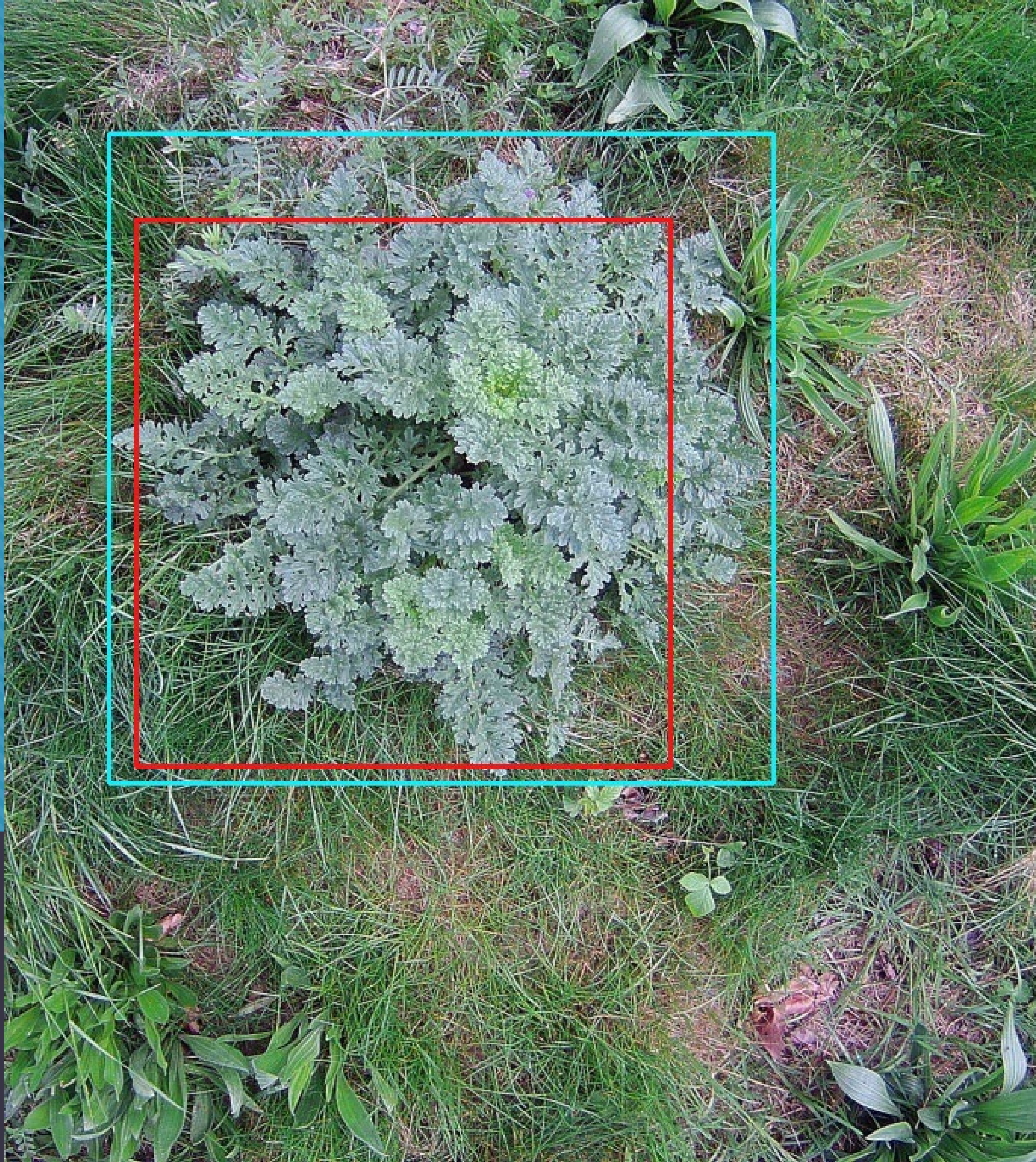
Object Detection Based On Convolutional Neural Networks Of *Senecio jacobaea* For Weed Control

Bachelor Thesis by Jonas Zender (21125)

- 1st Supervisor: Prof. Dr.-Ing. Rolf Becker
- 2nd Supervisor: Prof. Dr. Daniela Lud



