A TOOL TO ENHANCE THE CAPACITY FOR DEEP LEARNING BASED OBJECT DETECTION AND TRACKING WITH UAV DATA

A. Ancy Micheal [1,*], K. Vani [2], S. Sanjeevi [3], Chao-Hung Lin [4]

[1] Dept. of Information Science and Technology, College of Engineering, Anna University, Chennai, India – ncysus17@gmail.com
[2] Dept. of Information Science and Technology, College of Engineering, Anna University, Chennai, India – vanirrk@gmail.com
[3] Dept. of Geology, College of Engineering, Anna University, Chennai, India – ggjeevi@gmail.com
[4] Dept. of Geomatics, National Cheng-Kung University, No. 1, University Road, Tainan City 701, Taiwan – linhung@mail.ncku.edu.tw

KEY WORDS: UAV, Deep Learning, Object Detection and Tracking

ABSTRACT:

Currently, deployment of UAV has transformed from crucial to day-to-day scenarios for various purposes such as wastage collection, live entertainment, product delivery, town mapping, etc. Object tracking based UAV applications such as traffic monitoring, wildlife monitoring and surveillance have undergone phenomenal changeover due to deep learning based methodologies. With such transformation, there is also lack of resources to practically explore the UAV images and videos with deep learning methodologies. Hence, a deep learning-based object detection and tracking tool with UAV data (DL-ODT-UAV) is proposed to fill the learning gap, especially among students. DL-ODT-UAV is a resource to acquire basic knowledge about UAV and deep learning based object detection and tracking. It integrates various object annotators, object detectors and object tracker. Single object detection and tracking is performed with YOLO as object detector and LSTM as object tracker. Faster R-CNN is adopted in multiple object detection. With exploring the tool, the ability of students to approach problems related to deep learning methodologies will improve to a greater level.

1. INTRODUCTION

Unmanned Aerial Vehicle(UAV) deployment has created tremendous growth in fields such as disaster recovery (Erdelj and Natalizio, 2016), traffic surveillance (Khan et al., 2017), ecology monitoring (Madhavan et al., 2018), forest surveillance (Berie and Burud, 2018), land mapping, road mapping, town mapping ("Mapping India through drones," 2019), research in earth science, wildlife and maritime monitoring (Hodgson et al., 2018), product delivery (Haque et al., 2014), military purposes (Cyprian Aleksander, 2018)and police investigation (Ndna and Tss, 2017). The object detection and tracking methodologies along with UAV data have improved surveillance and security purpose. The transformation of feature engineering-based object detection and tracking to deep learning-based object detection and tracking enhance the accuracy of tracking based UAV applications. To perform detection in UAV images, features invariant to scale, affine transformation, rotation and translation are suitable. Traditional object detection methodologies such as SURF, SIFT (Micheal and Vani, 2018) , Harris corner operator (Yu et al., 2008) and Enhanced Viola-Jones (Xu et al., 2017) have been experimented by the researchers for object detection in UAV images. Deep learning (DL) based object detectors such as Faster R-CNN (Ren et al., 2017), You Only Look Once(YOLO) (Xu et al., 2018) and Single Shot Multibox Detector (Rohan et al., 2019)performs well with UAV data. Traditional object tracking algorithms such as Meanshift (Fang et al., 2011), Kanade Lucas Tomasi (Tong et al., 2013) and Kalman Filter (Teutsch and Krüger, 2012) have been implemented in UAV object tracking. DL based object trackers to adapt UAV videos are still evolving.

The capability of DL methodologies to learn invariant features automatically, eases the process of object detection and tracking with better accuracy. Hence, DL based object detection and tracking methodologies gain importance to adapt UAV videos over traditional methodologies. In the near future, there is a high scope for DL methodologies to provide solutions for various problems related to object detection and tracking with UAV videos. The technical features of existing UAV have improved to a vast extent to improve their service to society. Tremendous efforts are taken to reach UAV to the layman. UAV has reached farmers in a remote village of India to spray pesticides in agricultural farms("Farmers use drones to spray pesticide," 2019). With such efforts to reach UAV to a layman, lack of materials to handle UAV data among students and researchers still persist. In this work, a tool is proposed to enrich the experience of the students/researchers for DL-based object detection and tracking with UAV data (DL- ODT-UAV). The objectives of DL-ODT-UAV tool are :

1. To educate the students with the basics of UAV, object detection, object tracking and DL based object detection and tracking.

