ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Semi-synthetic Data for Automatic Drone Shadow Detection

Mohammed El Amine Mokhtari,[1] Virginie Vandenbulcke,[2] Sohaib Laraba,[3] Matei Mancas,[4] Elias Ennadifi,[5], Mohamed Lamine Tazir[6], and Bernard Gosselin[7]

University of Mons - Numediart
Boulevard Dolez, 31 - 7000 Mons - Belgium

Abstract. In this paper, we deal with the problem of shadow detection of UAVs, which impacts their navigation. We propose to generate synthetic images containing shadows in random locations, backgrounds, sizes, and opacities in order to augment our dataset. The generated data is used to train and compare several models to effectively detect, in real-time, UAVs shadows which will help to stabilize their localization and navigation. Deep learning models such as SSD, YOLOv3, and YOLOv5 are tested for the detection part. With our approach, we achieved 99% of the mean average precision when using the YOLOv5.

1 Introduction

Drones are now used almost everywhere; in very different domains such as data acquisition, cinema, rescue, and, most importantly, security. Our project entails the creation of a sophisticated indoor drone that will be used for security purposes without any internet connection. However, the drone itself has internal issues that must be addressed before progressing to the next level of sophistication. One of the main issues is that when a drone tries to stay steady and motionless in a given position (what we call an "absolute position"), it extracts keypoints from the ground to calculate the speed of the engines in order to stay still relatively to those keypoints. However, in some cases, when the drone's vertical camera takes images to extract keypoints, the captured frames contain the drone's shadow. As a result, a lot of the keypoints belong to the shadow which is not still but can be in motion. The drone will then stay still ... relatively to moving keypoints of the shadow which will imply a continuous and exponential motion relative to the ground. Here comes the idea of detecting shadows in frames to remove them and avoid potentially moving keypoints.

Problem: Classical image processing algorithms are used to detect or segment shadow in an image [3, 9]. But in our case, the drone's vertical camera is a monochrome camera with a resolution of 150x150 pixels, resulting in very poor image quality. So, if the current frame contains a drone shadow, which is a large black shape in the middle of a clearer background, the algorithm will automatically take keypoints within the pixels representing the drone shadow and use it as a corrupted reference. Fig. 1 shows some of the frames taken by the vertical camera.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Fig. 1: Examples of drone shadows (real data).

Limitations: The idea of detecting and then eliminating the shadow appears to be a good solution to the problem, however the amount of data available to train a deep learning model is insufficient. And this is a well-known issue, particularly when it comes to a very specific task for which no public dataset is available. In our case, the number of frames containing the drone shadow is too small to be used with a deep learning model.

Proposal: To address this issue, we propose to develop a deep learning model for object detection that is trained to detect the drone shadow from images in real time. When we know the position of the shadow in the image, we can simply exclude those pixels from the frame when attempting to extract keypoints.

2 Related Works

To date, all works done for shadow detection, whether for drone shadow or others, have used traditional image processing methods to remove the drone shadow from the image [9, 4]. S. Murali and VK. Govindan proposed converting the image into LAB space [3], which allows access to some important image parameters that are not available in other spaces. The L channel of the LAB space controls the luminance in the image, which helps in detecting darker areas thus defining shadowed pixels. Other authors [5, 6, 7, 8] attempted to incorporate convolutional networks (DNNs) for the extraction of relevant features. However, using classical image processing methods for this type of problems will be prohibitively computationally expensive. To avoid the speed issue, DNNs-based models are used for object detection such as SSD [1], YOLO [2], etc. can be used. However, in order to train these models, a large amount of data is required, which is a problem when dealing with shadows.

Synthetic data has previously been used for shadow detection, but in a different way. R. Guo et al. [10] used synthetic data generated by generative adversarial networks (GANs) for shadow detection and removal. For a real-time application, these algorithms are still too slow.

F. Nowruzi et al. [11] made an interesting comparison regarding the use of synthetic data for object detection. The purpose of their work was to determine the smallest amount of data that needed to be mixed with the synthetic data in

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

order to achieve the best object detection results possible. Table 4 displays the outcomes of various combinations of synthetic data such as 7D [13], P4B [14], and CARLA [12] with real data BDD [15]. The results in the table are for 95 % synthetic data and 5 % real data combinations. Their conclusion was that it is necessary to add more than 5 % of real data, but we show those results as on our side we used 5 % real data also.