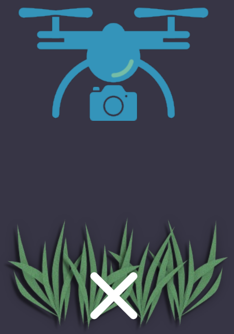
Introduction

Problem:

- *Senecio jacobaea* has spread massively
- *Contains hepatotoxic Pyrrolizidine Alkaloids (PAs)*
- *Causes substantial damages to farmers*
- *Is tedious to remove, especially in environmentally protected areas*

Vision: Two-step weed control method

1. Step
- Find candidates
- Map coordinates



2. Step
- Validate candidates
- Apply herbicide

Questions:

- Can *Senecio jacobaea* be detected at 1m height among other vegetation?
- How much does the dataset size affect the quality of the model?
- How much does the plant size influence the result?

Materials and Methods I
Collecting Data

- Collecting a dataset
  - 2128 images
  - 1m above ground
  - Akaso Brave 4 Action Cam
  - 5120x3840 (4:3)
- Resizing to 1024x768 for easier handling
- Introducing unique IDs to distinguish
  - Date
  - Location
  - Time of day
  - camera

ID/Date/Time	Location	Type	Weather	Imgs	Insts
kam_210515_n_acb4 15.05.21 12:00 – 14:30	KaLi: Stephans-wäldchen, LaGa entrance	Park	Mostly cloudy, Rainy	144	201
kle_210516_n_acb4 16.05.21 11:30 – 14:40	Kleve: Forstgarten, Joseph-Beuys Allee	Park	Cloudy, later sunny	469	1242
moe_210518_n_acb4 18.05.21 11:30 – 13:30	Moers: Schlosspark, Grafschafter Kampfbahn	Park, demolition area	Mostly sunny, cloudy	85	209
moe_210527_a_acb4 27.05.21 16:30 – 17:00	Moers: Grafschafter Kampfbahn	Demolition area	Cloudy, rainy	112	381
kam_210529_a_acb4 29.05.21 14:00 – 15:00	KaLi: Stephans-wäldchen, LaGa entrance	Park	Sunny	90	151
mil_210529_e_acb4 17:00 – 18:30	Rheinberg Millingen Heidestraße 15	Meadow, pasture	Sunny	260	517
leu_210601_n_acb4 01.06.21 12:00 – 15:30	Leucht: Strohweg, Bierweg, Stappweg	Forest, fields	Sunny	968	3022

Materials and Methods II
Creating Annotations

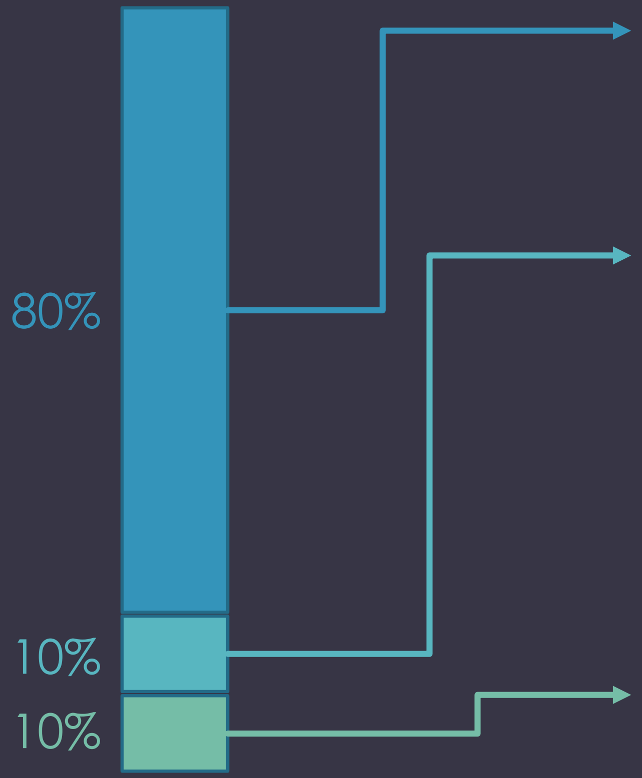
- Model output:
  - Fixed number of predictions
  - Each prediction contains
    - Class
    - Confidence
    - Bounding box coordinates
- Labels
  - Each image has its own XML label
  - For each specimen, the label contains
    - Class
    - Bounding Box coordinates
  - Done using CVAT
  - 5723 specimens in 2128 images

