UAV Detection and Classification

No Author Given

No Institute Given

Abstract. —Unmanned aerial vehicles (UAV) have been widely used in various fields, and their invasion of security and privacy has aroused social concern. Several detection and tracking systems for UAVs have been introduced in recent years, but most of them are based on radio frequency, radar, and other media. We assume that the field of computer vision is mature enough to detect and track invading UAVs.We use visible light mode dataset called Dalian University of Technology Anti UAV dataset, DUT Anti-UAV for short. It contains a detection dataset with a total of 10,000 images. This dataset to train several existing detection algorithms and evaluate the algorithms' performance. Several tracking methods are also tested on our tracking dataset.

1 Introduction

With the advancement of industrial technology, Unmanned Aerial Vehicles (UAVs) have increasingly become mainstream and are widely adopted in various sectors such as logistics, transportation, and surveillance due to their compact design, affordability, and ease of operation. However, despite offering numerous benefits, UAVs also raise significant concerns regarding public safety, personal privacy, and potential unauthorized or accidental intrusions into restricted areas. As a result, detecting and tracking UAVs—particularly those used for illegal or unintended purposes—has become critically important.

Currently, many UAV detection systems rely on radar, radio frequency (RF), and acoustic sensors. While these technologies can be effective, they often suffer from drawbacks such as high costs, sensitivity to environmental noise, and limited coverage, restricting their deployment primarily to sensitive zones like airports and major public venues. These limitations highlight the need for more accessible and scalable UAV detection solutions.

In this context, vision-based detection using camera sensors offers a promising alternative due to its cost-effectiveness and potential for wide-scale deployment. However, vision-based UAV detection presents its own challenges, including small target size, background complexity, and UAVs blending into noisy or cluttered scenes. UAVs may follow irregular flight paths and often become occluded or confused with visually similar objects, making reliable tracking even more difficult.

The emergence of Deep Learning (DL) has significantly improved performance across various computer vision tasks, particularly in object detection and tracking. Convolutional Neural Networks (CNNs) and architectures such as

Faster R-CNN, SSD, YOLO have demonstrated remarkable capabilities. Nevertheless, when directly applied to UAV detection, these generic models often fall short due to the unique characteristics of UAV imagery—mainly the small object size, dynamic motion, and complex environments.

Our motivation stems from the need to adapt state-of-the-art detection and tracking frameworks specifically for anti-UAV tasks. This involves addressing the problem at both the data and algorithmic levels. Deep learning methods require extensive and diverse training data to achieve robust and accurate results. However, acquiring such data—especially under varying environmental conditions and regulatory constraints—is challenging. Issues such as background shift (e.g., training on sky-dominated scenes while testing involves ground, trees, or mixed backgrounds), limited annotations, and distribution changes further complicate generalization.

Therefore, our work focuses on leveraging and adapting deep learning-based models to effectively tackle the unique challenges of UAV detection and tracking, ensuring reliable performance even in complex, real-world conditions.

2 Related Work

2.1 Model-Based Methods (Traditional Computer Vision Approaches)

Before the widespread use of deep learning, object detection primarily relied on handcrafted features and classical computer vision techniques. These model-based methods involved manual feature extraction and traditional machine learning algorithms for classification and localization:

- **Haar-like Features and Cascade Classifiers:** Employed integral images and the AdaBoost algorithm for efficient object detection.
- **Histogram of Oriented Gradients (HOG) + SVM:** Combined gradient-based feature descriptors with Support Vector Machines for detecting objects in images.
- **Deformable Parts Model (DPM):** Modeled objects as collections of parts with spatial constraints; this approach won the Pascal VOC challenge in 2009.
- **Background Subtraction + Classifiers:** Used motion-based background subtraction in static video settings, followed by CNN classifiers to distinguish between drones, birds, and backgrounds.

While effective in controlled environments, these techniques struggled with small object detection, background clutter, and object motion, limiting their applicability in real-world drone detection scenarios.

2.2 Learning-Based Methods (Deep Learning Approaches)

The emergence of Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), transformed object detection by enabling automatic feature extraction and end-to-end learning from data.

