Algorithm evaluation for parallel detection and tracking of UAVs

Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn, and Georg Schitter

Automation and Control Institute (ACIN), TU Wien, Gusshausstrasse 27-29, Vienna, Austria

1. ABSTRACT

This paper presents the evaluation of object detectors and trackers within a parallel software architecture to enable long distance UAV detection and tracking in real-time using a telescope-based optical system. The architecture combines computationally expensive deep learning-based object detectors with traditional object trackers to achieve a detection and tracking rate of 100 fps. Four object detectors, FRCNN, SSD, Retinanet and FCOS, are fine-tuned on a custom UAV dataset and integrated together with three trackers, Medianflow, KCF and MOSSE, into a parallel software architecture. The evaluation is conducted on a separate set of test images and videos. The combination of FRCNN and Medianflow shows the best performance in terms of intersection over union and center location offset on the video test set, enabling detection and tracking of UAVs at 100 fps.

Keywords: Deep learning, detection, tracking, real-time, UAV

2. INTRODUCTION

The usage of unmanned aerial vehicles (UAVs) has increased drastically over the past decade due the relatively cheap availability and versatility of the technology.[1] However, safety concerns and safety related incidents have soared equally in the past, as numerous examples of near or actual collisions with commercial air planes in various countries demonstrate.[2,3] Another perilous situation includes a near collision with an air ambulance in a height of 400 m.[4] Similarly, UAVs offer a menacing potential to endanger critical infrastructure like nuclear power plants[5] and prove useful to smuggle drugs across state borders and into prisons.[6] As these examples illustrate the emanating danger posed by UAVs, deployment of appropriate detection systems is crucial to ensure timely threat reconnaissance.

Generally, UAV detection systems combine multiple sensors into holistic systems, which consist of RADAR,[7] LiDAR,[8] acoustics,[9] radio frequency (RF)[10] and electro-optics (EO) sensors.[11] The latter is a key component, as it allows a profound situational evaluation through visual imagery. EO systems have a relatively narrow field of view (FOV) camera, which are typically attached to a pan and tilt mount to enable monitoring a larger area. Utilization of telescope systems can further increase the operational range to multiple kilometres for small UAVs.[11] These EO systems rely on efficient computer vision algorithms to extract the UAV position within video frames to steer the mount and keep the UAV within the FoV.

Over the past decade deep learning based methods have proven to be the most promising approach to detect and classify objects in challenging imagery. Algorithms like YOLO,[12] SSD,[13] Retinanet,[14] FRCNN[15] or FCOS[16] are prominent examples for state of the art object detection. However, neural networks come at a high processing cost and therefore, are limited in the achievable frame rates. A high frame rate and the resulting high number of UAV localizations per second is an important prerequisite for a dynamic pan and tilt system, which has to keep up with the fast and agile movements of a UAV.[17] Combinations of neural networks with faster object tracking algorithms in a parallel manner strive to benefit from both, high accuracy of the neural networks and the fast processing speeds of tracking algorithms.[18] Fast tracking algorithms include Mosse,[19] KCF,[20] MedianFlow[21] or CSRT.[22] Combining neural networks with Kalman Filters offers improved frame rates too, however, compared to the trackers, a filter only estimates the position between two detections.[23] Various studies exist comparing object tracking and object detection algorithms for the use case of UAV detection.[24–26] To enable an efficient detection and tracking of UAVs at high frame rates, a thorough analysis and comparison of different detector

Further author information: Send correspondence to Denis Ojdanić
E-mail: ojdanic@acin.tuwien.ac.at, Telephone:+43 (0) 1 58801 376 520

Post-Print version of the article: Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn and Georg Schitter, "Algorithm evaluation for parallel detection and tracking of UAVs," *In Optics, Photonics, and Digital Technologies for Imaging Applications VIII*, vol. 12998, pp. 327-333. SPIE, 2024. DOI: 10.1117/12.3017037

and tracker combinations within a parallel architecture is required.
The contribution of this paper is the evaluation and comparison of various deep learning based object detectors and traditional object trackers within a parallel architecture to enable detection and tracking of UAVs at 100 fps. A high number of UAV localizations enables a dynamic actuation of a pan-tilt mount to track the fast and agile movements of the UAV.