2. To provide a practical exploration of DL based object detection and tracking with UAV data.

2. COMPONENTS OF DL-ODT-UAV

The DL-ODT-UAV is comprised of :

1. Study resource about UAV and deep learning based object detection and tracking
2. DL based Single Object Detection and Tracking
3. DL based Multiple Object Detection

2.1 Study resource in DL-ODT-UAV

The study resource is designed to provide basic knowledge about :

1. UAV
2. Object detection and tracking
3. Deep learning
4. DL based object detectors and trackers

The materials showcase the transformation of UAV from pigeon based aerial imagery to UAV's available in 2019. The technical details of UAV at the year 2019 such as flight time, area coverage, camera resolution, maximum speed and its primary applications are mentioned in the study resource. Brief description of object detection and tracking, the need for transformation from feature engineering to deep learning and types of deep learning based object detection and tracking would provide basic knowledge about vision based deep learning concepts to the students/researchers.

2.2 Single Object Detection and Tracking

In this module, Recurrent YOLO (ROLO) model is adopted for DL based single object detection and tracking with UAV data (Ning et al., 2017). ROLO model exploits the spatiotemporal domain for accurate tracking. The framework of ROLO model is shown in Figure 1. The video frames are annotated and fed into YOLO object detector. YOLO collects the visual features and the spatial location of the objects. The object location consists of class, confidence, bounding box center, height and width of the image. In the next stage, the obtained spatial location and the visual features are fed into Long Short Term Memory(LSTM) for sequence processing. LSTM exploits the visual feature and spatial location for predicting the object location. Object tracking is performed with trajectory.

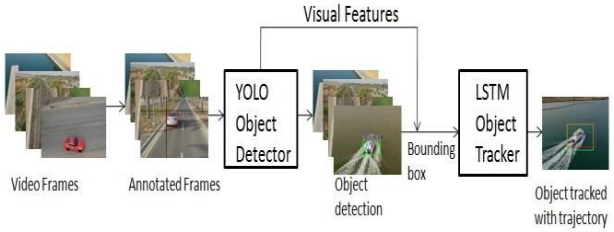

Figure 1. Architecture of single object detection and tracking

The tracking probability of ROLO model is given in Equation 1. In training, Mean Squared Error is adopted (Equation 2).

$$p(B_1, B_2, , B_T \mid X_1, X_2, , X_T) = \prod_{t=1}^{T} p(B_t \mid B_{<t}, X_{\leq t}) \quad (1)$$

where B_t = object location
X_t = input frame
$B_{<t}$ = previous object location
$X_{<t}$ = previous input frames

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n} \left\| B_{t\,\arg et} - B_{prediction} \right\|_2^2 \quad (2)$$

where $B_{prediction}$ = model prediction
B_{target} = target groundtruth value
$\|.\|$ = squared euclidean norm

2.3 Multiple Object Detection

Faster R-CNN is adopted for multiple object detection("TensorFlow-Object-Detection-API," 2019). The framework of multiple object detection in DL-ODT-UAV is shown in Figure 2. The objects in the image frames are annotated. In this module, pretrained Inception-V2 model is used ("Inception v2 model," 2017). The annotations along with pretrained model are trained for object detection using Faster R-CNN. Faster R-CNN is composed of region proposal network and a detector network. Region proposal network generates region proposals followed by Fast R-CNN to detect objects. Object detection will generate multiple bounding boxes around the same object with various scores. Hence, there is a need to remove the bounding boxes with low scores. Non-maximum suppression is applied to remove multiple bounding boxes around the same object. Multiple bounding boxes with low scores existed for threshold below 0.91%. With threshold as 0.91%, single bounding box around the respective objects are obtained. Hence, threshold is fixed as 0.91% thereby retaining the bounding boxes with scores higher than the threshold.