leu_210601_n_acb4_0436.xml


```xml
<annotation>
  <folder>leu</folder>
  <filename>leu_210601_n_acb4_0436.JPG</filename>
  <source>
    <database>Unknown</database>
    <annotation>Unknown</annotation>
    <image>Unknown</image>
  </source>
  <size>
    <width>1024</width>
    <height>768</height>
    <depth></depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>Senecio</name>
    <occluded>0</occluded>
    <bndbox>
      <xmin>308.88</xmin>
      <ymin>230.42</ymin>
      <xmax>738.71</xmax>
      <ymax>739.41</ymax>
    </bndbox>
  </object>
</annotation>
```


Materials and Methods III
Training, Validation, and Test Set

2128 images



Training Dataset

- Used for optimization of the model parameters

Validation Dataset

- Used during training to monitor
  - Model progress
  - Under- and Overfitting
- Evaluated after each iteration (epoch) over the training dataset

Test Dataset

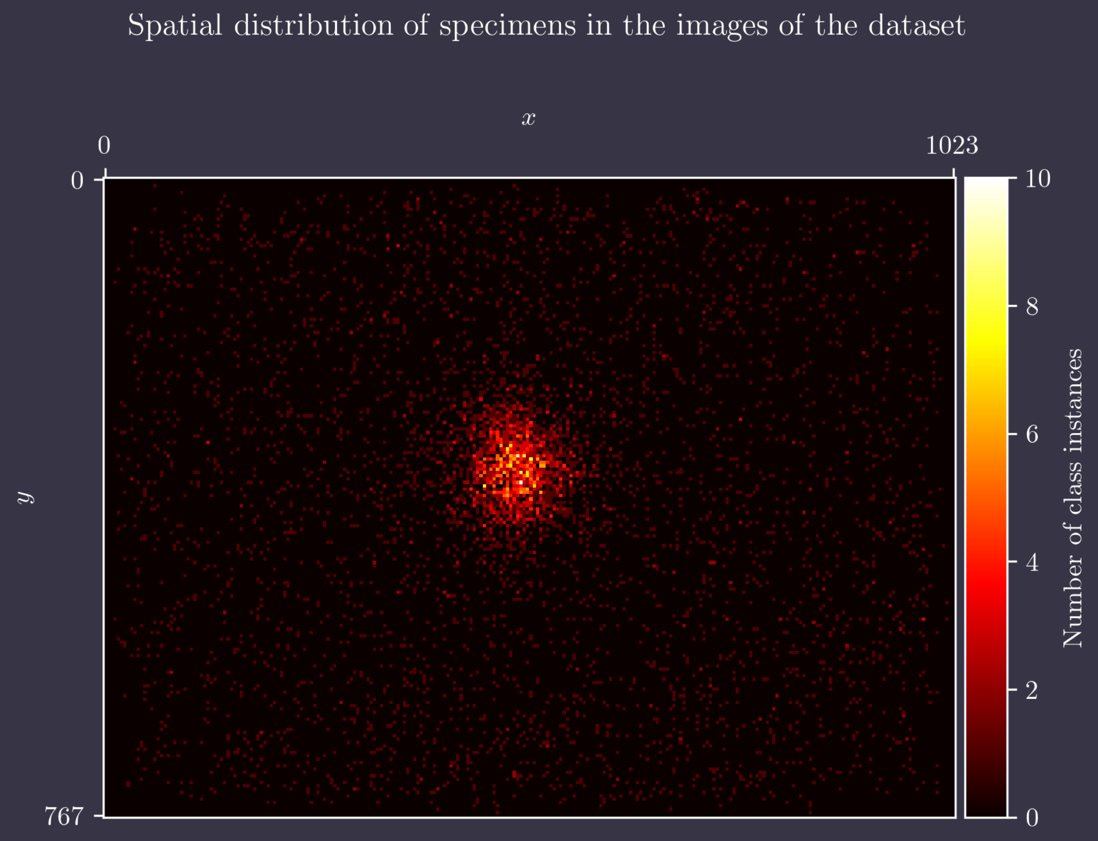
- Used after the training to evaluate the performance of the finished model