3. OBJECT DETECTION AND TRACKING

To implement a system detecting UAVs at 100 fps a parallel software architecture is used, where the detector and tracker are running within separate threads.[27] The slow, but accurate deep learning based detector initializes a fast, but less accurate object tracker to achieve a high frame rate, while maintaining a high accuracy.[27] This combination of a deep learning object detector with a fast object tracker enables detection and tracking of objects at 100 fps. For object detection FRCNN, FCOS, RetinaNet and SSD are selected to be trained and evaluated as they cover both single and dual stage network architectures as well as anchor box and anchor box free approaches. For object tracking, MedianFlow, Mosse and KCF are used, as these algorithms have low processing demands, which makes them perfect candidates to be paired with a deep learning algorithm in parallel to increase the UAV localization frame rate.

3.1 Training

In order to train the four deep learning object detectors a pre-labelled training dataset consisting of approximately 18000 images is used.[27] The dataset contains images of different quad-copter UAVs in front of various backgrounds. The distribution of the bounding boxes is depicted in Fig. 1, which illustrates that the UAVs within the dataset cover a small pixel area with respect to the image size. Nevertheless, to further decrease the number of false positives during inference, about 8 % of pure background images, which do not contain any UAV are added to the dataset extending it to about 20500 images. These background images do contain other flying objects, like air planes or birds, in order to avoid erroneous detections during inference. 8 % of the 18000 images, which contain UAVs, serve as validation dataset and the remaining images with all the added background images are used for training.

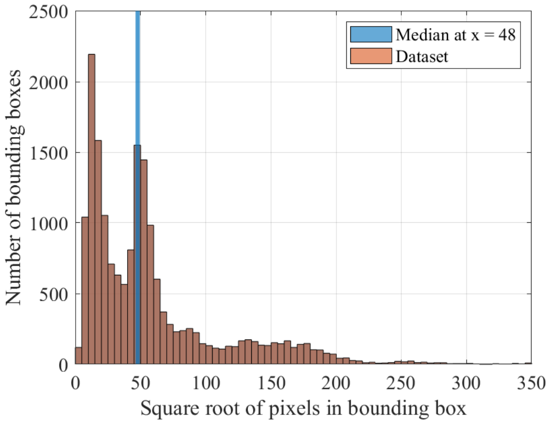


Figure 1: Distribution of the bounding box area in pixels within the dataset. Note that the pure background images are not depicted within this plot, as these images have no bounding box.

As a training strategy for the neural networks, fine-tuning is applied, whereas the networks are initialized with the weights trained on the COCO dataset.[28] The training and the experiments in Section 4 are conducted on an RTX 3080 (Nvidia Corporation, Santa Clara, California, USA) with 10 GB of GPU RAM. Additionally, the PC has an AMD Ryzen 3900 CPU (Advanced Micro Devices, Inc., Santa Clara, California, USA) with 12

Post-Print version of the article: Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn and Georg Schitter, “Algorithm evaluation for parallel detection and tracking of UAVs,” *In Optics, Photonics, and Digital Technologies for Imaging Applications VIII*, vol. 12998, pp. 327-333. SPIE, 2024. DOI: 10.1117/12.3017037

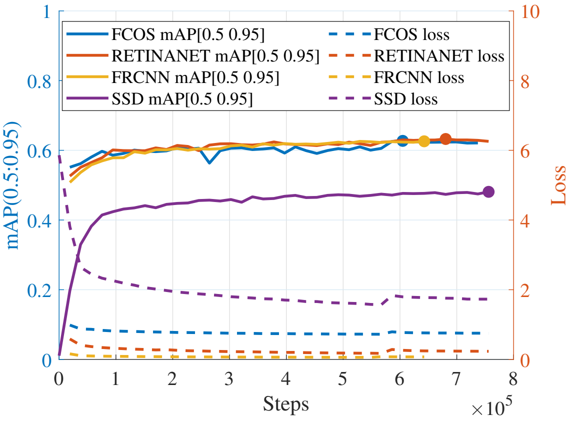


Figure 2: Overview of the training process showing the loss curves and the mAP(0.5:0.95) on the validation dataset for each deep learning object detector. The dots on the validation curve indicate the best performing model, which is later used for inference.

Table 1: Hyper-parameters for fine-tuning of the deep learning object detectors.

Algorithm	Learning rate	Weight decay	Momentum	Batch size
SSD	0.0001	0.0006	0.6	8
FCOS	0.0005	0.0007	0.8	4
FRCNN	0.0002	0.0007	0.8	2
RetinaNet	0.0003	0.0007	0.8	4

cores, 24 threads and 32 GB of RAM installed.

