TU WIEN Informatics

Flower State Classification for Watering System

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Software Engineering & Internet Computing

eingereicht von

Tobias Eidelpes, BSc
Matrikelnummer 01527193

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.-Prof. Dr. Horst Eidenberger

Wien, 20. Februar 2023

Tobias Eidelpes	Horst Eidenberger

TU WIEN Informatics

Flower State Classification for Watering System

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Software Engineering & Internet Computing

by

Tobias Eidelpes, BSc
Registration Number 01527193

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.-Prof. Dr. Horst Eidenberger

Vienna, 20th February, 2023

Tobias Eidelpes   Horst Eidenberger

Erklärung zur Verfassung der Arbeit

Tobias Eidelpes, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 20. Februar 2023

Tobias Eidelpes

Danksagung

Ihr Text hier.

Acknowledgements

Enter your text here.

Kurzfassung

Ihr Text hier.

Abstract

Enter your text here.

Contents

CHAPTER 1

Evaluation

The following sections contain a detailed evaluation of the model in various scenarios. First, we present metrics from the training phases of the constituent models. Second, we employ methods from the field of Explainable Artificial Intelligence (XAI) such as Local Interpretable Model Agnostic Explanation (LIME) and Gradient-weighted Class Activation Mapping (Grad-CAM) to get a better understanding of the models' abstractions. Finally, we turn to the models' aggregate performance on the test set and discuss whether the initial goals set by the problem description have been met or not.

1.1 Object Detection

The object detection model was pre-trained on the COCO [LMB+] dataset and fine-tuned with data from the Open Images Dataset (OID) [KRA+] in its sixth version. Since the full OID dataset contains considerably more classes and samples than would be feasibly trainable on a small cluster of GPUs, only images from the two classes *Plant* and *Houseplant* have been downloaded. The samples from the Houseplant class are merged into the Plant class because the distinction between the two is not necessary for our model. Furthermore, the OID contains not only bounding box annotations for object detection tasks, but also instance segmentations, classification labels and more. These are not needed for our purposes and are omitted as well. In total, the dataset consists of 91479 images with a roughly 85/5/10 split for training, validation and testing, respectively.

1.1.1 Training Phase

The object detection model was trained for 300 epochs on 79204 images with 284130 ground truth labels. The weights from the best-performing epoch were saved. The model's fitness for each epoch is calculated as the weighted average of mAP@0.5 and mAP@0.5:0.95:

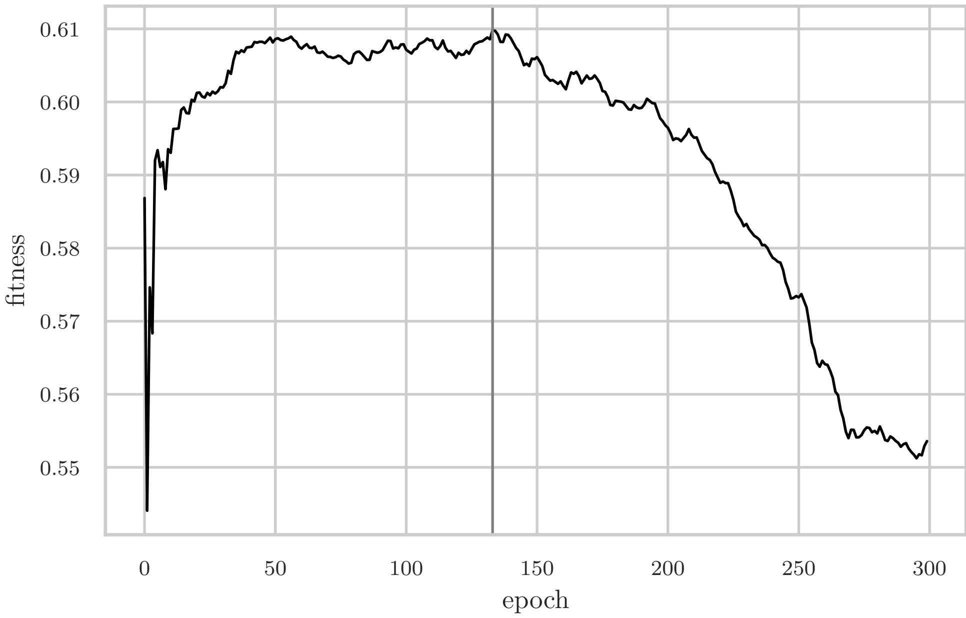

Figure 1.1: Model fitness for each epoch calculated as in equation 1.1. The vertical gray line at 133 marks the epoch with the highest fitness.