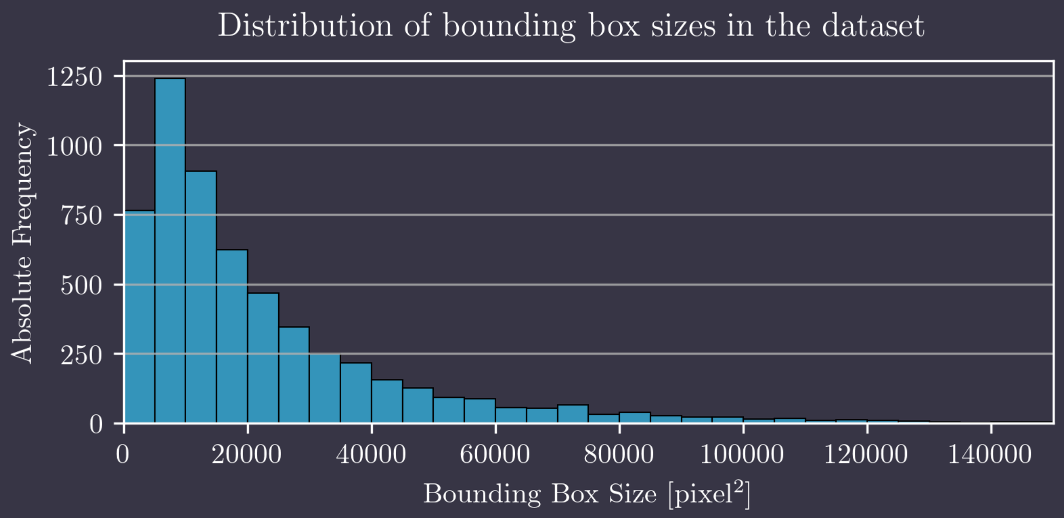
Materials and Methods IV
PASCAL VOC Dataset Format

- PASCAL VOC dataset format
  - Standardized file & directory structure
  - Suitable for training SSD-MobileNet-v1
- Python was used for automatizing the process
- Images were shuffled before being distributed to training, validation, and test set

```
<dataset name>
├── Annotations    (XML labels)
├── ImageSets
│   └── Main
│       ├── default.txt  ┐
│       ├── test.txt     │  (define training,
│       ├── train.txt    │  Validation, test set)
│       └── val.txt      ┘
├── JPEGImages     (Images)
└── labels.txt     (define classes)
```

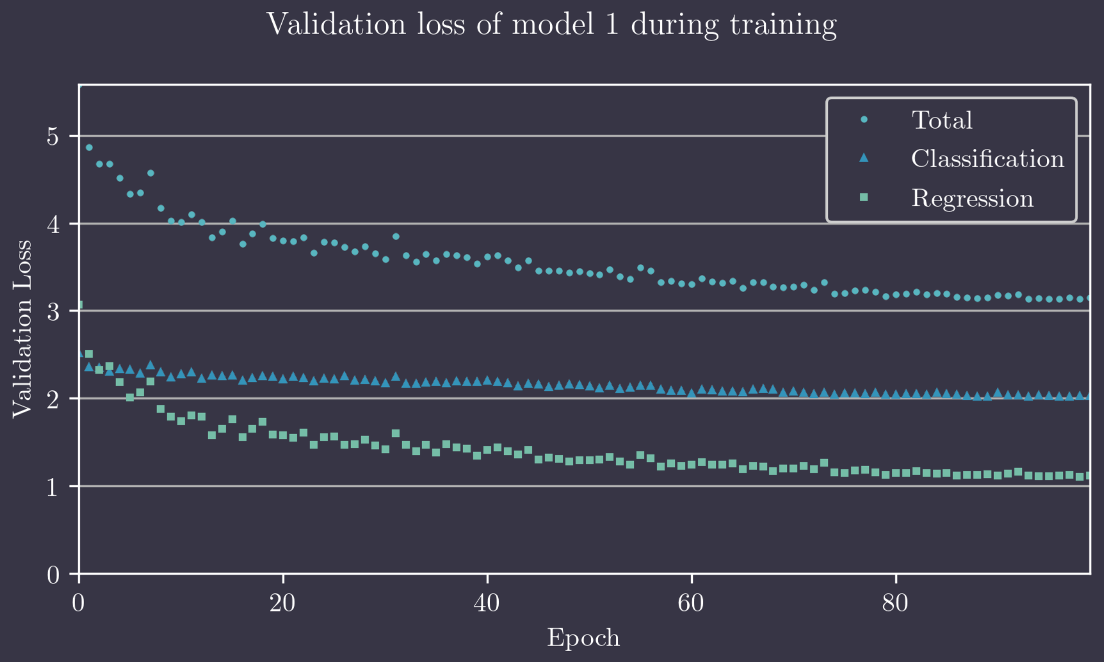
Materials and Methods V
Dataset Visualization





