TECHNISCHE UNIVERSITÄT WIEN

Institut für Computertechnik
Institute of Computer Technology

A MASTER THESIS ON

Optimizing deep neural networks for efficient dronebased ragweed detection

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Diplom-Ingenieur
(Equivalent to Master of Science)

in

Data Science E 066 645

by

BSc. Lukas Steindl

11743494

Supervisor(s):
Univ.Prof. Dipl.-Ing. Dr.techn. Axel Jantsch
Dipl.-Ing. Dr.techn. Alexander Wendt

Vienna, Austria
May 2021

Abstract

Ambrosia artemisiifolia also know as ragweed is an adventive weed species that aggressively spreads across Europe. 4-5 percent of the population suffers from strong allergic reactions to its pollen. Richter et al. [1] simulated the dispersion of ragweed in Austria and Bavaria for different climate scenarios and management regimes and found that taking no counteractions would result in a rise of mean annual allergy related cost from 133 Mio. EUR in 2005 to 422 Mio. EUR in 2050. For the same region they calculated that an investment of 15 Mio. EUR/year in traditional weed management would reduce the mean allergy related cost by 82 Mio. EUR per year. The effort for removing ragweed splits into the cost for detection (10%) and the cost for subsequent eradication (90%) and is estimated to be 8570 EUR per km^2 and year in total.

This work proposes a novel method that uses aerial drones equipped with highly optimized deep neural networks (DNNs) to scan large areas of vegetation with high speed and precision. Compared to the traditional manual approach with estimated survey-cost of roughly 860 EUR/km^2 the proposed automated technique has the potential to cut the cost for ragweed-detection and monitoring by orders of magnitudes. The challenge is that the range of aerial drones is limited by their battery capacity and energy consumption. In addition to that it might not be possible to deploy DNN Models with millions of parameters to embedded devices due to memory constraints. Offloading the workload into data centers introduces new issues as the drones may operate in rural areas with insufficient network connectivity and bandwidth. The problem that this thesis tries to solve is to find a proper state of the art DNN for recognizing ragweed from a drone camera and to analyze how this DNN can be compressed to be applicable on embedded hardware. An optimized system could provide a cost-effective solution to support the combat against the spread of Ambrosia artemisiifolia and alleviate the medical conditions experienced by millions of sensitized people in Europe.

Kurzfassung

Ambrosia artemisiifolia auch als Ragweed bekannt ist eine invasive Pflanzenart die sich schnell in Europa ausbreitet. 4-5 Prozent der Bevölkerung reagieren allergisch auf ihre Pollen. Richter u.a haben die weitere Ausbreitung von Ragweed in Österreich und Bayern unter der Annahme verschiedener Klimaentwicklungen und Aktionspläne untersucht und herausgefunden, dass mit einem Anstieg der durchschnittlichen jährlichen Allergiekosten von 133 Mio. im Jahr 2015 auf 422 Mio. EUR im Jahr 2050 zu rechnen ist wenn keinerlei Gegenmaßnahmen ergriffen werden. Für dieselbe Region wurde berechnet, dass bereits eine Investition von 15 Mio. EUR / Jahr die durchschnittlichen jährlichen Allergiekosten um 82 Mio. EUR reduzieren würde. Die Kosten der Ragweedentfernung teilen sich auf in 10% für die Erkennung und 90% für die tatsächliche Beseitigung und werden insgesamt auf 8570 EUR pro km^2 und Jahr geschätzt.

Diese Arbeit befasst sich mit der Frage, ob es mit dem heutigen Stand der Technik möglich ist, mittels Flugdrohne und eingebetteter künstlicher neuronaler Netzwerke größere Vegetationsflächen mit hoher Geschwindigkeit und Präzision auf Ragweed zu untersuchen. Im Vergleich zu traditionellen Methoden, könnte diese Technik die Kosten der Ragweed-Erkennung von derzeit rund 860 EUR/km^2 um Größenordnungen reduzieren. Da sich Ragweed in Regionen mit unzureichender Breitbandinfrastruktur ausbreitet, ist eine Verlagerung der Bilddatenverarbeitung in Rechenzentren häufig nicht praktikabel. Die Herausforderung besteht somit darin, eine möglichst optimale Kombination aus eingebetteter spezialisierter Inferenzhardware und künstlichen neuronalem Netzwerk zu finden, sodass eine größtmögliche Vegetationsfläche für ein gegebens Qualitätsniveau und Energiebudget untersucht werde kann. Ein optimiertes System würde eine kosteneffiziente Lösung im Kampf gegen die Ausbreitung von Ambrosia artemisiifolia darstellen und könnte helfen, die Beschwerden von Millionen allergiegeplagter Menschen in Europa lindern.