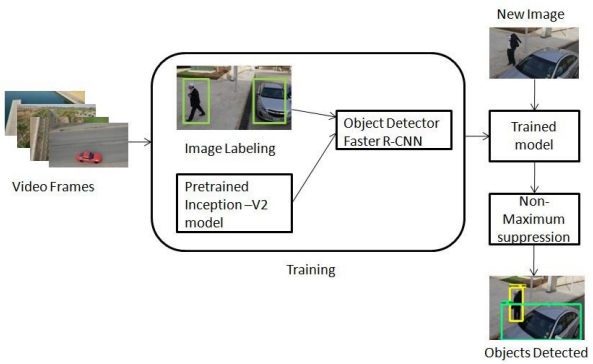

Figure 2. Architecture of multiple object detection

3. IMPLEMENTATION OF DL-ODT-UAV TOOL

The DL-ODT-UAV tool is implemented in Python with libraries such as Tkinter (Grayson, 2000), Tensorflow and Keras (Ballard, 2018). With the components mentioned in Section 2, an interactive graphical user interface has been developed to carry out activities for deep learning-based object detection and tracking.

3.1 Interactive Module 1: Study Resource

The study resource is provided as a scrollable pdf format. The front-end of study resource sample is shown in Figure 3(a) and (b).

(a)

(b)

Figure 3(a) and (b). Sample of study resource in DL-ODT-UAV tool

The manual about the DL-ODT-UAV tool is provided to guide the students/researchers. Following the manual, the type selection is provided to the user on the next page as shown in (Figure 4). The type selection is provided for:

1. Single Object Detection and Tracking
2. Multiple Object Detection

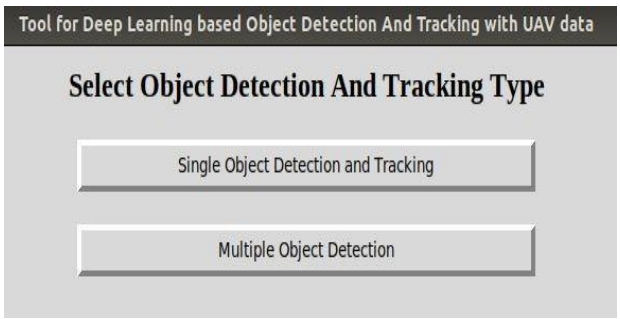

Figure 4. Button options of Type selection

3.2 Interactive Module 2: Single Object Detection and Tracking

Single object detection and tracking module is composed of the following steps:

1. Video to frame conversion.
2. The object in the frame is labeled semi-automatically.
3. The ground truth file for the labeled images is generated.
4. With the generated ground truth, object detection is performed with YOLO object detector.
5. The visual features and the spatial location obtained from the detected objects are fed into LSTM for training.
6. Object tracking is performed with trajectory.

The input is obtained with three options (Figure 5):

1. Input video followed by frame conversion
2. Image folder
3. Select the existing dataset

Figure 5. Loading data for single object tracking.

BBox-Label-Tool is used for object annotation ("BBox-Label-Tool," 2017). The tool is modified for annotating single object per image frame. Manually labeling the object in each frame for 10 videos was a tedious job. Hence, a semi-automatic single object annotator with YOLO has been implemented in the tool to reduce the labeling time. YOLO object detector performs object detection. The user checks the obtained bounding box in every frame. In case of false detection, the user rectifies by drawing the bounding box on the object. The groundtruth for the annotated images is generated as a text file. The groundtruth file contains the bounding box locations of the annotated objects. The modified semi-automatic single object annotator is shown in Figure 6. The generated ground truth is fed into YOLO for object detection(Figure 7). The visual features and the spatial location obtained from YOLO detection are fed into LSTM for object tracking. The number of training iteration is obtained from user (Figure 8). Finally tracking demo is performed for object tracking with trajectory as shown in Figure 9.