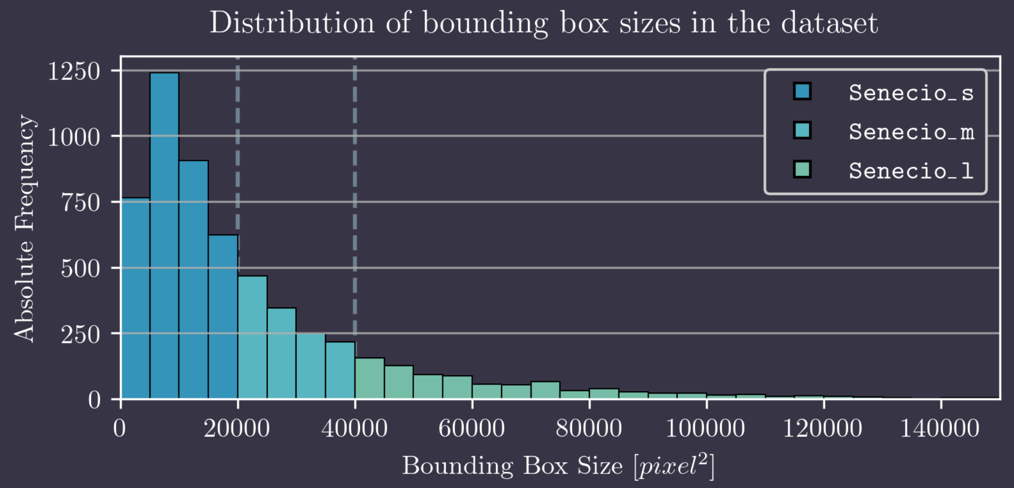
Materials and Methods VI
Training Model 1

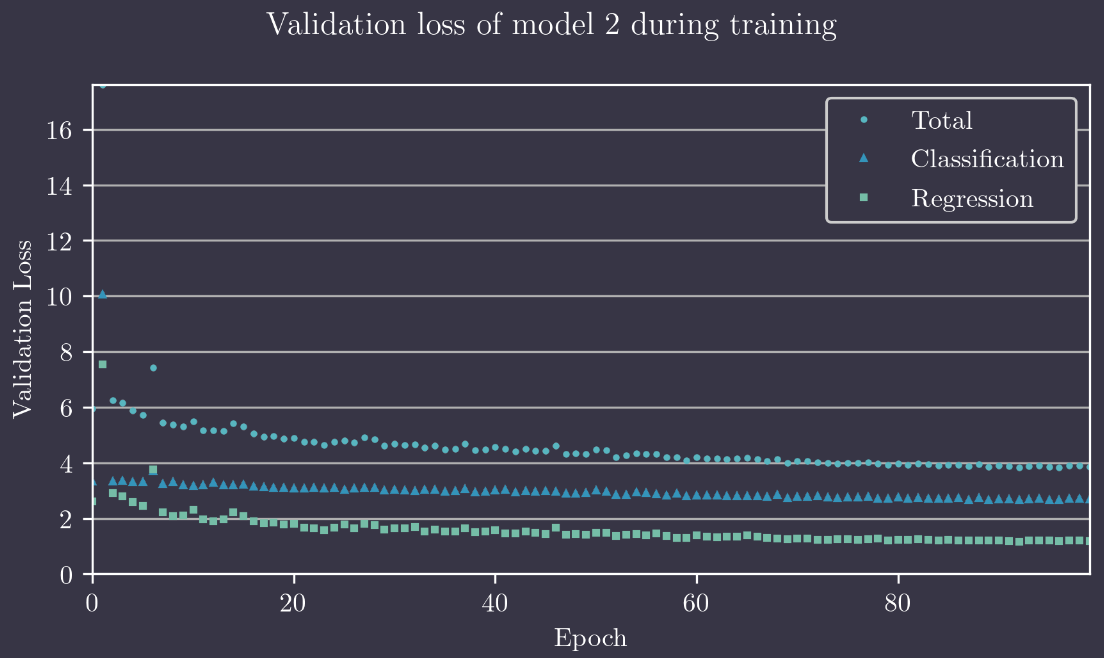
- Training parameters:
  - NVIDIA Jetson Xavier NX
  - Modified training script to write logging messages to log file
  - 100 epochs
  - 851 batches (batch size 2)
  - Learning rate: 0.01
- Log file plot:
  - Validation loss is calculated after each epoch
  - Validation loss is composed of
    - Classification loss
    - Regression loss



