Technische Hochschule Brandenburg
University of Applied Sciences
Fachbereich Informatik und Medien

Multi-staged Deep Learning approach for automatic counting and detecting banana trees in UAV images using Convolutional Neural Networks

Herval Bernice Nganya Nana
Master Thesis • Computer Science Program • Department of Computer Science and Media • 10.09.2018

Background and aim

Agricultural production forecasting has always been a big challenge for farmers and agribusiness industries, leading to a climate of uncertainty in the sector. For many decades, researchers have been looking for ways to improve the accuracy of agricultural production forecasting. This research aims to develop a multi-staged deep learning approach for automatic counting and detecting of banana trees in UAV images using convolutional Neural Networks.

Objectives

To fully achieve the aim of this study, a couple of objectives were outlined on the target system. The target system has to perform an accuracy on the counting task of at least 90%. In case of doubt, an **underestimation** is more important than an overestimation. Regarding the localization task, the system should be as accurate as possible. It should process a 4000×4000 pixels UAV image in less than 12 minutes. Assuming that a 40×40 px square in a UAV image corresponds to a square of 2.5×2.5 m, an image of size 4000×4000 px is exactly 6.25 ha.

Method

Given an n×m pixel image of a banana plantation, the sliding window technique was first applied in order to scan the image. This was applied with a window size of 40×40 px and a stride of 10px. Then, a classifier was applied on each of these windows to determine whether it contained at least one tree crown (class 1) or not (class 0). A confidence *c* was also returned for each window. After that, a regression model was applied on each class 1 labelled window in order to predict the coordinates *(x,y)* of the best visible tree crown (candidate) in a given window. The windows classified as 0 were consequently ignored. Finally, the candidates were aggregated using NMS, so that the best performance on the counting task at first and then on the localization task. See figure 1.

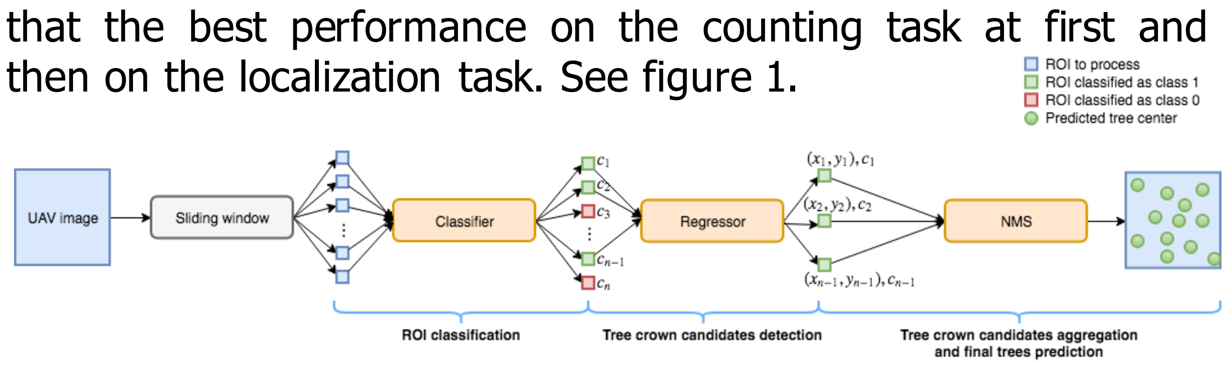

Fig. 1: Workflow of the multi-staged method

Data collection

To build the dataset, volunteers annotated as many templates as possible using a web application developed in the context of this research. During the annotation process, if there was at least one tree crown on a given template, volunteers must have selected the best visible one with high precision. In addition to the outcoming data, 840 marked templates from Christoph Gresch's Master Thesis of [Gre07] were provided at the beginning of this study. Figure 2 shows examples of templates. Figure 3 shows the class distribution in the dataset.

(a) (b)

Fig. 2: Template examples (a) No tree crown (class 0), (b) Tree crown (class 1).

Fig. 3: Class distribution in the training set (blue) and in the test set (green). Training set (85%), test set (15%). 39780 marked templates gathered after using data augmentation

Implementation

This step was carried out using Anaconda version 1.6.9 with Python 3.6.4 on a computer running Ubuntu 16.04 LTS 64 Bit. In addition, an NVIDIA GeForce GPU with 940M and 2GB dedicated VRAM was used. An Intel Core i7-5500U CPU with 2.5GHz and Turbo Boost up to 3.0GHz was used too. The RAM's capacity of this computer was 8GB DDR3 L Memory.

-Sliding window: This method was used in order to scan the input image. A stride of 10px and a window size of 40×40px were applied.

-Window Classification: A CNN [Lec89] model which achieved an accuracy of 85.02% and an F1-Score of 0.89 was trained. Figure 4 shows the classifier's architecture.

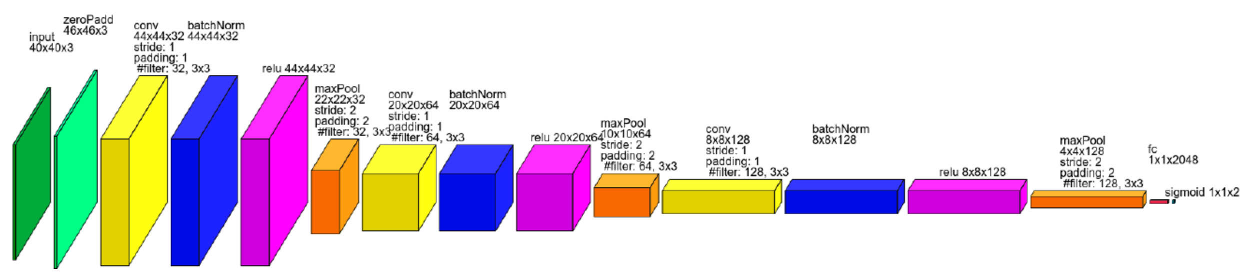

Fig. 4: Classifier architecture

-Candidates prediction (regression): A regressor achieving a MAE of 0.1397 was reached in this study. Its architecture was similar to the classifier's architecture except from the last layer which contained 2 units instead of 1.

Experiment (Candidates aggregation)

Three test images were created. The tree crowns on the image were manually marked. While the test images A and B (optimization images) were used to find the best parameters of the aggregation, test image C (850×296px) was used to evaluate the final model. The Non-Maximum-Suppression (NMS) method was used here to aggregate the candidates. The NMS parameters were found by selecting the best parameter set from multiple parameter sets generated using a brute force approach.

Results and conclusion

On the counting task, a MOE of 0.0821 was achieved, which corresponds to an accuracy of 91.79% on the test case. The final model performed an Average Distance Error of 43cm on this test case and could process the test image in 8 seconds.

References

[Gre07] Gresch, Christoph: *Einsatz von Data Mining zur Identifikation und Schätzung der Anzahl von Bananenpflanzen in einem Bild.* THB, Master Thesis, 2007.

[Lec89] Lecun, Yann: *Generalization and network design strategies.* Elsevier, 1989.

Supervisors: Prof. Dr.-Ing. Sven Buchholz • Technische Hochschule Brandenburg • Dipl.- Inf. Ingo Boersch • Technische Hochschule Brandenburg • Jan Vogt • Orca Geo Services