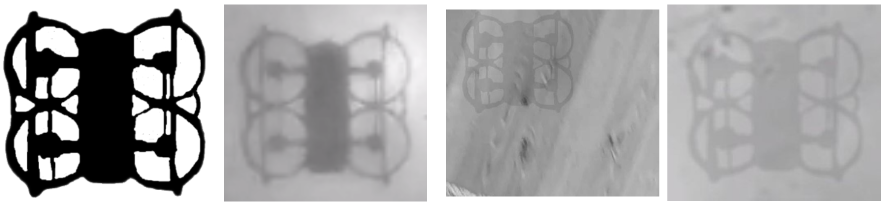
Training Dataset	Recall (person detection)	Recall (car detection)
7D + BDD	0.077	0.269
P4B + BDD	0.088	0.310
CARLA + BDD	0.034	0.287

Table 1: Metrics comparison for synthetic and real data [11].

3 Methodology

3.1 Synthetic augmented Data

We propose to generate synthetic data from pre-existing real frames. This solution entails extracting the drone shadow from the image and then placing it in various backgrounds with a variety of random parameters. The first parameter is the size of the shadow, which represents the distance between the drone and the ground. The second is the opacity, which varies at random from 20 to 100%. The third is a random position that controls the position of the drone in the image while adhering to all of the conditions.



(a) (b) (c) (d)

Fig. 2: (a) The drone shadow extracted from the image. (b) Real image with real shadow. (c) Background from a phone camera with the drone shadow (v1). (d) Background from the drone's vertical camera with the drone shadow (v2).

3.2 Shadow Detection Model

After creating the synthetic data, a deep learning model for shadow detection is built. The traditional image processing approaches can not be efficient for our problem due to the slowness of the existing algorithms. Detecting only the objects bounding boxes is easier than performing the precise object segmentation

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

(in our case of an indoor drone detecting its shadow the bounding bow is enough as we just need to eliminate the shadow are from frame where the shadow is present). For this purpose, we tested various object detection models such as SSD, YOLOv3 [2], and YOLOv5 [16].
The purpose of using these algorithms is that they have faster versions, and since our use is in the real time on a low-end device the computational time aspect is crucial.

We tried various dataset combinations to train the deep learning models. 5 % real images captured by the drone's vertical camera (~30 images) and 95 % synthetic images (~950 images). However, 95 % of the generated images have different combinations. The first combination is to create the synthetic data using the drone's shadow and backgrounds captured by a phone camera, which is the v1 of the synthetic data. The second combination is to create 500 images as the v1 and another 500 images using backgrounds taken by the drone's vertical camera as the v2.

4 Results and Discussion

We tested SSD, YOLOv3 and YOLOv5 for the task, but we do not present their results here because the YOLOv3 and SSD were unable to detect the shadow in the real dataset based on the training set. However, YOLOv5 (small version) was trained and is shown in Table 2 for a variety of dataset combinations. The results are all tested on the real data test set.

Dataset	Recall	mAP @.5	mAP @.5 : .9
100% RD	0.75	0.875	0.665
100% SD	0	0	0
95% SD + 5% RD (v 1)	1	0.982	0.869
95% SD + 5% RD (v 2)	**1**	**0.995**	**0.923**

Table 2: Metrics of multiple datasets (RD: real data and SD: synthetic data combinations. Results tested on the real images test set.

The results for the 100 % real datasets are presented in the first row of Table 2. It has been trained on 30 real images. The results are good, but also overfeeted for the specific dataset. In ddition, the results for an industrial project are insufficient. The results of the model trained on 100 % synthetic dataset are shown in the second row. The model was tested on a simulated dataset and yielded a mAP of 0.914, but when tested on real data, the mAP is 0% which means 0 detections. It was unable to detect shadows in real-world images. The model was trained on a mixed dataset in the third row, with 5 % real (30 images) and 95 % synthetic datasets. The synthetic datasets (v1) were created using a phone camera's background and a random placement of the shadow as described

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

in 3.1. The results appear to be far superior to training on real-world datasets (since we got hundreds of the training images). Finally, the last row shows the same architecture trained on the dataset v2. In this case, there is an additional improvement of the results.

(a) (b) (c) (d)

Fig. 3: Qualitative result. (a) 100% real data. (b) 100 % synthetic data. (c) Mixed data (v1). (d) Mixed data (v2).

