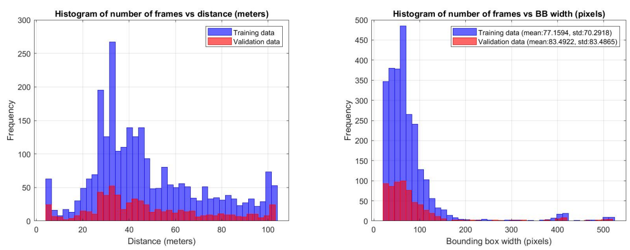
IV. Simulation results

The first part of the simulation section involves training the CNN network for 3D localization using two test cases in terms of the type of drones and background, and with different levels of complexity (explained in the upcoming lines). The dataset used for training includes multiple simulation conditions. To train the CNN network, the stochastic gradient descent optimizer is employed with an initial learning rate of 1×10^{-5}. The network is trained for 10 epochs to optimize its performance.

[1] You Only Look Once

(a) Training data (b) Testing data

Fig. 4: Some sample input images to the regression network for the first simulated test case



(a) Distance (b) Bounding box width

Fig. 5: Histogram of data for the first simulated test case for localization

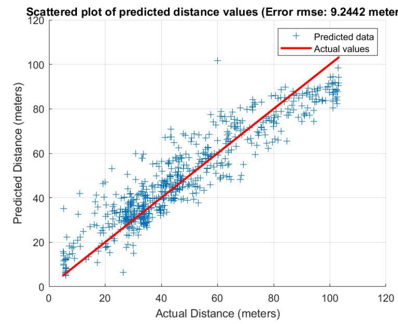
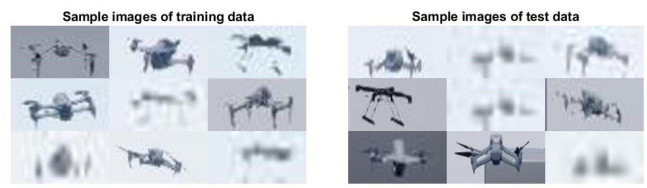


Fig. 6: Scatter plot of the estimated distance vs. the actual value for the first simulated test case



(a) Training data (b) Testing data

Fig. 7: Some sample input images to the regression network for the second simulated test case

In each case, 80% of the dataset is utilized for training, while the remaining 20% is reserved for testing/validation. In the first case, we consider the loaded and unloaded Quadrotor model in both Blocks and City Park environments. Examples of the captured images can be seen in Fig. 4.

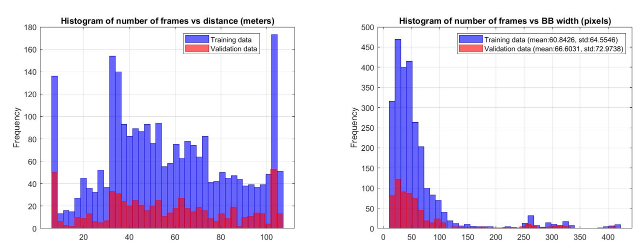
This data exhibits diverse backgrounds, and the distribution of backgrounds may differ between the training and test datasets. However, despite these challenges, the CNN network was able to estimate the distance, as shown in Fig. 6. The RMSE for distance estimation was calculated to be approximately 9.24 meters in this case. The histogram of the drone's distance and width is presented in Fig. 5 for both the training and testing datasets. The analysis reveals that the majority of drone distances fall within the range of 20 to 100 meters, with a particular concentration of around 30 to 40 meters. Consequently, the network achieved better results for distances around 40 meters due to the abundance of data for this particular case. The reduced deviation from the actual distance in Fig. 6 reflects this improved performance.

The inclusion of various types of drones with different sizes and shapes adds another layer of complexity to the case. In the second simulated case, we simulate four distinct drone models: DJI Mavic, DJI FPV, DJI Inspire, and Quadrotor. Each of these drone types possesses its own unique characteristics, such as differing dimensions and shapes. By incorporating this diverse range of drones into the Blocks environment, we create a more challenging distance regression problem. Samples of input images for this test case can be seen in Fig. 7.

