PhD thesis topic – September 2022

Object Detection and Tracking Approaches for Aerial Images Using Deep Learning-Based Algorithms

FR Math-STIC – Axe 1

Hanane AZZAG (LIPN) and Fangchen FENG (L2TI)

May 9, 2022

1 Context and topic description

Aerial images can be considered as special type of images captured from airplanes, drones or satellites with different altitudes and resolutions. The object detection benchemarking on different datasets shows that the performance is higher on natural images (e.g. PASCAL VOC, MS COCO) than on aerials images such as DOTA and DIOR datasets. Indeed, the recent research reveals that several challenges still remain with current techniques for aerial images. One example is that, for training-based applications, a huge computational time is required during the training stage due to the potentially high resolution of the aerial images, which often leads to performance degradation with limited computational resources.

In this thesis topic, we are interested in the development of Object Detection (OD), Object Tracking (OT) and Multi-Object Tracking (MOT) deep learning-based algorithms in aerial images. In particular, we will address the following challenges:

1. **Detection of Small Size Objects** – As mentioned in [1], the object size on the image has a strong impact on the detection performance because of the amount of detail representing the object. The development of models such as Faster R-CNN, a deep convolutional network used for object detection, appears to the user as a single, end-to-end, unified network. It is one of the most important models employed in OD because of its architecture and its good speed/accuracy relation [1, 2]. single-shot Detector (SSD) is another important object detection model. Without the delegated region proposal network, it predicts the boundary boxes and the classes directly from feature maps in one single pass [3]. Moreover, the architectures with different feature extractors [1] can guarantee a better equilibrium between the accuracy and the speed for the detection problem. However, these methods are not well suited for the detection of small objects.

2. **Temporal Analysis of Videos** – The object tracking is usually associated with video analysis which is difficult to implement in real-time, requiring significant hardware power to identify and track each object [2]. Camera movement during the detection is another big challenge as it is shown in [2] that the influence of the camera movement decreases the tracking performance in the framework of video analysis. It is therefore interesting to exploit different models and strategies to search for the best architecture for both the detection and the continuous tracking problem. The link between the object tracking and the video analysis will also be investigated.

3. **Scalability of Models for Real-Time Applications** – The tracking performance usually depends on the computational power which limits the number of images to be analyzed per second (FPS) [1]. GPU is often used in OD to accelerate the training process and to increase the FPS for implementation. It is therefore possible to include the with/without implementation comparison while analyzing the tracking process. Another interesting research track is the application of different strategies to avoid the temporal analysis directly by the model. One possible way is to adapt the STAR-RGB method [4] for the object tracking.

4. **Quantity of objects detected** – Another important challenge is the simultaneous detection and tracking of multiple objects which generally limit the FPS [5]. As mentioned previously, the MOT application requires the GPU implementation for the FPS to be acceptable. It is therefore important to analyze the performance of different models combined with and without the use of GPUs. A qualitative evaluation on the influence of FPS in real applications will also be performed.

2 Work Plan

We expect the work to be developed along the following plan:

1. Year 1: Bibliographic research will first be carried out. Preliminary work towards objective 1 is expected. The best models will then be selected and will be followed by work towards objective 2.
2. Year 2: With the results of the first year, the best models will be upgrade
3. Year 3: Once all the objectives have been achieved, it will be possible to open up to perspectives and carry out more tests for real situations. The writing of the manuscript will then be expected

3 Supervisors

- Supervisor: Hanane AZZAG - Maître de conférences HDR (LIPN, CNRS UMR 7030)
- Co-supervisor: Fangchen FENG - Maître de conférences (L2TI, UR 3043)

References

[1] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al., "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7310–7311.

[2] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu, "A survey of deep learning-based object detection," *IEEE access*, vol. 7, pp. 128837–128868, 2019.

[3] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[4] Wyctor Fogos Da Rocha, Pablo Pereira E Silva, Luiza Emília VN Mazzoni, Clebeson Canuto, Rafhael Milanezi De Andrade, Antônio Bent, Mariana Rampinelli, and Douglas Aumonfrey, "The use of convolutional neural network and StarRGB technique for gait movements recognition in remote physiotherapy," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. IEEE, 2021, pp. 01–06.

[5] Haogang Feng, Gaoze Mu, Shida Zhong, Peichang Zhang, and Tao Yuan, "Benchmark analysis of Yolo performance on edge intelligence devices," *Cryptography*, vol. 6, no. 2, pp. 16, 2022.