Image processing for detection and geolocation of objects using UAV

M Yelis*, M Ospanova

Satbayev University, Kazakhstan, Almaty, Satpayev str., 22A

**Corresponding author's e-mail: k.marina92@gmail.com*

Abstract

Unmanned aerial vehicles (UAVs) have become popular in recent years in a huge number of applications, especially for object detection, localization and tracking of objects in some areas. However, these tasks can be difficult to solve for small objects or complex scenes. In this work, we aim to analyse and generate an algorithm for post-object detection and post-GPS localization of the detected object.

Keywords: unmanned aerial vehicles, object detection, object classification, computer vision, machine learning, convolutional neural networks.

1 General

The successful use of UAVs together with image processing algorithms lead to the expansion of UAV application areas. The use of UAVs to locate an object, determine its location and track movements has long attracted the interests of various scientists. But the most projects focus on motion detection and visual tracking [1]. In this paper, we will use machine learning algorithms for scenes from UAVs to detect objects. The development of computer vision systems leads to the appearance of better algorithms with open and big datasets by training and classifying the data with the help of machine learning approaches.

In our research, we will use CNN for target detection from aerial images using UAV. Then GPS coordinates of the detected object have to be determined. The post-processing module automatically will load that image including the necessary information for that image. The next step it will calculate and provide the GPS coordinates of the target on the image. We will focus on enabling UAVs to continuously collect geo-referenced data and detect objects.

Almost all CNN approaches remodel classifiers to perform detection, applying them in different locations and scales. Each individual component of those pipelines is trained separately, that makes them slow to compute and hard to implement.

One of the popular algorithms, for now, is YOLO. An architecture of YOLO is presented by a single neural network made of 24 convolutional layers and 2 fully connected layers. The core of the idea is that You Only Look Once to predict what objects are present and where they are [2]. That means object detection is considered as a single regression problem. Prediction of bounding boxes and class probabilities are made directly from full images in one evaluation.

Yolo has some limitations, such as: each grid cell predicts only 2 boxes and can only have 1 class; there are still problems with small objects detection and incorrect localizations. Despite this, Yolo outperforms other methods in learning very generalizable representation of objects, so it could be applied in a wide domain: from natural images to artwork.

References

[1] Kadnichanskiy S A 2016 Jingxuan Sun, Boyang Li, Yifan Jiang and Chih-yung Wen *Sensors* **16**(11) 1778

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi 2016 *You Only Look Once: Unified, Real-Time Object Detection*