The histogram in Fig. 8 depicts the distribution of distance values in the dataset. It shows that we still have a considerable number of observations with distances ranging from 10 meters up to approximately 100 meters. The RMSE is calculated to be around 7.8 meters for this complex case.

The aforementioned results highlight the efficacy of the CNN regression network in estimating the distance of the detected drone within the bounding box. The calculated RMSE indicates that this method can achieve accuracy levels comparable to GPS measurements. However, it is important to note that the simulated dataset includes various other backgrounds and conditions. As a path for future research, this allows a comprehensive evaluation of the method's performance and opens up the possibility for its effectiveness and robustness in more diverse scenarios.

For the next part, we addressed the binary loaded vs. unloaded drone problem using the deep network structure described earlier. The simulation parameters of the ResNet-50 and 101 models were set to 25 epochs, a batch size of 64,



(a) Distance (b) Bounding box width

Fig. 8: Histogram of data for the second simulated test case for localization

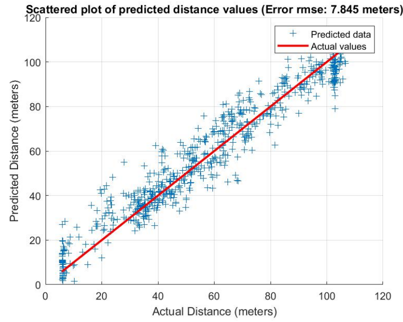


Fig. 9: Scatter plot of the estimated distance vs. the actual value for the second simulated test case

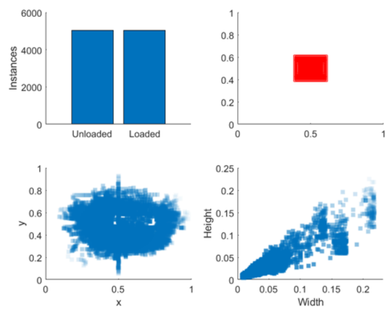


Fig. 10: Distribution analysis of the dataset for the test case used for the payload detection problem

and a learning rate of 0.03. The Stochastic Gradient Descent (SGD) optimizer has been selected for this problem. As an sample case, we consider the binary problem for a dataset that includes unloaded data from Quadrotor, DJI Mavic, and DJI FPV drones, and loaded data consisting of Quadrotor with an orange box, DJI Mavic, and DJI FPV with a gray box. The dataset covers various weather conditions such as sunny, rainy, and snowy. For each scenario, a total of 800 frames were captured, with 70% allocated for training, 15% for validation, and 15% for testing. This results in a total of 7,200 frames (5,040 for training, 1,080 for validation, and 1,080 for testing) for each unloaded and loaded case, amounting to 10,080 frames in total for training the loaded/unloaded classification task. Figure 10 presents an analysis of the collected dataset. The histogram in the top-left subfigure demonstrates a balanced distribution of loaded and unloaded classes, indicating an equal number of samples for each class. The bottom-left scatter plot visualizes the drone's location within the normalized frame region, showing that the drone is distributed across various parts of the frame. The top-right subfigure displays the normalized width and height of the ground-truth bounding boxes for different frames, providing insights into the size of the drones in the dataset. The bottom-right subplot illustrates the relationship between the bounding box size and the drone size, presenting a scatter plot that depicts the vertical and horizontal dimensions.