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder in ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Copyright Statement

I, BSc. Lukas Steindl, hereby declare that this thesis is my own original work and, to the best of my knowledge and belief, it does not:

- Breach copyright or other intellectual property rights of a third party.
- Contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
- Contain material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.
- Contain substantial portions of third party copyright material, including but not limited to charts, diagrams, graphs, photographs or maps, or in instances where it does, I have obtained permission to use such material and allow it to be made accessible worldwide via the Internet.

Signature: _____

Vienna, Austria, May 2021 BSc. Lukas Steindl

Contents

Chapter 1

Introduction

Over the last 20 years the rise of Ambrosia Artemisiifolia extended the hayfever season for millions of senitized people in Europe from early summer to late autumn. The plant originally native to north america came to our continent as a result of upcoming global trade in the 19th century.

Figure 1.1: Ambrosia artemisiifolia. Mario Lešnik

Initially not spreading too much, several factors such as industrialized agriculture and climate change speeded up the infestation in europe. [2] Figure 1.2 shows the rapid distribution of this alien species across Europe.

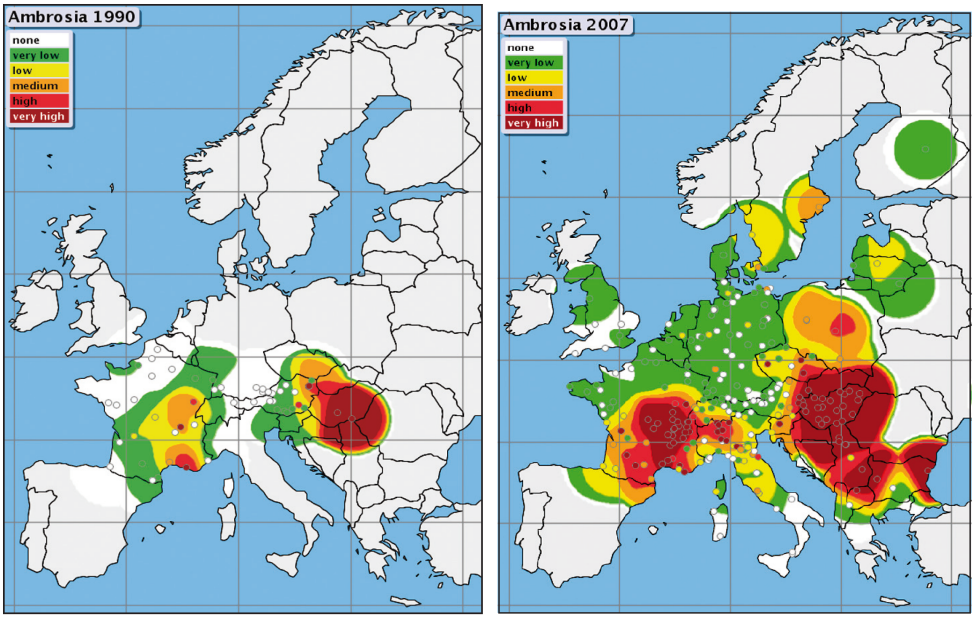

Figure 1.2: Ragweed pollen concentration 1990 (left) and 2007 (right). Favored by climate change Ambrosia Artemisiifolia spreads out roughly 25km from east to west every year. Data provided by the European Aeroallergen Network. [3]

Richter et al. simulated the dispersion of ragweed in Austria and Bavaria for different climate scenarios and management regimes and found that taking no counteractions would result in a rise of mean annual allergy related cost from 133 Mio. EUR in 2005 to 422 Mio. EUR in 2050. For the same region they calculated that an investment of 15 Mio. EUR/year in traditional weed management would reduce the mean allergy related cost by 82 Mio. EUR per year.

The effort for removing ragweed splits into the cost for detection (10%) and the cost for subsequent eradication (90%) and is estimated to be 8570 EUR per km^2 in total. One way to lower the detection cost is to crowd-source the task. Federal state governments implemented laws to make it obligatory to report find-spots and demand landowners and public entities to remove the plant upon discovery. Citizens are asked to take pictures of ragweed with their mobile phones and send them to the authorities for verification and removal coordination. In an interdisciplinary project [4] a ResNet50-Image-Classifier was trained to show that deep neural networks (DNNs) are capable of distinguishing ragweed from similar looking plants in close-up images from mobile phones.