Two-Stage Detectors Two-stage detectors first generate region proposals, which are then classified and refined:

- **R-CNN Family:** Introduced region-based detection, but suffered from slow inference and complex training.
- **SPPNet and Fast R-CNN:** Improved the speed and accuracy of R-CNN using spatial pyramid pooling and streamlined processing.
- **Faster R-CNN:** Added a Region Proposal Network (RPN), significantly boosting detection speed and accuracy.
- **Mask R-CNN:** Extended Faster R-CNN with instance segmentation capabilities, though not suitable for real-time applications.

Single-Stage Detectors Single-stage detectors approach object detection as a regression task, offering better inference speed:

- **YOLO Series:** The "You Only Look Once" (YOLO) family pioneered single-stage, real-time object detection by reframing detection as a regression problem. Successive versions introduced improvements in backbone networks (e.g., DarkNet-53), multi-scale detection using FPN, and training strategies such as data augmentation, label smoothing, CIoU loss, and batch normalization to enhance accuracy and speed, especially for small object detection.
- **SSD (Single Shot MultiBox Detector):** SSD performs object detection in a single forward pass by applying convolutional filters at multiple feature scales, allowing real-time performance while maintaining competitive accuracy.
- **RetinaNet:** Known for introducing the Focal Loss function to address class imbalance, RetinaNet achieves high accuracy while preserving the simplicity and efficiency of a single-stage architecture.

Applications in Competitions

- **Modified YOLOv3:** Widely used by teams for real-time UAV detection tasks.
- **Cascade R-CNN with ResNeXt-101 + FPN:** Adopted by winning teams for high-accuracy detection of small aerial targets.
- **Alexis Team:** Combined YOLOv with Spatial Pyramid Pooling (SPP) and synthetic data to enhance training robustness.

3 Methodology

In this work, we focus on the task of UAV detection using two deep learning-based object detection frameworks: Single Shot MultiBox Detector (SSD) and YOLOv11. Both models are selected for their real-time performance and strong generalization capabilities in aerial imagery.

3.1 Single Shot MultiBox Detector (SSD)

The SSD model is a one-stage detector that discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. Unlike two-stage detectors that rely on region proposal steps, SSD performs object classification and bounding box regression directly on feature maps, making it highly efficient. In our implementation, we use SSD300, which utilizes a VGG16 backbone followed by multiple convolutional feature layers to predict objects at multiple scales. SSD achieves a good balance between accuracy and speed, making it suitable for embedded UAV detection scenarios.

3.2 YOLOv11

YOLOv11 is an improved version of the YOLO (You Only Look Once) family, designed to enhance detection accuracy while maintaining real-time inference speeds. It incorporates architectural refinements such as expanded convolutional layers, improved neck designs, and optimized loss functions to better capture small object features—critical in UAV detection. YOLOv11 processes the input image in a single forward pass, making predictions for bounding boxes and class probabilities simultaneously. In our setup, YOLOv11 has been fine-tuned on a custom UAV dataset to improve precision and recall in complex aerial environments.

By leveraging both SSD and YOLOv11, our methodology aims to evaluate the trade-offs between detection speed and accuracy under varying object sizes and flight conditions. The complementary strengths of both models provide robust performance for real-world UAV monitoring applications.

References

1. Jie Zhao, Jingshu Zhang, Dongdong Li, and Dong Wan, "Vision-based Anti-UAV Detection and Tracking."
2. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, 2015.
3. Fardad Dadboud, Vaibhav Patel, Varun Mehta, and Miodrag Bolic, "Single-Stage UAV Detection and Classification with YOLOv5: Mosaic Data Augmentation and PANet," Computational Analysis and Acceleration Research Group (CARG), University of Ottawa.
4. 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 (ECCV)*, 2016.
5. Bing Cao, Haiyu Yao, Pengfei Zhu, and Qinghua Hu, "Visible and Clear: Finding Tiny Objects in Difference Map."
6. Rahima Khanam and Muhammad Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements."
7. Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun, "YOLOX: Exceeding YOLO Series in 2021," arXiv preprint arXiv:2107.08430.
8. Bailin Liu and Huan Luo, "An Improved YOLOv5 for Multi-Rotor UAV Detection," 2023.