The histogram of the bounding box width for the training

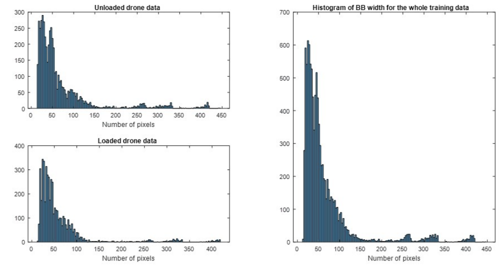


Fig. 11: Histogram of bounding box width for the training data in the test case utilized for the payload detection problem

data can also be seen in Fig. 11 which shows the maximum frequency of width to be around 20 to 30 pixels. The simulation results show an accuracy of around 97% and 99% for the ResNet-50 and 101, respectively.

A. Experimental results

Validation experiments were performed on the A2A-SSD dataset to evaluate its similarity to real-world data. This involved training networks using simulated data and subsequently testing their performance on real-world recorded data. The practical data was captured using a quadrotor at varying distances ranging from 20 to approximately 100 meters from the camera (ground-truth value was recorded using GPS sensor). In order to simulate a loaded drone, a box-shaped payload measuring $7.8 \times 9 \times 15cm^3$ was attached beneath the drone. For the payload detection problem, the performance was calculated based on around 2000 frames for the loaded and unloaded cases. Notably, the camera was mounted on a chasing drone, resulting in an air-to-air test case. For the distance estimation problem, the ground truth and estimated distances are plotted in Fig. 12 with an RMSE of around 8.4 meters. For the payload detection problem, the classification accuracy achieved is 88% and 92% for the unloaded drone and 87% and 89% for the loaded case for the ResNet-50 and 101, respectively. The results demonstrate the efficacy of our simulated dataset in training the classification and regression network. The recorded practical data will be made available soon.

V. Conclusion

This study presented a simulated multi-view air-to-air drone dataset, offering a comprehensive collection of various drone types, backgrounds, and weather conditions. The dataset stands out from existing datasets due to two distinct features: its multi-view perspective and the inclusion of location information for all cameras and drones. Deep networks were employed to address specific problems within the dataset, and the trained networks were tested on practical datasets. The preliminary results demonstrated the promising performance of the trained networks on previously unseen data, showcasing the dataset's wealth of information and its potential for solving real-world drone-related challenges using optical sensors.

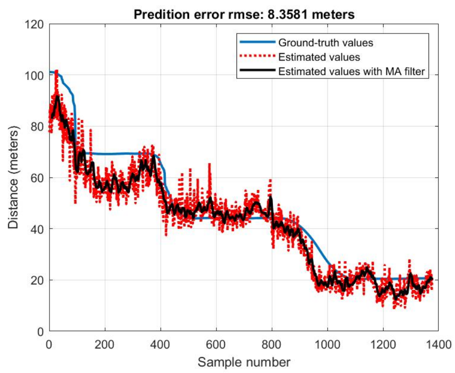


Fig. 12: Ground truth and estimated distance for a typical scenario

ACKNOWLEDGMENT

This work is supported by the Canadian Safety and Security Program, which is led by Defence Research and Development Canada's Centre for Security Science (DRDC CSS), in partnership with Public Safety Canada. The Canadian Safety and Security Program is a federally-funded program to strengthen Canada's ability to anticipate, prevent/mitigate, prepare for, respond to, and recover from natural disasters, serious accidents, crime, and terrorism through the convergence of science and technology with policy, operations, and intelligence. Partial funding is provided from the NSERC Discovery Grant RGPIN-2020-04417, and complementary support has also been provided by National Research Council Canada through *Artificial Intelligence for Logistics* and *Integrated Aerial Mobility* programs.

REFERENCES

[1] S. A. H. Mohsan, N. Q. H. Othman, Y. Li, M. H. Alsharif, and M. A. Khan, "Unmanned aerial vehicles (uavs): Practical aspects, applications, open challenges, security issues, and future trends," *Intelligent Service Robotics*, vol. 16, no. 1, pp. 109–137, 2023.

[2] J.-P. Huttner and M. Friedrich, "Current challenges in mission planning systems for UAVs: A systematic review," in *2023 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. IEEE, 2023, pp. 1–7.