Fig. 3 depicts a qualitative result of the model. The model trained on both real and synthetic data performs better than previous models, but still makes some errors when some objects are present in the scene. As a result, the next model trained on real backgrounds (taken by the drone camera) significantly improved the results. A large amount of synthetic data, 950 randomly generated images, was used. The best combination when creating synthetic data is to use 50 % real backgrounds (preferably with objects), because the model will be trained on data with the same image quality when using backgrounds from the exact camera that will be used in the real world.
It is also interesting to compare Table 2 with which are computed with the same mix of synthetic/real data. The is an important difference in the results which is probably due to several factors : 1) the complexity of the object detection : in our case we focus on a single drone so the shadow has very low variability while in several object with much higher variability are taken into account, 2) The synthetic data is based on real images background in our case while in in the data is fully synthetic. Our work show that for a precise use-case with low variability, the YOLO v5 version can be efficiently trained with very few real-data images if semi-synthetic images can be added to the training dataset.

5 Conclusion

The possibility of using synthetic data for shadow detection is discussed in this paper. The lack of training data can be resolved by adding synthetic or semi-synthetic data to the training set. However, having only synthetic data is insufficient as shown in previous work, but mixing the synthetic data with a small amount of real data is enough to have a good model which can detect objects with a limited variability.
In our case, the best model was obtained by randomly placing the drone's shadow on different real backgrounds.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

6 Acknowledgments

The authors acknowledge the Region Wallonne (Belgium) for the research funding for DroneSecurity project and TRAIL/ARIAC project and the company Sky-Hero for their collaboration.

References

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, SSD: Single Shot MultiBox Detector. *Computer Vision and Pattern Recognition* (CVPR 2016),

[2] J. Redmon, and A. Farhadi, YOLOv3: An Incremental Improvement. *Computer Vision and Pattern Recognition* (CVPR 2018),

[3] S. Murali, and VK. Govindan, Shadow Detection and Removal from a Single Image Using LAB Color Space. *Cybernetics and Information Technologies* (2013),

[4] V. Arevalo, J. Gonzalez, and G. Ambrosio, Shadow detection in colour highâresolution satellite images. *Cybernetics and Information Technologies* (2013),

[5] X. Zou, R. Zhou, M. El Helou, and S. Susstrunk, DRONE SHADOW TRACKING. *CoRR* (2019),

[6] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, Automatic Feature Learning for Robust Shadow Detection. *CVPR* (2014),

[7] S. Hosseinzadeh, M. Shakeri, and H. Zhang, Fast Shadow Detection from a Single Image Using a Patched Convolutional Neural Network. *International Conference on Intelligent Robots and Systems (IROS)* (2018),

[8] P. Corke, R. Paul, W. Churchill, and P. Newman, Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. *International Conference on Intelligent Robots and Systems (IROS)* (2018),

[9] G. D. Finlayson, M. S. Drew, and C. Lu, Entropy minimization for shadow removal. *International Journal on Computer Vision (IJCV)* (2009),

[10] R. Guo, B. Ayinde, and H. Sun, Efficient Shadow Detection and Removal using Synthetic Data with Domain Adaptation. *International Conference on Pattern Recognition (ICPR)* (2021),

[11] Nowruzi F. E., Kapoor P., Kolhatkar D., Hassanat F. A., Laganiere R., and Rebut J., How much real data do we actually need: Analyzing object detection performance using synthetic and real data. *arXiv preprint arXiv:1907.07061.* (2019),

[12] Dosovitskiy A., Ros G., Codevilla F., Lopez A., and Koltun V., CARLA: An open urban driving simulator. *In Conference on robot learning (pp. 1-16).* (2017),

[13] Wrenninge M., and Unger J., Synscapes: A photorealistic synthetic dataset for street scene parsing. *arXiv preprint arXiv:1810.08705.* (2018),

[14] Richter S. R., Hayder Z., and Koltun V., Playing for benchmarks. *In Proceedings of the IEEE International Conference on Computer Vision (pp. 2213-2222).* (2017),

[15] Yu F., Chen H., Wang X., Xian W., Chen Y., Liu F., ... and Darrell, T., Bdd100k: A diverse driving dataset for heterogeneous multitask learning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2636-2645).* (2020),

[16] Jocher G., Chaurasia A., Stoken A., Borovec J., Yonghye K., ... and Thanh Minh M.,Playing for benchmarks. *ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference, Zenodo, 10.5281/zenodo.6222936, https://doi.org/10.5281/zenodo.6222936* (2022),