This thesis extends the idea of automated detection and proposes a novel method that uses aerial drones equipped with highly optimized DNNs to scan large areas of vegetation with high speed and precision. Compared to the traditional manual approach with estimated survey-cost of roughly 860 EUR/km^2 the proposed automated technique has the potential to cut the cost for ragweed-detection and monitoring by orders of magnitudes.

The challenge is that the range of aerial drones is limited by their battery capacity and energy con-

sumption. It also might not be possible to deploy DNN Models with millions of parameters such as the ResNet50 to embedded devices due to memory constraints. Offloading the workload into data centers introduces new issues as the drones may operate in rural areas with insufficient network connectivity and bandwidth.

The problem that this work tries to solve is to find a proper state of the art DNN for recognizing ragweed from a drone camera and to analyze how this DNN can be compressed to be applicable on embedded hardware. An optimized system could provide a cost-effective solution to support the combat against the spread of Ambrosia artemisiifolia and to alleviate the medical conditions experienced by millions of people in Europe.

Chapter 2

State of the Art

Plascak et al. [5] used helicopters and drones to collect RGB and NGB images flying at high altitudes. They created orthographic maps and manually engineered features such as tone, color, texture, pattern, form, size, height/altitude, and location to interpret the images. They highlight that the spacial distribution of ragweed can be estimated well by a strong green leaf color in early summer as the weed consumes more water than other plants nearby.

Figure 2.1: DJI Phantom Drone with NGB Camera.

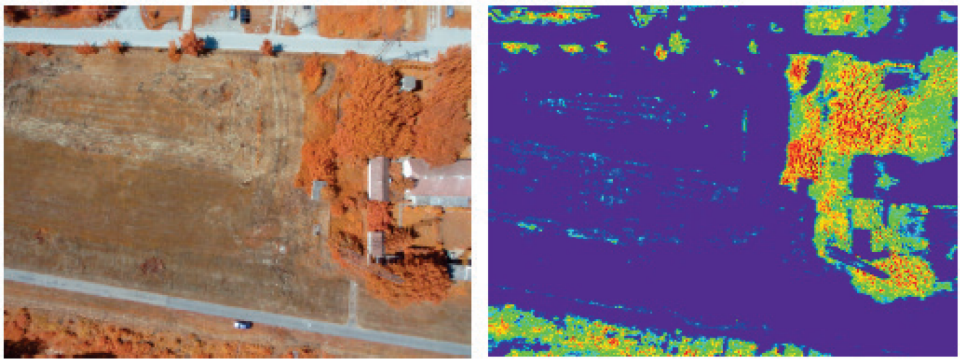

Figure 2.2: Remote Sensing and classification with traditional computer vision methods from high altitude platform.

Budda et al. [6] suggested a combination of robotic sprayer and drone system as a solution to automatically eradicate different kinds of weeds. They trained an object detector to recognize different weed types in an early stage to apply small amounts of pesticides.

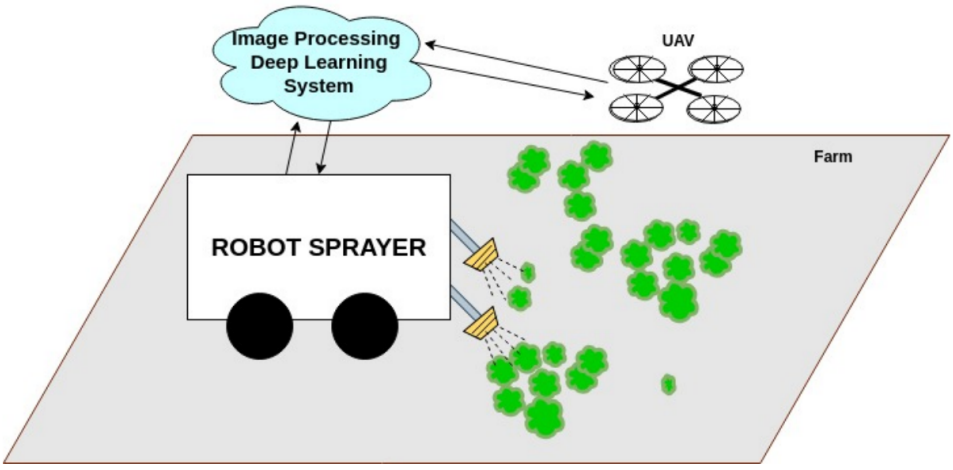

Figure 2.3: Drone works in tandem with ground robot.

