Get started   Open in app

Christos Kyrkou

Follow   179 Followers   About

Training a deep-learning classifier for aerial top view detection of vehicles

Christos Kyrkou   Mar 27, 2018 · 5 min read

> *Deep learning approaches have demonstrated state-of-the-art performance in various computer vision tasks such as object detection and recognition. In this post I provide details on how to develop and train a Convolutional Neural Network (CNN) to detect top-view vehicles from UAV footage.*

Introduction

Unmanned Aerial Vehicles (drones) are emerging as a promising technology for both environmental and infrastructure monitoring, with broad use in a plethora of applications. In particular, Road Traffic Monitoring (RTM) constitutes a domain where the use of UAVs is receiving significant interest. Under the above deployments, UAVs are responsible for searching, collecting and sending, in real time, vehicle information for traffic regulation purposes from on-board camera sensors. For this purpose a deep convolutional neural network (CNN) is developed that can detect vehicles in images and an appropriate detection algorithm is formed. The algorithm is based on the sliding window approach for which more details can be found here.

Convolutional Neural Network Architecture

The network architecture is inspired from the VGG16 paradigm of sequential blocks of 3x3 convolutional layers followed by max-pooling, and has a total of 960,000

parameters which are much less than pre-trained models such as VGG16. Different experiments were performed to find the appropriate network depth and image size. Initial experiments were done using 32x32 pixel images akin to the cifar-10 benchmarks, however, due to the small amount of detail provided by top-view vehicle images the network did not perform as well as it was expected. Increasing the image resolution to 50x50 and capturing finer details reduced the amount of false positives and thus resulted in a higher accuracy network.
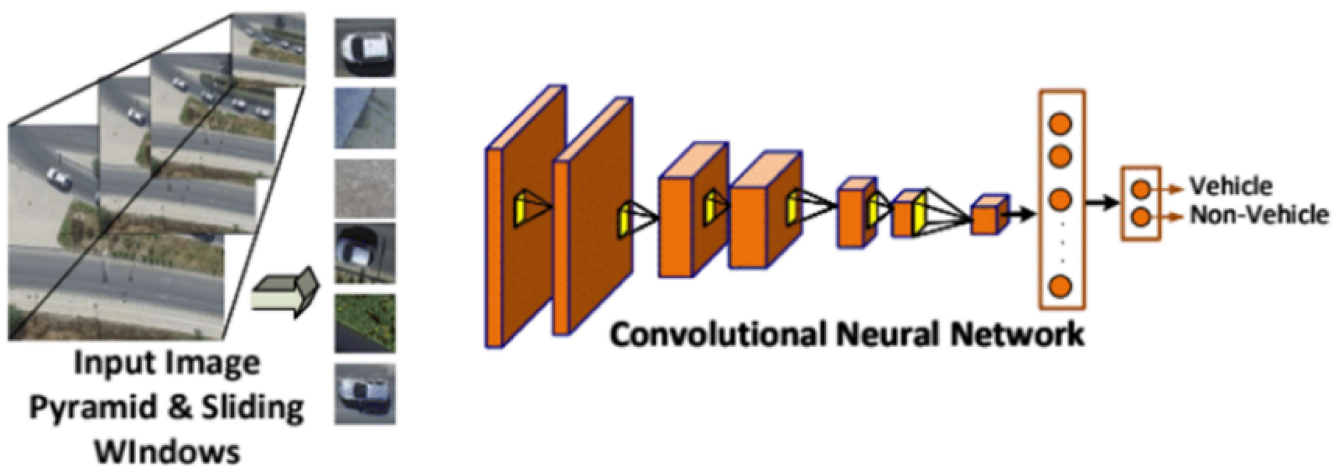