Materials and Methods VII
Training Model 2

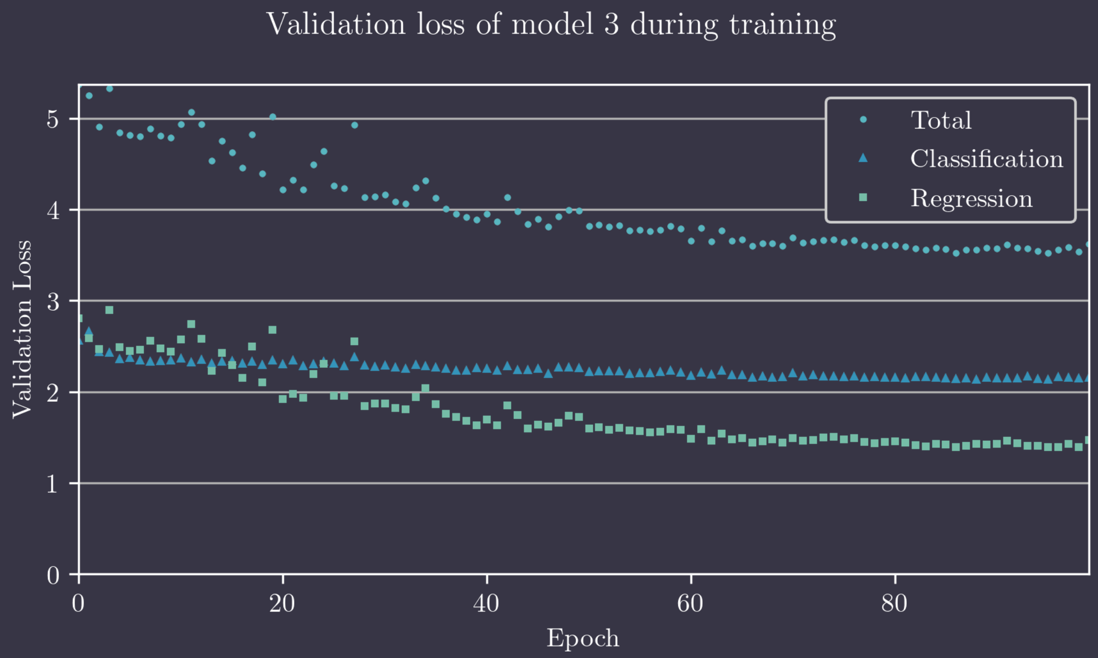
- Distribution of the `Senecio` class into three separate size classes:
  - `Senecio_s` (3532)
  - `Senecio_m` (1281)
  - `Senecio_l` (907)
- Training parameters were the same as before





Materials and Methods VIII
Training Model 3

- The size of the training dataset was reduced by 50%. Besides that, it is identical with dataset 1
- Training parameters were the same as before