[3] D. Fortune, H. Nitsch, G. Markarian, D. Osterman, and A. Staniforth, "Counter-unmanned aerial vehicle systems: Technical, training, and regulatory challenges," *Security Technologies and Social Implications*, pp. 122–148, 2022.

[4] S. A. H. Mohsan, M. A. Khan, F. Noor, I. Ullah, and M. H. Alsharif, "Towards the unmanned aerial vehicles (UAVs): A comprehensive review," *Drones*, vol. 6, no. 6, pp. 1–27, 2022.

[5] J. Farlík and L. Gacho, "Researching UAV threat–new challenges," in *2021 International Conference on Military Technologies (ICMT)*. IEEE, 2021, pp. 1–6.

[6] M. Hassanalian and A. Abdelkefi, "Classifications, applications, and design challenges of drones: A review," *Progress in Aerospace Sciences*, vol. 91, pp. 99–131, 2017.

[7] G. Singhal, B. Bansod, and L. Mathew, "Unmanned aerial vehicle classification, applications and challenges: A review," 2018. [Online]. Available: https://doi.org/10.20944/preprints201811.0601.v1

[8] F. Svanström, C. Englund, and F. Alonso-Fernandez, "Real-time drone detection and tracking with visible, thermal and acoustic sensors," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 7265–7272.

[9] A. Coluccia, G. Parisi, and A. Fascista, "Detection and classification of multirotor drones in radar sensor networks: A review," *Sensors*, vol. 20, no. 15, p. 4172, 2020.

[10] F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Single-stage uav detection and classification with yolov5: Mosaic data augmentation and panet," in *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2021, pp. 1–8.

[11] R. Jiang, Y. Zhou, and Y. Peng, "A review on intrusion drone target detection based on deep learning," in *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, vol. 4. IEEE, 2021, pp. 1032–1039.

[12] F. Herzog, J. Chen, T. Teepe, J. Gilg, S. Hörmann, and G. Rigoll, "Synthehicle: Multi-vehicle multi-camera tracking in virtual cities," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1–11.

[13] J. Li, D. H. Ye, M. Kolsch, J. P. Wachs, and C. A. Bouman, "Fast and robust uav to uav detection and tracking from video," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, pp. 1519–1531, 2021.

[14] M. W. Ashraf, W. Sultani, and M. Shah, "Dogfight: Detecting drones from drones videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7067–7076.

[15] J. Li, D. H. Ye, T. Chung, M. Kolsch, J. Wachs, and C. Bouman, "Multi-target detection and tracking from a single camera in unmanned aerial vehicles (UAVs)," in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016, pp. 4992–4997.

[16] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting flying objects using a single moving camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 879–892, 2016.

[17] Autodesk 3ds max. [Online]. Available: https://www.autodesk.ca/en/products/3ds-max/overview/

[18] H. Stuckey, A. Al-Radaideh, L. Escamilla, L. Sun, L. G. Carrillo, and W. Tang, "An optical spatial localization system for tracking unmanned aerial vehicles using a single dynamic vision sensor," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3093–3100.

[19] Y.-C. Lai and Z.-Y. Huang, "Detection of a moving UAV based on deep learning-based distance estimation," *Remote Sensing*, vol. 12, no. 18, p. 3035, 2020.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[21] U. Seidaliyeva, M. Alduraibi, L. Ilipbayeva, and N. Smailov, "Deep residual neural network-based classification of loaded and unloaded UAV images," in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, 2020, pp. 465–469.

[22] U. Seidaliyeva, M. Alduraibi, L. Ilipbayeva, and A. Almagambetov, "Detection of loaded and unloaded UAV using deep neural network," in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, 2020, pp. 490–494.

[23] I. Ku, S. Roh, G. Kim, C. Taylor, Y. Wang, and E. T. Matson, "UAV payload detection using deep learning and data augmentation," in *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, 2022, pp. 18–25.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.