$$f_{epoch} = 0.1 \cdot \mathsf{mAP@}0.5 + 0.9 \cdot \mathsf{mAP@}0.5{:}0.95 \tag{1.1}$$

Figure 1.1 shows the model's fitness over the training period of 300 epochs. The gray vertical line indicates the maximum fitness of 0.61 at epoch 133. The weights of that epoch were frozen to be the final model parameters. Since the fitness metric assigns the mAP at the higher range the overwhelming weight, the mAP@0.5 starts to decrease after epoch 30, but the mAP@0.5:0.95 picks up the slack until the maximum fitness at epoch 133. This is an indication that the model achieves good performance early on and continues to gain higher confidence values until performance deteriorates due to overfitting.

Overall precision and recall per epoch are shown in figure 1.2. The values indicate that neither precision nor recall change materially during training. In fact, precision starts to decrease from the beginning, while recall experiences a barely noticeable increase. Taken together with the box and object loss from figure 1.3, we speculate that the pre-trained model already generalizes well to plant detection because one of the categories in the COCO [LMB+] dataset is *potted plant*. Any further training solely impacts the confidence of detection, but does not lead to higher detection rates. This conclusion is supported by the increasing mAP@0.5:0.95 until epoch 133.

Further culprits for the flat precision and recall values may be found in bad ground truth data. The labels from the OID are sometimes not fine-grained enough. Images

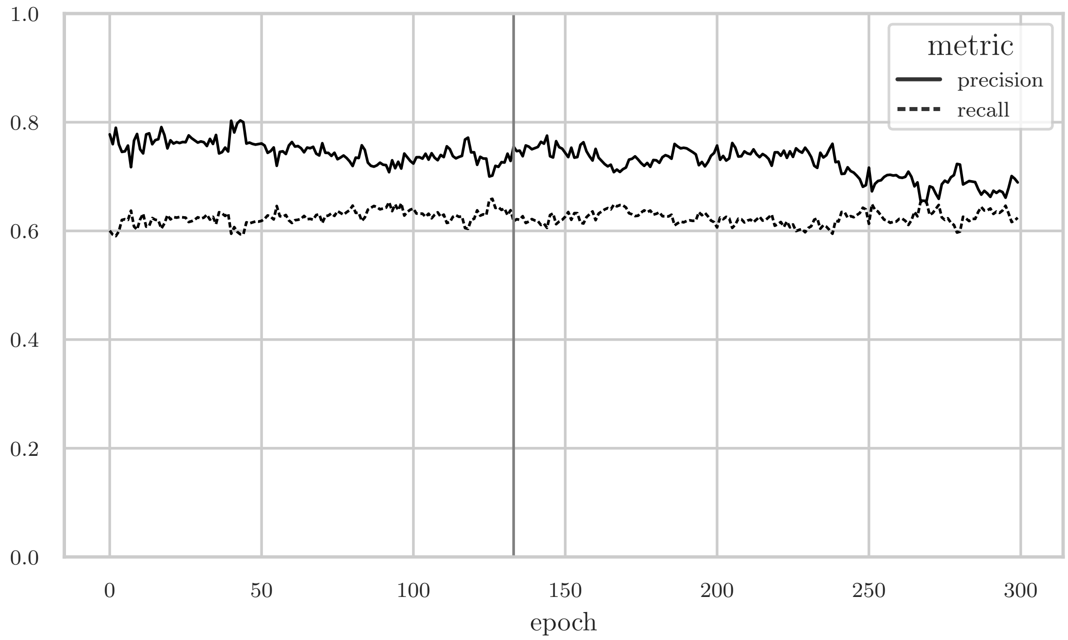

Figure 1.2: Overall precision and recall during training for each epoch. The vertical gray line at 133 marks the epoch with the highest fitness.

which contain multiple individual—often overlapping—plants are labeled with one large bounding box instead of multiple smaller ones. The model recognizes the individual plants and returns tighter bounding boxes even if that is not what is specified in the ground truth. Therefore, it is prudent to limit the training phase to relatively few epochs in order to not penalize the more accurate detections of the model. The smaller bounding boxes make more sense considering the fact that the cutout is passed to the classifier in a later stage. Smaller bounding boxes help the classifier to only focus on one plant at a time and to not get distracted by multiple plants in potentially different stages of wilting.