Overall Detection Process with Convolutional Neural Network

Training

For the training of the CNN classifier appropriate image data were collected. In particular, the size of the images to train the neural network is chosen to be of 50x50 pixel resolution to trade-off between the visual detais of the vehicle and the computational requirements of the network. In total 9000 vehicle images were cropped and extracted from aerial image footage captured using a DJI Matrice 100 UAV.
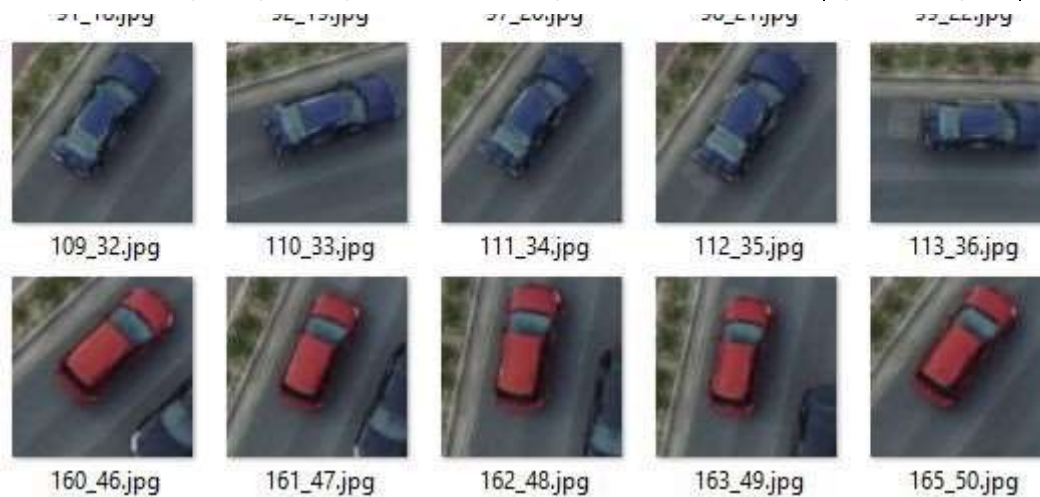
Example of images used for training the vehicle detector

Cropping positive image samples (vehicles) from video footage using motion detection in OpenCV

In particular to accelerate the data collection process a motion detection algorithms is used to extract all moving vehicles by appropriately cropping the moving objects and filtering based on the expected size of the vehicle. This was possible since the UAV was stationary above a road segment and where majority of moving objects where vehicles. To form the negative example dataset image patches of predetermined size where automatically extracted, using OpenCV from UAV footage that did not contain vehicles.

Later the hard negative mining process was used to iteratively retrain the network with patches that were erroneously classified as vehicles. Furthermore, to enhance the training set the KERAS image generator was used to augment the data with random translations, rotations, and mirroring, effectively increasing the dataset to a total of 95,000 positive and 245,000 negative examples.

```python
datagen = ImageDataGenerator(rotation_range = 180,
                featurewise_center = False,
                featurewise_std_normalization=False,
                width_shift_range=0.1,
                height_shift_range=0.1,
                zoom_range = [0.9, 1.1], # less than 1 is zoom in, larger than one is zoom out
                shear_range=0.0,
                fill_mode = 'wrap',
                horizontal_flip=True,
                vertical_flip=True,
                rescale=None,
                cval=0)
```

Example of how the KERAS image data generator was used

In addition to the classifier, non-maximum suppression is also applied to remove multiple detections, also a heat-map strategy is applied to filter out some low confidence detections and noise. The heat map is composed by taking the detections over 3 consecutive frames and adding 1s to those locations in the image where a vehicle is detected. In practice, what is achieved with a heat map over several frames of video, is that areas of multiple detections get "hot", while transient false positives stay "cool". The detections are kepted only for the regions where the summation exceeds a threshold of 2. That is if a patch is classified as a vehicle only once in the consecutive frames then it is discarded. In this way we only keep the high confidence detections and remove false positives.

Performance

The final performance of the classifier was tested on various set of full sized images. The vehicle detector is able to reliably recognize cars in various conditions as is seen in the images below. The most important aspect that helped in getting the best performance from the network was the augmentation strategy that resulted to both more positive samples and increased variability in the vehicle appearance.

Detection using a heat-map on an image frame

Conclusions

This post has detailed how to apply a sliding window approach with a deep learning classifier (convolutional neural network) to detect top-view vehicles in images and can be applied to UAV applications such as traffic monitoring or search and rescue missions.This particular approach is based on the sliding-window paradigm, in which a classifier is applied on a dense image grid. For this reason it is computationally very demanding and hardware accelerators need to be employed to speedup the whole process. Later blog posts will show how recent techniques such as SSD and YOLO can improve on the sliding window technique and apply a convolutional neural network once over the whole image and detect all objects.

Detections on a 3000×3000 resolution image (Images taken from
http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html)

Video Footage

Deep Convolutional Neural Network Classifier for Car Detection from a UAV

Relevant Papers

[1] **Christos Kyrkou**, Stelios Timotheou, Panayiotis Kolios, Theocharis Theocharides, Christos Panayiotou, "Optimized Vision-Directed Deployment of UAVs for Rapid Traffic Monitoring", accepted to appear in proceedings of International Conference on Consumer Electronics (ICCE), January, 2018.

Deep Learning   Drones   Object Detection   Aerial Imaging   Computer Vision

About   Help   Legal

Get the Medium app