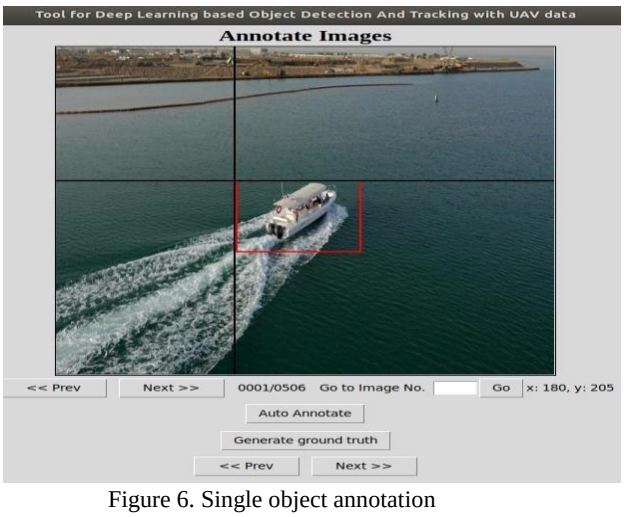

Figure 6. Single object annotation

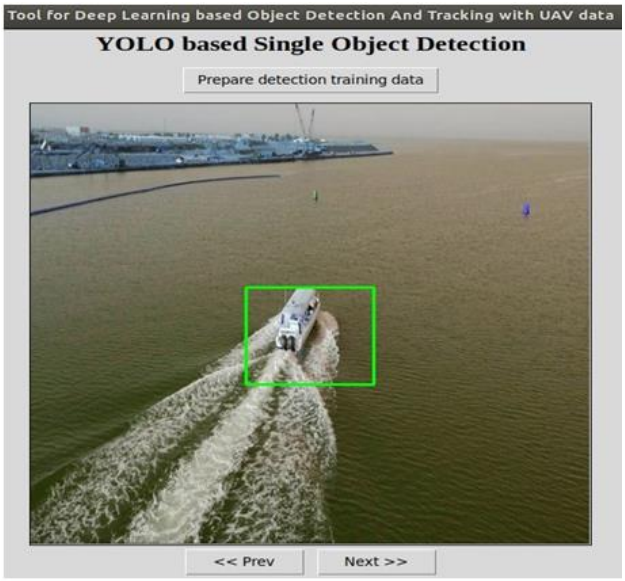

Figure 7. Object detection with YOLO

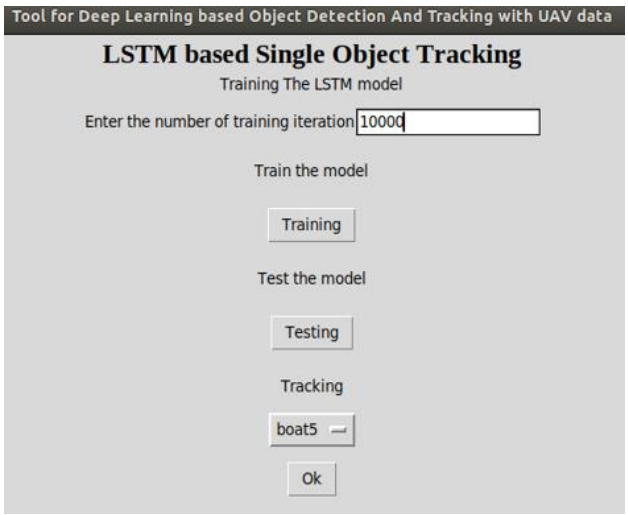

Figure 8. Training LSTM

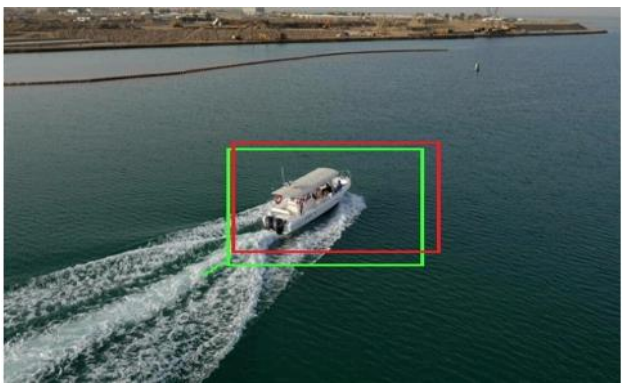

Figure 9. Tracked object with trajectory. The red box indicates the groundtruth and green box indicates the tracked object

3.3 Interactive Module 3: Multiple Object Detection

The steps involved in multiple object detection are:

1. Video to frame conversion.
2. Image annotation.
3. The .xml files obtained from annotated images are converted into .csv files.
4. Faster R-CNN configuration file is modified according to the number of classes and training iteration.
5. Frozen inference graph is generated with the highest numbered trained checkpoint.
6. Multiple object detection is performed.