The algorithms are trained for approximately 40 epochs and the further hyper-parameters are summarized in Table 1. To diversify the dataset, data augmentation techniques are applied during the training process such as random horizontal flipping and color transformations of the images including greyscale transformation or changing the brightness, contrast, saturation and hue.[29] Furthermore, random cropping is applied after epoch 30, as activating the cropping earlier has led to an exploding loss. Fig. 2 depicts the loss and the mean average precision (mAP) on the validation dataset during the training process. A slight increase of the loss is visible for all models after random cropping is activated. The best model according to a mAP of 0.5 to 0.95 is saved during the training process and the dots on the validation curves represent the best performing models, which are selected for the experiments.

4. EXPERIMENTS AND RESULTS

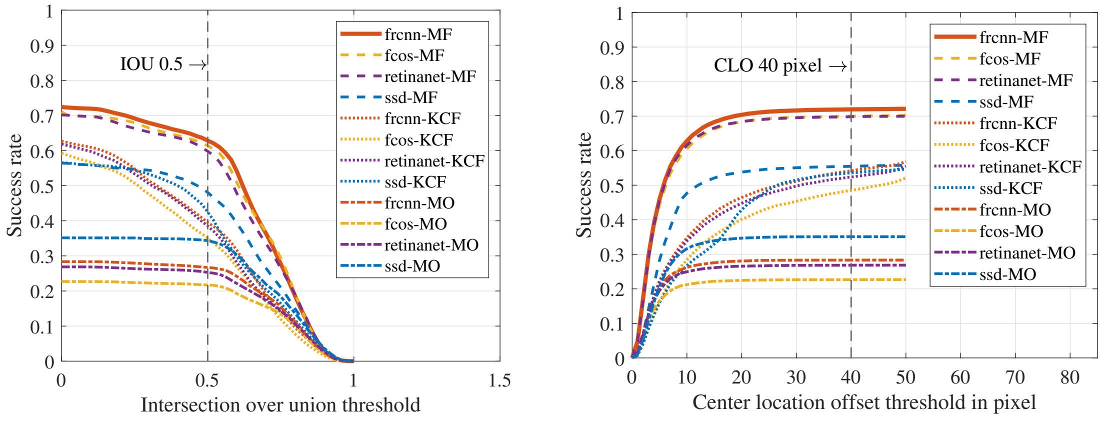
For experimental evaluation a test video dataset of approximately 25000 video images of UAVs in front of different backgrounds is used.[11] Most of the images are captured using telescope systems. Additionally, of each video sequence every fifth image is extracted to form a test image dataset of 5000 images to analyse the mAP of the detectors.

Table 2 shows the results of applying the four deep learning object detectors onto the test image dataset. FRCNN scores the highest mAP(0.5:0.95) of 40.7 % closely followed by FCOS at 37.9 % and RetinaNet at 35.6 %. SSD, as the fastest and least complex algorithm, achieves 34.2 %.

To evaluate the combined results of detector and tracker running in parallel, the test videos serve as input at a frame rate of 100 fps. To achieve this detection and tracking frame rate, the slow object detectors are combined with fast object trackers.[27] To verify, whether the algorithms are localizing the object correctly, the detector and tracker output are merged based on their prediction probability.[27] The intersection over union (IOU) and

Table 2: mAP for the deep learning algorithms.

Algorithm	mAP(0.5)	mAP(0.5:0.95)
FRCNN	88.7 %	40.7 %
FCOS	89.6 %	37.9 %
RetinaNet	85.6 %	35.6 %
SSD	75.9 %	34.2 %



(a) IOU for all detector and tracker combinations.

(b) CLO for all detector and tracker combinations.

Figure 3: The IOU and CLO of all algorithm combinations on the test video dataset.

the center location offset (CLO) serve as evaluation metrics,[30] whereas an IOU of 0.5 or larger is considered as successful and and CLO of 40 pixels or less.

Fig. 3 depicts the results of applying all detector and tracker combinations to the video dataset and Table 3 summarizes the IOU at a threshold of 0.5 and the CLO at a pixel threshold of 40. The best combined performance is achieved by FRCNN and the MedianFlow tracker, which score an IOU of 62.8 % and an CLO of 72 % at a threshold of 0.5 and 40 pixels. The latter tracking algorithm also achieves the best results when compared to the other object trackers. Among the detectors, SSD scores the lowest results, when combined with MedianFlow, which ensues from the evaluation on the image test set as presented in Table 3. However, in combination with lower performing trackers like KCF or Mosse, it performs better compared to the other deep learning detectors. SSD consists of a simpler network architecture and therefore, processes approximately twice as many images per second as the other deep learning detectors. For the less reliable trackers, more frequent reinitialisations as facilitated by SSD, result in an higher combined performance compared to the other detector, when paired with Mosse or KCF. Finally, Fig. 4 shows an example video sequence of a detection and track using the parallel architecture with FRCNN and MedianFlow.