They emphasize the importance of flying at high altitudes to improve efficiency. Increasing the distance between camera and surface reduces the number of pixels per plant and degrades recognition quality. To mitigate this problem they propose to upsample images using generative adversarial networks (GANs) as a first step and then passing the output of the GAN as input to the object detector.

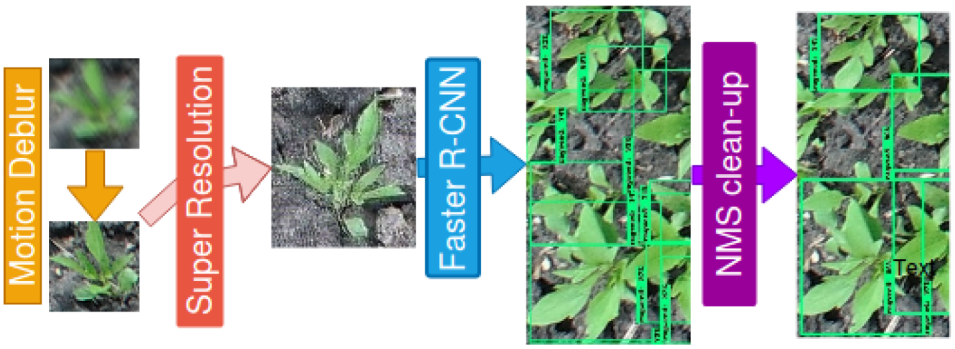

Figure 2.4: Upsamling the images with GANs before feeding the images into the object detector.

The authors reported a 93.8 % classification accuracy for the detected instances of the three weed classes that they consider in their model. However 50 out of 170 Ragweed examples were not detected by the Fast R-CNN Object Detector. Their solution also requires offloading the workload to the cloud. This limits the application to areas with sufficient network connectivity and bandwidth. They also work with object detectors only. When looking at dense vegetation it becomes difficult to properly draw boundingboxes around the plants of interest only. The experiments described in chapter 3 showed high rates of false positives using object detectors when applied to dense vegetation.

Chapter 3

Design Space

This chapter presents a theoretical model of the problem. For a given level of battery capacity the system should scan as much vegetation area $A = d * a$ as possible. The variable d denotes the distance a drone flies in horizontal direction and a is the surface-area scanned at each position when looking down in a 90 degree incidence angle from the platform. Figure 3.1 illustrates the idea.

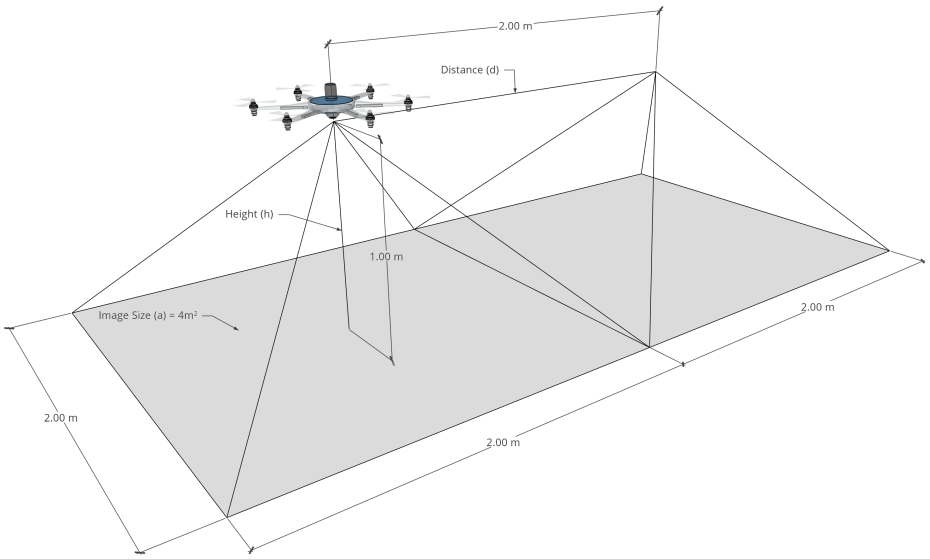

Figure 3.1: The drone flies at altitude h and makes an image in position d that covers area a.

Quantitatively the distance d can also be seen as the number of positions from which the drone can take a non-overlapping image. This number is limited by the battery-capacity b the energy consumption of the inference model $e(s)$, the energy consumption of the drone c and the velocity v.

$$d = \frac{b}{(e(s) + c)} * v \tag{3.1}$$

Note that the first factor in this equation is just the time t that the drone can fly and perform

inference. The altitude or height h of the drone above ground determines the size of the image $a = 2h^2$ in square meters. The sampling-rate s expresses the number of inferences the systems can perform per position.