In multiple object detection, the mode of obtaining input is either video or image folder. LabelImg annotator ("LabelImg," 2015) is integrated with the tool for annotating multiple objects as shown in Figure 10.

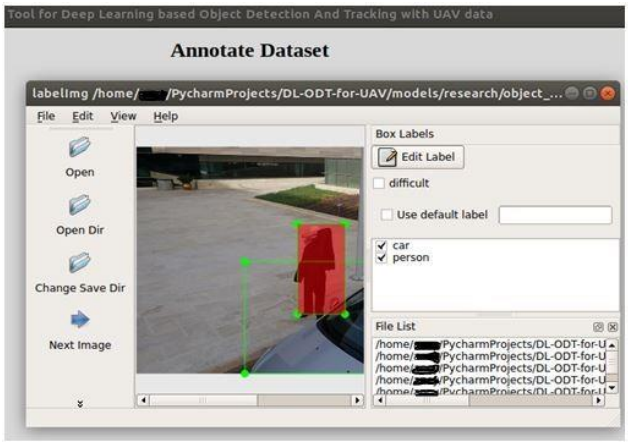

Figure 10. LabelImg annotator

The annotated files contain the class name and the bounding box locations of the respective objects. The annotated .xml files are converted into .csv files. The class name of the annotated objects is obtained from the user to update the Faster R-CNN configuration (Figure 11). The number of training iteration is specified by the user to initiate training (Figure 12). After the training, object detection is performed with trained checkpoints. The multiple objects detected with the tool is shown in Figure 13.

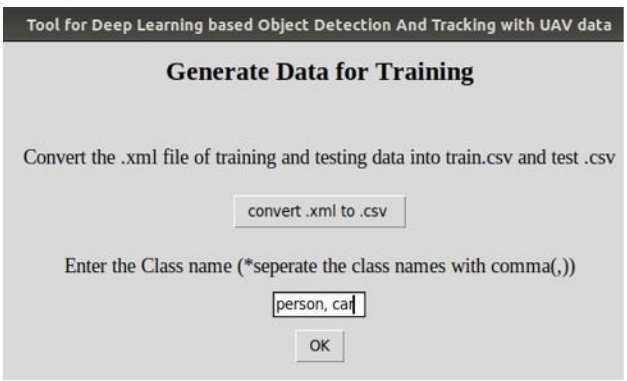

Figure 11. Generate data for Faster R-CNN training

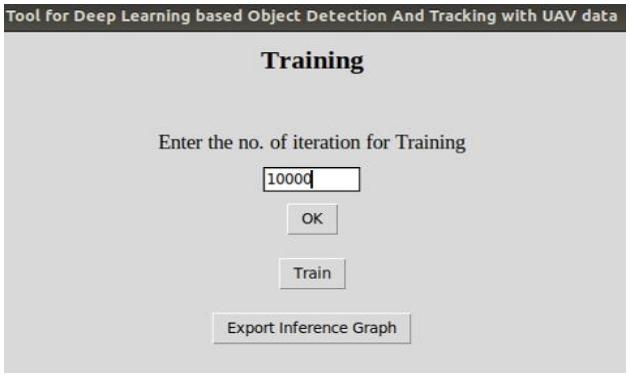

Figure 12. Training Faster R-CNN

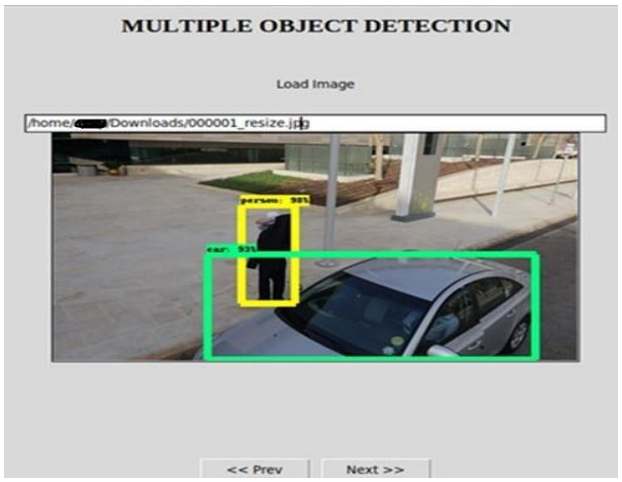

Figure 13. Multiple object detection

4. EVALUATION

The tool performs well with high definition images and videos. The UAV123 dataset is used in the DL-ODT-UAV (Mueller et al., 2016). Metrics such as precision and recall is used to evaluate interactive module 2 and 3 (Equation 3 and 4). The precision and recall obtained for single object detection and tracking module is 90.83% and 92.09% and for multiple object detection module is 91.23% and 93.51% (Table 1).

