Journal of Engineering and Applied Sciences 14 (Special Issue 1): 3946-3949, 2019
ISSN: 1816-949X

Human Detection using Improved YOLOv2: Images Captured by the UAV

[1]Juwon Kwon and [2]Soonchul Kwon
[1]Department of Electronic Engineering,
[2]Department of Smart Systems, Kwangwoon University, Seoul, Korea

Abstract: In recent years, the technology of Unmanned Aerial Vehicles (UAV) has developed. The UAV which were initially used for military purposes have recently begun to be commercialized. The market for the UAV has grown and small UAV have been created. Many people can enjoy a variety of hobbies such as taking pictures using the UAV. Therefore, the technology utilizing the images captured by the UAV is also developing. Existing human detection algorithms have resulted in unsatisfactory results in images taken by the UAV. There is a problem that distortions that do not occur in ordinary images occur due to the difference of angle of view. When the distance between the UAV and the human is long, there is a problem that the human is small and has a low resolution. In this study, we improve the YOLOv2 algorithm which shows good performance for object detection, to improve human detection performance in images taken by the UAV.

Key words: UAV, human detection, deep learning, algorithm, unsatisfactory, commercialized

INTRODUCTION

In recent years as the market for the UAV has increased, more and more people have been controlling the UAV as a hobby. They take pictures and videos through the UAV. However, the image obtained from the UAV is different from the image obtained by the general camera. Distortion occurs due to differences in angle of view because the image is taken above a human. Figure 1 shows the difference of images according to this differences in angle of view. The angle depends on the altitude of the UAV. And there are problems as the altitude of the UAV increases. When the altitude of the UAV is increased, the size of objects to be detected becomes smaller in the image. When the ratio of objects in an image is small it has difficulty in detecting because it has a low resolution.

In order to solve this problem, we apply the deep learning algorithm which is undergoing many studies recently. Deep learning which requires a lot of computations has made considerable progress by accelerating computation speed with the development of hardware. Deep learning has been studied in various fields and has had a great influence on the image processing field. Image processing using artificial neural network has been actively studied. As a result, the accuracy of object classification has surpassed human in 2015. Figure 2 shows the ImageNet top-5 error (%) of

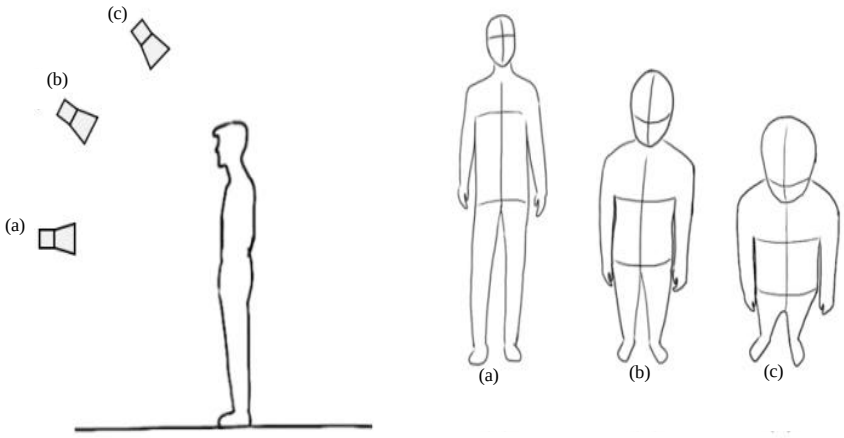

Fig. 1: a-c) Distortion of human shape by camera angle

Corresponding Author: Soonchul Kwon, Department of Smart Systems, Kwangwoon University, Seoul, Korea

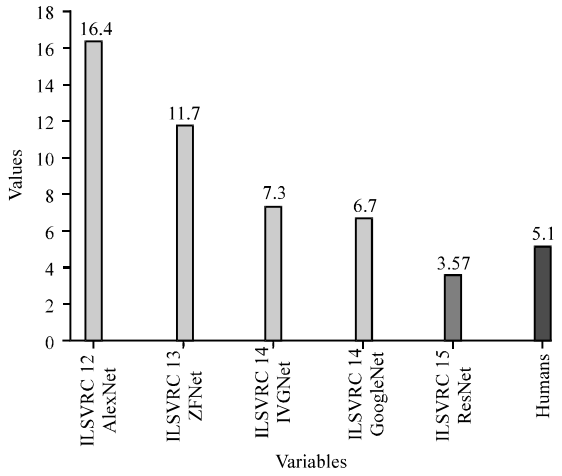

Fig. 2: ImageNet top-5 error of object classification

Fig. 3: a) The original image and b) is the result image of human detection with YOLOv2

Table 1: Detection frameworks on PASCAL VOC 2007

Detection frameworks	mAP	FPS
Fast R-CNN	70.0	0.5
Faster R-CNN	73.2	7
SSD300	74.3	46
SSD500	76.8	19
YOLOv2 288×288	69.0	91
YOLOv2 416×416	76.8	67
YOLOv2 544×544	78.6	40

classification task. ResNet by He *et al*. (2016), released in 2015, demonstrates performance that exceeds human. We describe human detection using YOLOv2 Model among state-of-the-art algorithms based on deep learning. The YOLO algorithm shows good performance when the object size is large in an image and the image resolution is high. Table 1 shows the performance of the existing YOLOv2 algorithm. FPS and precision are good. However, in Fig. 3 it can be seen that the detection fails in the image taken by the UAV. In this study, we improved the YOLOv2 Model to show better results in aerial images.