To summarize, among the analysed algorithms, FRCNN and MedianFlow achieve the best performance, when detecting and tracking UAVs at 100 fps scoring an IOU(0.5) of 62.8 %. This enables detection and tracking of UAVs flying at fast velocities in dynamic trajectories.

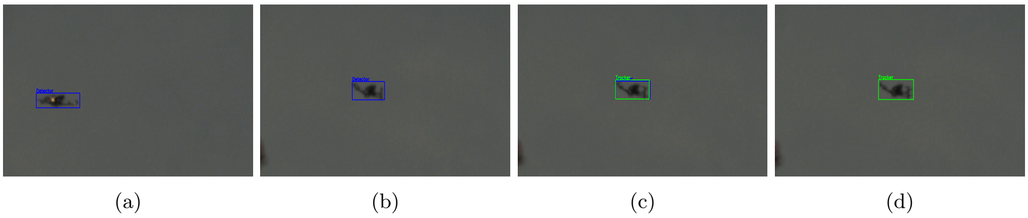
5. CONCLUSION

An algorithm analysis is presented for a real-time parallel software architecture, which combines accurate deep neural network detectors with traditional trackers to enable UAV localization at 100 fps. FRCNN, RetinaNet, SSD and FCOS are selected as detectors and fine-tuned for the task of UAV detection. The detectors together with three trackers, Medianflow, KCF and Mosse, are analysed on a separate test image and video dataset. An evaluation is performed to determine, which combination of neural network and object tracker yields the best

Post-Print version of the article: Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn and Georg Schitter, "Algorithm evaluation for parallel detection and tracking of UAVs," *In Optics, Photonics, and Digital Technologies for Imaging Applications VIII*, vol. 12998, pp. 327-333. SPIE, 2024. DOI: 10.1117/12.3017037

Table 3: The results of the IOU(0.5) and the CLO(40 pixels) for each algorithm combination. The results are sorted in a descending order according to the IOU(0.5).

Algorithm	IOU(0.5)	CLO(40 px))
FRCNN - MedianFlow	62.8 %	72.0 %
FCOS - MedianFlow	61.4 %	70.0 %
RetinaNet - MedianFlow	59.8 %	69.8 %
SSD - MedianFlow	48.1 %	55.5 %
SSD - KCF	42.5 %	53.4 %
FRCNN - KCF	39.5 %	54.0 %
RetinaNet - KCF	38.5 %	52.2 %
FCOS - KCF	35.0 %	48.3 %
SSD - Mosse	34.3 %	35.1 %
FRCNN - Mosse	26.7 %	28.3 %
RetinaNet - Mosse	25.3 %	26.8 %
FCOS - Mosse	21.7 %	22.6 %



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

Figure 4: Image sequence of a detection and track of a UAV. The blue and green bounding boxes indicate the detector the tracker output.

results in terms of IOU at an threshold of 0.5 and the CLO at an pixel threshold of 40. FRCNN and MedianFlow prove to be the best performing algorithm combination by score an IOU of 62.8 % and a CLO of 72 % when detecting and tracking UAVs at a frame rate of 100 fps. By localizing UAVs at 100 fps, a dynamic control of the pan-tilt mount for the telescope is possible, which enables tracking and keeping fast and agile UAVs within the FoV.

ACKNOWLEDGMENTS

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).
The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

REFERENCES

[1] Liew, C. F., DeLatte, D., Takeishi, N., and Yairi, T., “Recent developments in aerial robotics: A survey and prototypes overview,” *ArXiv* **abs/1711.10085** (2017).

[2] “Drone collides with commercial aeroplane in canada.” BBC News (Oct. 2017). Accessed Feb 2024.

[3] “Drone over windsor came close to british airways plane, report says.” BBC News (Dec. 2023). Accessed Feb 2024.

[4] “Drone in ’near-miss’ with air ambulance.” BBC News (Sept. 2019). Accessed Feb 2024.

[5] Phillips, C. and Gaffey, C., “Most French Nuclear Plants ’Should Be Shut Down’ Over Drone Threat.” Newsweek Magazine (Feb. 2015). Accessed Feb 2022.

[6] Daly, M., “Cheap and They Don’t Snitch: Drones Are the New Drug Mules.” VICE (Jan. 2024). Accessed Feb 2022.

[7] Drozdowicz, J., Wielgo, M., Samczynski, P., Kulpa, K., Krzonkalla, J., Mordzonek, M., Bryl, M., and Jakielaszek, Z., "35 GHz FMCW drone detection system," in [*2016 17th International Radar Symposium (IRS)*], 1–4, IEEE (2016).