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{3}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{4}$$

Where TP - True Positive(actual object), FP - False Positive(false alarm), FN - False Negative(missed object)

Module	Technique	Precision	Recall
Single Object Detection and Tracking	YOLO + LSTM	90.83%	92.09%
Multiple Object Detection	Faster R-CNN	91.23%	93.51%

Table 1. Performance metrics of interactive modules

5. OUTCOMES OF DL-ODT-UAV TOOL

With the practical exploration of DL-ODT-UAV tool, students/researchers will possess knowledge about:

1. Preliminary steps required for DL based object detection.
2. Image annotation.
3. How different image annotating formats work with object detectors.
4. The performance difference between single-shot and region-based based object detectors.
5. How the number of samples and training iterations will affect the final outcome.
6. Object detection and tracking with UAV data.

The tool will enable the user to perform object detection and tracking in various applications like surveillance, crowd monitoring and traffic analysis.

6. CONCLUSION

The growth in UAV and deep learning methodologies have been significant in the last decade. With technological advancement in both fields, there is a need for practical exploration of UAV data with deep learning methodologies. This paper proposes a tool for students/researchers to work with deep learning methodologies in UAV data. Single object detection and tracking is performed with YOLO and LSTM. The tool is designed with a semi-automatic YOLO based single object annotator to reduce annotation time. Faster R-CNN performs multiple object detection. LabelImg annotator is integrated with the tool to annotate multiple objects. Single object detection and tracking exhibits precision as 90.83% and recall as 92.09%. Multiple object detection exhibits precision as 91.23% and recall as 93.51%. The tool serves as an efficient resource for students/researchers who are eager to explore UAV data with deep learning-based object detection and tracking.

ACKNOWLEDGEMENTS

The authors thank the International Society for Photogrammetry and Remote Sensing(ISPRS) for funding the project under ISPRS Scientific Initiatives 2019.

REFERENCES

Ballard, W., 2018. Hands-on deep learning for images with TensorFlow : build intelligent computer vision applications using TensorFlow and Keras.

BBox-Label-Tool, 2017. https://github.com/puzzledqs/BBox-Label-Tool (19 March 2020).

Berie, H.T., Burud, I., 2018. Application of unmanned aerial vehicles in earth resources monitoring: Focus on evaluating potentials for forest monitoring in Ethiopia. Eur. J. Remote Sens. https://doi.org/10.1080/22797254.2018.1432993

Cyprian Aleksander, K., 2018. Military Use of Unmanned Aerial Vehicles – A Historical Study. Saf. Def. 4, 17–21. https://doi.org/10.37105/sd.4

Erdelj, M., Natalizio, E., 2016. UAV-assisted disaster

management: Applications and open issues, in: 2016 International Conference on Computing, Networking and Communications, ICNC 2016. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCNC.2016.7440563

Fang, P., Lu, J., Tian, Y., Miao, Z., 2011. An improved object tracking method in UAV videos. Procedia Eng. 15, 634–638. https://doi.org/10.1016/j.proeng.2011.08.118

Farmers use drones to spray pesticide, 2019. https://www.thehindu.com/news/cities/Hyderabad/now-farmers-use-drones-to-spray-pesticide/article30342501.ece (8 April 2020).

Grayson, J.E., 2000. Python and Tkinter programming. Manning.

Haque, M.R., Muhammad, M., Swarnaker, D., Arifuzzaman, M., 2014. Autonomous Quadcopter for product home delivery, in: 1st International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2014. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICEEICT.2014.6919154

Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D., Raja Segaran, R., Reid, I., Terauds, A., Koh, L.P., 2018. Drones count wildlife more accurately and precisely than humans. Methods Ecol. Evol. 9, 1160–1167. https://doi.org/10.1111/2041-210X.12974

Inception v2 model, 2017. https://github.com/tensorflow/models/tree/master/research/slim (12 April 2020).