MATERIALS AND METHODS

Design and simulation

Existing object detection algorithms: State-of-the-art deep learning based object detection algorithms include SSD (Liu *et al*., 2016), YOLO (Redmon *et al*., 2016; Redmo and Farhadi, 2017), Fast-RCNN (Girshick, 2015) and Faster-RCNN (Ren *et al*., 2015). We used YOLOv2 which has FPS of about 40 and mAP of 78.6 for 544×544 input image among various algorithms. This is the measurement result for PASCAL VOC 2007 data and it shows different results depending on hardware specification. YOLOv2 calculates the reliability of each grid by dividing the image into S×S grids. At first it shows inaccurate reliability but it adjusts the position of the bounding box and increases the accuracy by leaving the bounding box with the highest reliability. This provides about twice the performance of existing real-time object detection algorithms.

Image distortion problem: In the case of an image taken by the UAV other than a general shooting method, a difference in angle of view occurs and the object is distorted. Figure 1 shows why as the camera's height increases, the proportion of objects will be different from normal images. And this is a challenging problem in the existing deep learning based object detection algorithm. As the altitude of the UAV increases, the ratio of the human head increases relatively and differs from the normal image. So, it does not perform well in deep learning based algorithms. Figure 3 shows that when the input of the existing YOLOv2 Model is not a normal human image but an image input called an aerial view image or a bird's eye view image, the object is misidentified or misdetected.

Proposed method: In the case of the YOLOv2 Model which has been trained from existing data, there is a limit

to the lack of a human dataset captured by the UAV. To solve this problem, we created training data using publicly available dataset. In the training data, the dataset of the human images from a normal camera and the UAV is included and the image of the human corresponding to the top view is excluded. The reason for this is that the head ratio has become too large to lower the performance of the model. The proposed method which is an improvement of YOLOv2, follows the structure of Darknet-19 but since, there are many unnecessary classes in human detection. So, I reduced the number of classes and adjusts the number of layers and features of the network accordingly. At first, we only want to detect one object, so, we left only one class. However, there was sometimes a problem of detecting shadows as people. So, we proceeded to add classes for shadow to reduce false positives. In conclusion, we modified the class value to 2 and the filter value to 35 through the calculation of 5x (classes+5).

RESULTS AND DISCUSSION

In this sudy, the experiment was conducted using the collected dataset in the improved YOLOv2 Model. The dataset was created by capturing the images in the building and the video of the mini-drone video dataset (Bonetto *et al*., 2015). The experiment was conducted on Ubuntu 16.04 with GeForce GTX 1080 graphics card and Intel® Core™ i7 CPU.

Data: We created a dataset including 4000 images taken as the bird's eye view in the building and the mini-drone video datasets. The mini-drone video dataset is 38 videos taken with the DJI's Phantom 2 drone. Each video has a full HD resolution of 16-24 sec long. Figure 4 and 5 show an example of our dataset and the mini-drone video dataset.

Experimental results: The proposed method shows better performance than the existing YOLOv2 algorithm. In the existing YOLOv2 Model when the aerial view image is inputted, the object is not detected correctly or is detected as another object. Table 2 shows that our method has better recall and precision than previous methods:

$$\text{Recall} = \frac{\text{Detected true}}{\text{Total number of existing true}} \quad (1)$$

$$\text{Precision} = \frac{\text{True detections}}{\text{Whole detections of an algorithm}} \quad (2)$$

The recall and precision were calculated according to Eq. 1 and 2. Figure 6 shows the original images and the results of human detection using the proposed method. Figure 3 detects well in the example image that the existing YOLOv2 did not detect.

Fig. 4: a-c) Example images of a dataset we shot in a building

Fig. 5: Example images of mini-drone video dataset

Fig. 6: The original images and the results of human detection using the proposed method: a) The original images; b) Result of human detection using proposed method and c) A comparison of the results between the existing YOLOv2 and the proposed method

Table 2: The results of experiment

Detection frameworks	Recall	Precision
YOLOv2	0.431	0.674
Proposed method	0.914	0.878

CONCLUSION

In this study, we experimented to detect human by adjusting the Darknet-19 structure of object detection algorithm YOLOv2 for images taken by the UAV. In the existing YOLOv2 which has low performance when the object is distorted and the size is small or the objects are close to each other, a new training dataset is added, the number of classes is reduced and the model is modified. As a result, the human detection performance is better in the images taken by the UAV.

REFERENCES

Bonetto, M., P. Korshunov, G. Ramponi and T. Ebrahimi, 2015. Privacy in mini-drone based video surveillance. Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) Vol. 4, May 4-8, 2015, IEEE, Ljubljana, Slovenia, ISBN:978-1-4799-6026-2, pp: 1-6.

Girshick, R., 2015. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV'15), December 7-13, 2015, IEEE, Computer Society Washington, DC, USA., ISBN:978-1-4673-8391-2, pp: 1440-1448.

He, K., X. Zhang, S. Ren and J. Sun, 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 26-July 1, 2016, IEEE, Las Vegas, Nevada, USA., ISBN:9781509014385, pp: 770-778.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy and S. Reed *et al*., 2016. SSD: Single shot multibox detector. Proceedings of the European Conference on Computer Vision, October 11-14, 2016, Springer, Cham, Switzerland, ISBN:978-3-319-46447-3, pp: 21-37.

Redmon, J. and A. Farhadi, 2017. YOLO9000: Better, faster, stronger. J. Comput. Sci., 1: 1-9.

Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, IEEE, Las Vegas, Nevada, USA., ISBN:978-1-4673-8852-8, pp: 779-788.

Ren, S., K. He, R. Girshick and J. Sun, 2015. Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. In: Advances in Neural Information Processing Systems, Cortes, C., N.D. Lawrence, D.D. Lee. M. Sugiyama and R. Garnett (Eds.). Curran Associates, Inc., New York, USA., pp: 91-99.