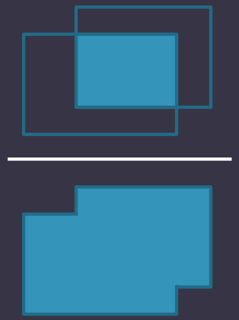


Materials and Methods IX
Inference, Predictions, and IoU

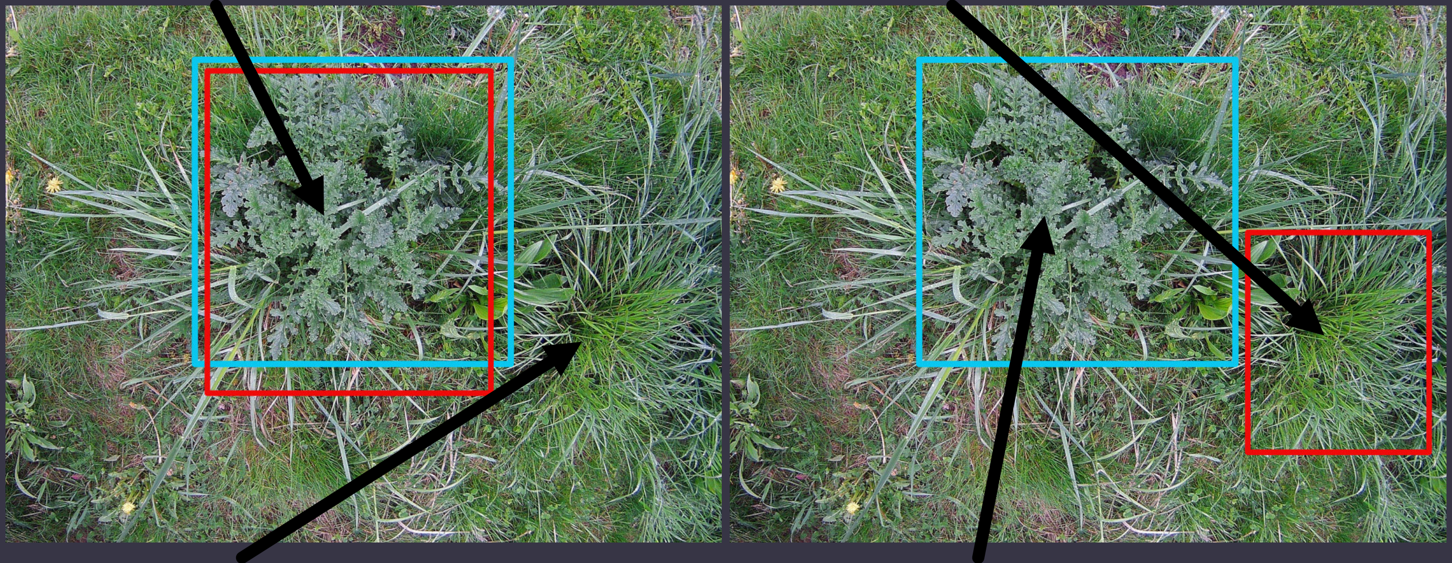
- Inference on test datasets was done with detectnet
- Model output consisted of confidence, class and bounding box coordinates
- Confidence thresholds of 20%, 30%, 40%, and 50% were used

Intersection over Union (IoU)

$$\mathrm{IoU} = \frac{\mathrm{area}(B_{gt} \cap B_p)}{\mathrm{area}(B_{gt} \cup B_p)} =$$



True Positive (TP) False Positive (FP)



True Negative (TN) False Negative (FN)

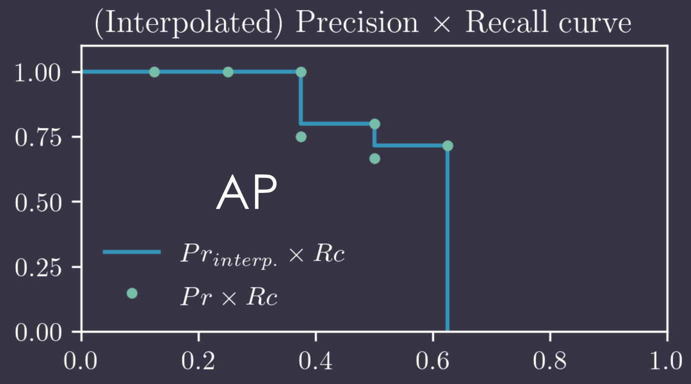
Ground Truth / Prediction

Materials and Methods X
Precision and Recall, AP, mAP

- $Precision = \frac{TP}{TP+FP} = \frac{TP}{all\ predictions}$
- $Recall = \frac{TP}{TP+FN} = \frac{TP}{all\ ground\ truths}$
- Average Precision (AP):
  - Area under interpolated precision x recall curve
  - Class-specific metric
- Mean Average Precision (mAP):
  - Arithmetic mean of all APs
  - Metric for model as a whole
- Evaluation was done with an open-source object detection metrics toolbox