To introduce a quality factor one has to assume that the drone is equipped with a camera that provides unlimited optical zoom without loss of image-quality and with no alignment-time. With that the average number of pixels per plant can remain constant when flying at a higher altitude by just increasing the sampling-rate.

We define quality q as the ratio between s and a.

$$q = \frac{s}{a} \tag{3.2}$$

This means the qualityfactor is 1 if the model can process just as many images of size 1 as there are units of area in the observed ground image.

Example: Imagine the drone flies $1m$ above ground then the area observed is $4m$. Assume the model can process 4 images each covering $1m$. Then the qualityfactor will be $\frac{4}{4} = 1$. If the drone flies at $2m$ the observed area will be 16 squaremeters. Now the model has to process 16 samples at a time to provide the same level of quality as before.

With (3.2) the flight altitude h can be expressed as a function of s and q.

$$h = \frac{(\frac{\sqrt{s}}{q})}{2} \tag{3.3}$$

The same idea could be used to quantify the scan-quality related to the velocity of the platform. In this theoretical model we do not allow a loss of quality due to velocity and say that the drone-speed is directly limited by the samplingrate s. So the drone can only move to the next position when every image tile in the current position got processed.

$$v = \sqrt{s} \tag{3.4}$$

Combining all equations above provides the following objective function as a function of s:

$$A = d * a = t * v * a = \frac{b}{(e(s) + c)} * \sqrt{s} * \frac{s}{q} \tag{3.5}$$

So far there is no function $e(s)$ that expresses the cost for faster inference. In reality one cannot infinitely improve the latency of a system. To model this limitation the power consumption for faster inference shall be a quadratic function of s: $e(s) = s^2$.

This function can be solved analytically and gives a maximum for:

$$s = Sqrt(3) * Sqrt(c)/b \tag{3.6}$$

Example: Assuming the drone requires 182.4 Watt per second airtime and the inference system takes 2.65 Watt per image and second with a quadratic increase of power consumption for faster inference, the systems optimal throughput would be 8.8 images per second. $s = Sqrt(3) * Sqrt(182.4)/2.65 = 8.82$ This is equal to an inference latency of 113ms. The drone would fly at a speed of 3 meters per second in an altitude of 1.5 meters above ground. Assuming a battery capacity of 6000mAh it would be possible to scan 11808 m² of vegetation.

These theoretical considerations abstract from the models actual detection quality that depends on the network design. Instead it considers the pixels per plant that relates to the input resolution and the distance between the camera and the plant as a measure of image-recording-quality. The model also abstracts from the weight of the embedded device and many other factors such as aerodynamic drag that would limit the speed of the drone. As the neural network design itself is $NP - Complete$ as shown by Avrim et al. [7] the whole problem is at least $NP - Complete$.

Therefore this work aims to find a heuristic solution that provides a high recognition quality using only a small number of pixels per plant. At the same time the system should be fast and its power consumption and hardware-weight should be low.

Chapter 4

Data Collection and Preparation

In September 2020, roughly 130 minutes of videodata and 200 high-resolution images of ragweed-invested dense vegetation were collected in the eastern part of Austria. Figure 4.2 shows an example image of an uncultivated field in Breitenbrunn am Neusiedler See. Multiple configurations of camera-systems and perspective were tested to obtain image data with the desired level of quality.

Figure 4.1: Image (3840 x 2160) 4k-resolution sampled from video taken with Huawei P20 Pro Mobile Phone from approximately 2.5 meters above ground. The image shows two ragweed plants in full bloom.

A Huawei-P20-Pro-mobile-phone-camera was used to collect the image data. It turned out to be challenging or even impossible to annotate dense vegeation images taken with this camera from 4 meters above ground. A tradeoff between observed area and resolution per plant was found at 2.5

meters above ground. This ground-sampling distance and a camera angle of roughly 45 degrees results in images that allow human annotators to properly identify the ragweed.

4.1 Object Detection Dataset

Inspired by the work of Buddha et al. a first dataset for object detection was annotated. Based on their proposed method the original 4k videos where chopped into 8 non-overlapping video tiles.