[8] Dogru, S. and Marques, L., "Drone detection using sparse lidar measurements," *IEEE Robotics and Automation Letters* **7**(2), 3062–3069 (2022).

[9] Mezei, J., Fiaska, V., and Molnár, A., "Drone sound detection," in [*16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*], 333–338, IEEE (2015).

[10] Nguyen, P., Ravindranatha, M., Nguyen, A., Han, R., and Vu, T., "Investigating cost-effective rf-based detection of drones," in [*Proceedings of the 2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*], 17–22 (2016).

[11] Ojdanić, D., Sinn, A., Naverschnigg, C., and Schitter, G., "Feasibility analysis of optical uav detection over long distances using robotic telescopes," *IEEE Transactions on Aerospace and Electronic Systems* **59**(5), 5148–5157 (2023).

[12] Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M., "Yolov4: Optimal speed and accuracy of object detection," (2020).

[13] Wei, L., Anguelov, D., Erhan, D., et. al., "SSD: Single Shot MultiBox Detector," in [*Computer Vision – ECCV 2016*], 21–37, Springer International Publishing, Cham (2016).

[14] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P., "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(2), 318–327 (2020).

[15] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in [*Advances in Neural Information Processing Systems*], **28**, Curran Associates, Inc. (2015).

[16] Tian, Z., Shen, C., Chen, H., and He, T., "Fcos: Fully convolutional one-stage object detection," (2019).

[17] Naverschnigg, C., Ojdanic, D., Sinn, A., and Schitter, G., "Analysis and Control of a Robotic Telescope System for High-Speed Small-UAV Tracking," *IEEE Transactions on Aerospace and Electronic Systems Magazine* (submitted 12.2023).

[18] Lee, D.-H., "CNN-based single object detection and tracking in videos and its application to drone detection," *Multimedia Tools and Applications* **80**, 34237–34248 (Oct 2020).

[19] Bolme, D., Beveridge, J. R., Draper, B. A., and Lui, Y. M., "Visual object tracking using adaptive correlation filters," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], IEEE (jun 2010).

[20] Henriques, J. F., Caseiro, R., Martins, P., and Batista, J., "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**, 583–596 (Mar. 2015).

[21] Kalal, Z., Mikolajczyk, K., and Matas, J., "Forward-Backward Error: Automatic Detection of Tracking Failures," in [*2010 20th International Conference on Pattern Recognition*], IEEE (aug 2010).

[22] Danelljan, M., Hager, G., Shahbaz Khan, F., and Felsberg, M., "Learning spatially regularized correlation filters for visual tracking," in [*Proceedings of the IEEE international conference on computer vision*], 4310–4318 (2015).

[23] Opromolla, R. and Fasano, G., "Visual-based obstacle detection and tracking, and conflict detection for small uas sense and avoid," *Aerospace Science and Technology* **119**, 107167 (2021).

[24] Park, J., Kim, D. H., Shin, Y. S., and Lee, S.-h., "A comparison of convolutional object detectors for real-time drone tracking using a ptz camera," in [*17th International Conference on Control, Automation and Systems (ICCAS)*], 696–699 (2017).

[25] Oh, H. M., Lee, H., and Kim, M. Y., "Comparing convolutional neural network(cnn) models for machine learning-based drone and bird classification of anti-drone system," in [*19th International Conference on Control, Automation and Systems (ICCAS)*], 87–90 (2019).

[26] Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., Kämäräinen, J.-K., Chang, H. J., Danelljan, M., Zajc, L. Č., Lukežič, A., et al., "The tenth visual object tracking vot2022 challenge results," in [*European Conference on Computer Vision*], 431–460 (2022).

[27] Ojdanić, D., Sinn, A., Naverschnigg, C., Zelinskyi, D., and Schitter, G., "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," *IEEE Transactions on Aerospace and Electronic Systems* (submitted 07.2023).

[28] Lin, T., Maire, M., Belongie S., et. al., "Microsoft COCO: Common Objects in Context," in [*Computer Vision–ECCV 2014, Proceedings, Part V 13*], 740–755, Springer (2014).

[29] Shorten, C. and Khoshgoftaar, T. M., "A survey on image data augmentation for deep learning," *Journal of big data* **6**(1), 1–48 (2019).

[30] Fan, H. and Ling, H., "Parallel Tracking and Verifying," *IEEE Transactions on Image Processing* **28**, 4130–4144 (aug 2019).