Khan, M.A., Ectors, W., Bellemans, T., Janssens, D., Wets, G., 2017. UAV-Based Traffic Analysis: A Universal Guiding Framework Based on Literature Survey, in: Transportation Research Procedia. Elsevier B.V., pp. 541–550. https://doi.org/10.1016/j.trpro.2017.03.043

LabelImg, 2015. https://github.com/tzutalin/labelImg (21 January 2020).

Madhavan, R., Silva, T., Farina, F., Wiebbelling, R., Renner, L., Prestes, E., 2018. Unmanned Aerial Vehicles for Environmental Monitoring, Ecological Conservation, and Disaster Management, in: Technologies for Development. Springer International Publishing, pp. 31–39. https://doi.org/10.1007/978-3-319-91068-0_3

Mapping India through drones, 2019. https://economictimes.indiatimes.com/news/economy/policy/technology-topography-mapping-india-through-drones/articleshow/72179542.cms (11 April 2020).

Micheal, A.A., Vani, K., 2018. Comparative analysis of SIFT and SURF on KLT tracker for UAV applications. Proc. 2017 IEEE Int. Conf. Commun. Signal Process. ICCSP 2017 2018-Janua, 1000–1003. https://doi.org/10.1109/ICCSP.2017.8286523

Mueller, M., Smith, N., Ghanem, B., 2016. A benchmark and simulator for UAV tracking, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 445–461. https://doi.org/10.1007/978-3-319-46448-0_27

Ndna, M., Tss, D., 2017. Use of unmanned aerial vehicles in crime scene investigations - novel concept of crime scene investigations. https://doi.org/10.15406/frcij.2017.04.00094

Ning, G., Zhang, Z., Huang, C., Ren, X., Wang, H., Cai, C., He, Z., 2017. Spatially supervised recurrent convolutional neural networks for visual object tracking. Proc. - IEEE Int. Symp. Circuits Syst. 1–4. https://doi.org/10.1109/ISCAS.2017.8050867

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rohan, A., Rabah, M., Kim, S.H., 2019. Convolutional Neural Network-Based Real-Time Object Detection and Tracking for Parrot AR Drone 2. IEEE Access 7, 69575–69584. https://doi.org/10.1109/ACCESS.2019.2919332

TensorFlow-Object-Detection-API,2019. https://github.com/EdjeElectronics/TensorFlow-Object-Detection-API-Tutorial-Train-Multiple-Objects-Windows-10 (12 April 2020).

Teutsch, M., Krüger, W., 2012. Detection, segmentation, and tracking of moving objects in UAV videos. Proc. - 2012 IEEE 9th Int. Conf. Adv. Video Signal-Based Surveillance, AVSS 2012 313–318. https://doi.org/10.1109/AVSS.2012.36

Tong, X., Zhang, Y., Yang, T., Ma, W., 2013. Automatic object tracking in aerial videos via spatial-temporal feature clustering. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8261 LNCS, 78–85. https://doi.org/10.1007/978-3-642-42057-3-11

Xu, S., Savvaris, A., He, S., Shin, H.S., Tsourdos, A., 2018. Real-time Implementation of YOLO+JPDA for Small Scale UAV Multiple Object Tracking. 2018 Int. Conf. Unmanned Aircr. Syst. ICUAS 2018 1336–1341. https://doi.org/10.1109/ICUAS.2018.8453398

Xu, Y., Yu, G., Wu, X., Wang, Y., Ma, Y., 2017. An Enhanced Viola-Jones Vehicle Detection Method from Unmanned Aerial Vehicles Imagery. IEEE Trans. Intell. Transp. Syst. 18, 1845–1856. https://doi.org/10.1109/TITS.2016.2617202

Yu, W., Yu, X., Zhang, P., Zhou, J., 2008. a New Framework of Moving Target Detection and Tracking for. Archives XXXVII, 606–614.

APPENDIX

The DL-ODT-UAV tool can be found in the link https://github.com/ancymicheal/DL-ODT-for-UAV