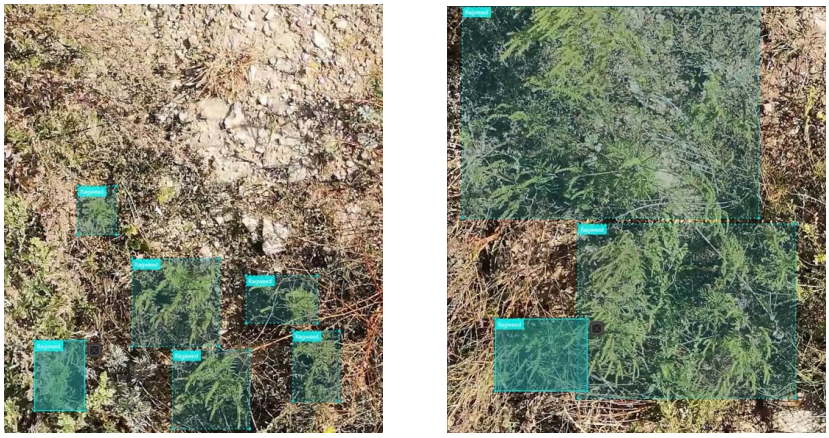

Figure 4.2: Image Tile of size 960x1080 with bounding boxes containing Ragweed

The final Object-Detection-Dataset consist of 971 image-tiles with a resolution of 960x1080 pixels. 481 Tiles show at least one region of interest. The data was split into 80% Trainingset and 20% Testset. As the results obtained by the object detection model showed a large fraction of false positives another dataset for semantic segmentation was created.

4.2 Semantic Segmentation Dataset

Providing semantic segmentation masks is a tedious and time consuming task. Various tools and platforms exist to support the process. Figure 4.3 shows an efficient way to create segmentation masks using a modern touchscreen with stylus pen.

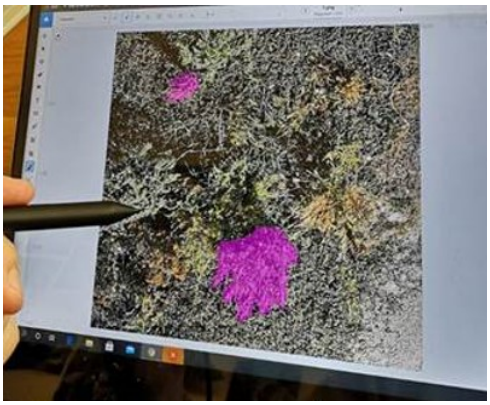

Figure 4.3: Annotating Images using a Touchdisplay with Stylus Pen directly in a modern webapplication [8]).

The segmentation-dataset consist of 103 train, 11 validation and 25 test images. Figure 4.4 gives an impression of the resulting annotation masks that form arbitrary shapes and sizes.