The box loss decreases slightly during training which indicates that the bounding boxes become tighter around objects of interest. With increasing training time, however, the object loss increases, indicating that less and less plants are present in the predicted bounding boxes. It is likely that overfitting is a cause for the increasing object loss from epoch 40 onward. Since the best weights as measured by fitness are found at epoch 133 and the object loss accelerates from that point, epoch 133 is probably the correct cutoff before overfitting occurs.

1.1.2 Test Phase

Of the 91479 images around 10% were used for the test phase. These images contain a total of 12238 ground truth labels. Table 1.1 shows precision, recall and the harmonic mean of both (F1-score). The results indicate that the model errs on the side of sensitivity

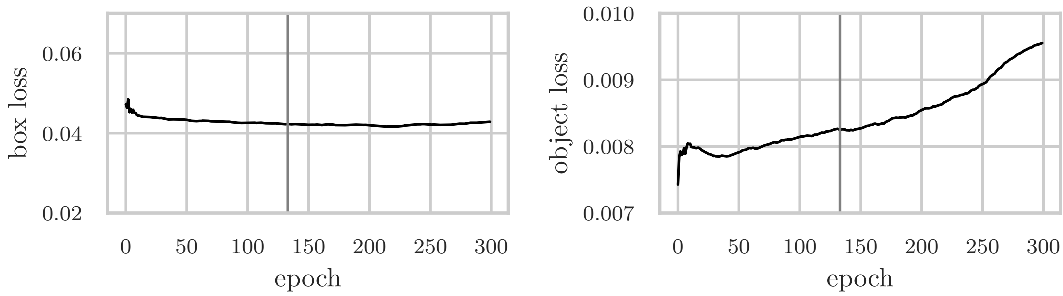

Figure 1.3: Box and object loss measured against the validation set of 3091 images and 4092 ground truth labels. The class loss is omitted because there is only one class in the dataset and the loss is therefore always zero.

because recall is higher than precision. Although some detections are not labeled as plants in the dataset, if there is a labeled plant in the ground truth data, the chance is high that it will be detected. This behavior is in line with how the model's detections are handled in practice. The detections are drawn on the original image and the user is able to check the bounding boxes visually. If there are wrong detections, the user can ignore them and focus on the relevant ones instead. A higher recall will thus serve the user's needs better than a high precision.

	Precision	Recall	F1-score	Support
Plant	0.547571	0.737866	0.628633	12238.0

Table 1.1: Precision, recall and F1-score for the object detection model.

Figure 1.4 shows the Average Precision (AP) for the Intersection over Union (IOU) thresholds of 0.5 and 0.95. Predicted bounding boxes with an IOU of less than 0.5 are not taken into account for the precision and recall values of table 1.1. COCO's [LMB+] main evaluation metric is the AP averaged across the IOU thresholds from 0.5 to 0.95 in 0.05 steps. This value is then averaged across all classes and called mean average precision (mAP). The object detection model achieves a state-of-the-art mAP of 0.5727 for the *Plant* class.

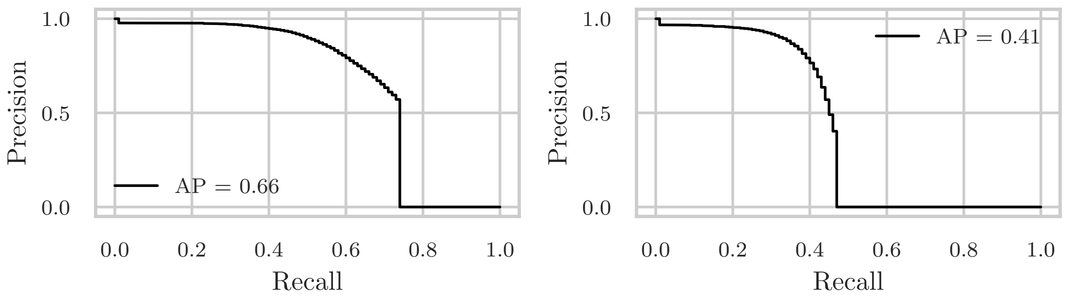

Figure 1.4: Precision-recall curves for IOU thresholds of 0.5 and 0.95. The AP of a specific threshold is defined as the area under the precision-recall curve of that threshold. The mAP across IOU thresholds from 0.5 to 0.95 in 0.05 steps mAP@0.5:0.95 is 0.5727.

List of Figures

List of Tables

List of Algorithms

Bibliography

[KRA+] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerig, and Vittorio Ferrari. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. 128(7):1956–1981.

[LMB+] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context.