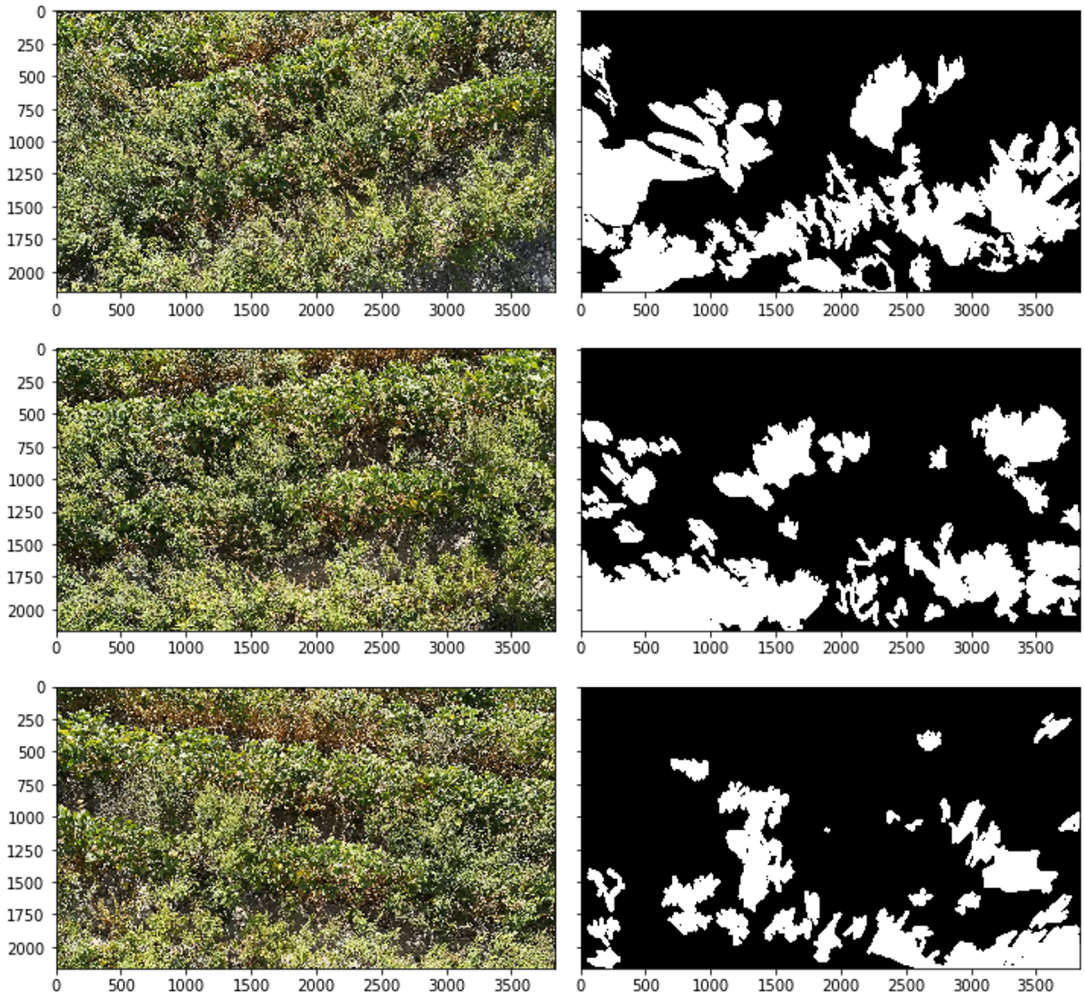

Figure 4.4: Annotating Images using a Touchdisplay with Stylus Pen directly in the browser.

The input images as well as the annotation masks were converted into numpy arrays and serialized

as google Tensorflow protobuffers (tfrecords). This format allows an efficient streaming of data during the training and validation phase.

Chapter 5

Preliminary Results

5.1 Object Detector

As a baseline model a pretrained Single-Shot-Detector with a ResNet50 Backbone and an input resolution of 640x640 was taken from the Tensorflow model zoo [9]. To finetune the weights for the ragweed-dataset a momentum optimizer with a base learning rate of 0.04 was executed for 25000 steps.

On first glance the qualitative results looked promising.

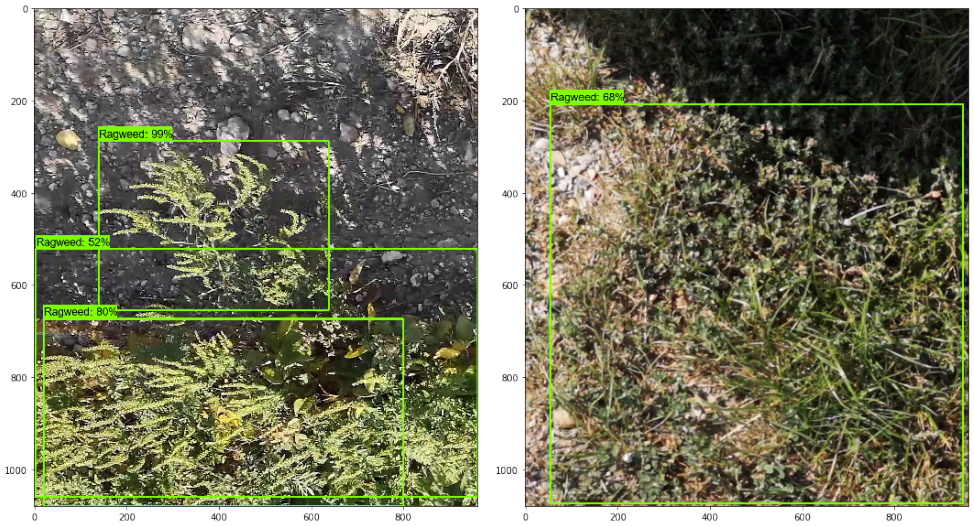

Figure 5.1: First Object Detection results. The model predicts reasonable bounding boxes if the plant is well separated from the rest of the vegetation. However it suffers from poor labeling when vegetation is dense leading to high false positive rates.

However when testing the model on unseen data it did not work as well. While the model memorized the training data really well it did not generalize:

On the testset the following precision and recall values were obtained:

Figure 5.2: The original image was split into eight tiles (four in each of the two rows). The last tile in the second row was not part of the dataset. One can recognize the tiles that were used for training as the model overfitted the trainingset and perfectly memorized the bounding boxes in these tiles. The last tile in the first row and the second last tile in the second row show unseen testdata. Here the model fails and predicts that the entire tile shows ragweed with a 40 to 48% confidence where it should not report anything.

Metric	**at IoU**	**Areas**	**maxDets**	**Score**
Average Precision (AP)	0.50:0.95	all	100	0.252
Average Precision (AP)	0.50	all	100	0.507
Average Precision (AP)	0.75	all	100	0.195
Average Recall (AR)	0.50:0.95	all	1	0.219
Average Recall (AR)	0.50:0.95	all	10	0.366
Average Recall (AR)	0.50:0.95	all	100	0.422

A mAP@[.5:.95] of just 0.252 and a fairly high average recall of 0.422 indicates that this particular object detector is ineffective for our dataset. The main issue might be that the groundtruth bounding boxes are already noisy containing lots of vegetation that is not ragweed. Another problem could be that the number of training images is too low resulting in overfitting the model to our trainingdata.

Despite the model beeing a feature pyramid network and extensive randomized rescaling of images during training (as part of the data augmentation pipeline) the detection performs even worse when applied to the entire 4k input image. The reason might be that rescaling the entire 3840x2160 input image to 640x640 leads to a lower number of pixels per plant as when rescaling a single tile of 960x1080

to the detectors input resolution of 640x640.

Tiling the input images produces another problem, some plants may not be discovered when they are at the border between two tiles. Budda et al. suggested to feed overlapping image tiles into the detector and then merge the resulting bounding-boxes again. They present an algorithm to do this in their paper. This additional complexity as well as the unconvincing detection results on our dense-vegetation-dataset led to another experiment looking at semantic segmentation models.

5.2 Segmentation Model

Semantic segmentation is the task of assigning a label to every pixel in an image. The current state of the art for this task are models of the deeplab family. [10]

A recent implementation of deeplabv3 for tensorflow 2.x from Haas [11] required just minor customizations to train the original deeplabv3 model on the ragweed-segmentation dataset. Similar to the original deeplab paper Haas uses a mobilenetv3 as feature extraction backbone.

The model was trained for 100 epochs, keeping the original input resolution (3841 x 2161) as the fine details of the ragweed leafs are considered important for high quality recognition results. Due to memory constraints of our training hardware the batch size was limited to 1 Image only.

As backbone a MobileNetV3Small pretrained on the cityscapes dataset was used.

Figure 5.3 shows primliminary results from the deeplab v3 semantic segmentation model on unseen testdata.

The model scores at 72% mIoU. on the ragweed-testdata. The inference latency is roughly 1.5 seconds per 4k input image. Another experiment was performed where the Input resolution was scaled down to 541 x 961 allowing a larger batch size during training. However this version of the model scored much lower not exceeding 50% mIoU.

In a next step an optimized version of the model will be trained for deployment on embedded hardware.

Figure 5.3: Purple shows the groundtruth, yellow marks the models prediction.

Bibliography

[1] R. Richter, U. Berger, S. Dullinger, F. Essl, M. Leitner, M. Smith, and G. Vogl, "Spread of invasive ragweed: Climate change, management and how to reduce allergy costs," *Journal of Applied Ecology*, 08 2013.

[2] R. Buttenschøn, Waldispühl, and B. C., "Guidelines for management of common ragweed, ambrosia artemisiifolia." 01 2010. [Online]. Available: http:/www.EUPHRESCO.org

[3] "Ean." [Online]. Available: https://ean.polleninfo.eu/Ean/

[4] "rwphoneclassifier." [Online]. Available: https://github.com/LukasSteindl/interdisciplinaryproject

[5] I. Plaščak, M. Jurišić, A. Šiljeg, L. Jeftić, D. Zimmer, and B. Željko, "Remote detection of ragweed (ambrosia artemisiifolia l.)," *Tehnički Glasnik*, vol. 12, pp. 226–230, 12 2018.

[6] K. Buddha, H. J. Nelson, D. Zermas, and N. Papanikolopoulos, "Weed detection and classification in high altitude aerial images for robot-based precision agriculture," in *2019 27th Mediterranean Conference on Control and Automation (MED)*, 2019, pp. 280–285.

[7] A. L. Blum and R. L. Rivest, "Training a 3-node neural network is np-complete," *Neural Networks*, vol. 5, no. 1, pp. 117–127, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608005800103

[8] "supervisely." [Online]. Available: https://supervise.ly/

[9] "tfmodelzoo." [Online]. Available: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

[10] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017.

[11] B. Haas, "Compressing mobilenet with shunt connections for nvidia hardware," Gusshausstrasse 27–29 / 384, 1040 Wien, May 2021.