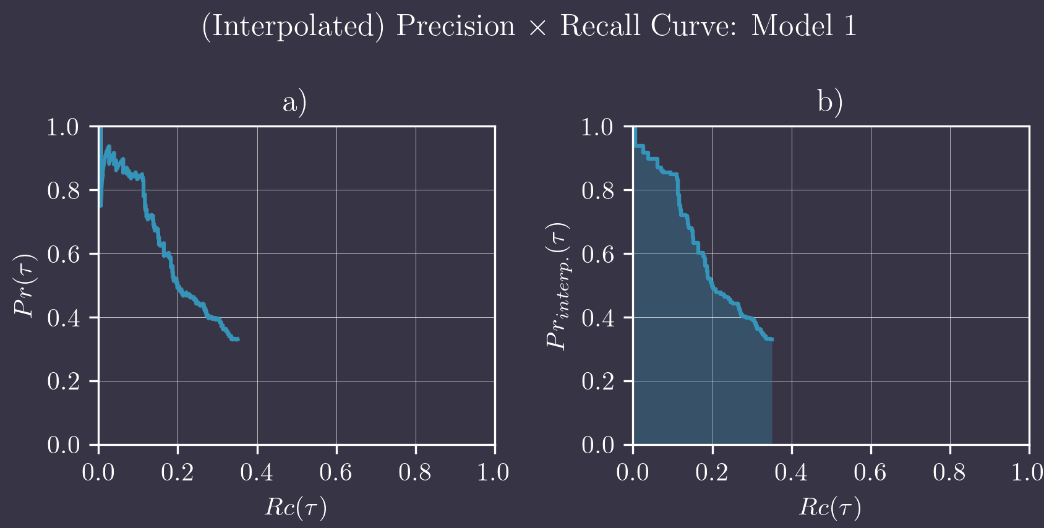
Example Model Evaluation (8 ground truth labels)

Bbox	τ	IoU	IoU>0.5?	ΣTP	ΣFP	$Pr(\tau)$	$Rc(\tau)$	$Pr_{interp}(\tau)$
B	99%	0.80	True	1	0	1.000	0.125	1.000
G	95%	0.93	True	2	0	1.000	0.250	1.000
D	94%	0.71	True	3	0	1.000	0.375	1.000
C	84%	0.44	False	3	1	0.750	0.375	0.800
A	76%	0.65	True	4	1	0.800	0.500	0.800
F	72%	0.00	False	4	2	0.667	0.500	0.714
E	67%	0.53	True	5	2	0.714	0.625	0.714



Results I
Model 1 – Original Dataset

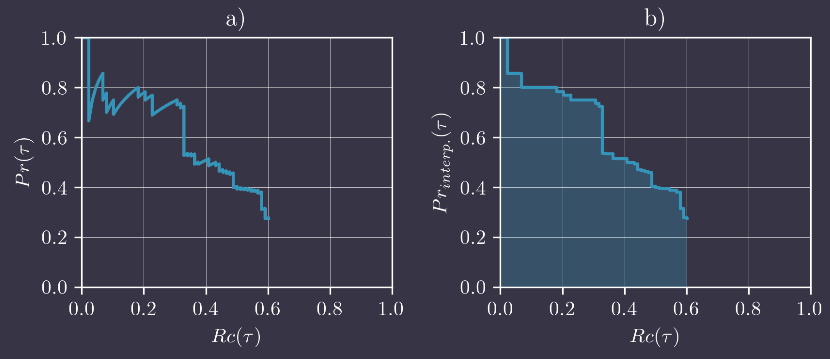
- Confidence threshold: 20%
- Metrics
  - Precision: 33.06%
  - Recall: 35.11%
  - AP/mAP: 21.93%
- Average Bounding Box Area
  - Ground Truth: 23823.78 pixel²
  - Predictions: 52032.08 pixel²



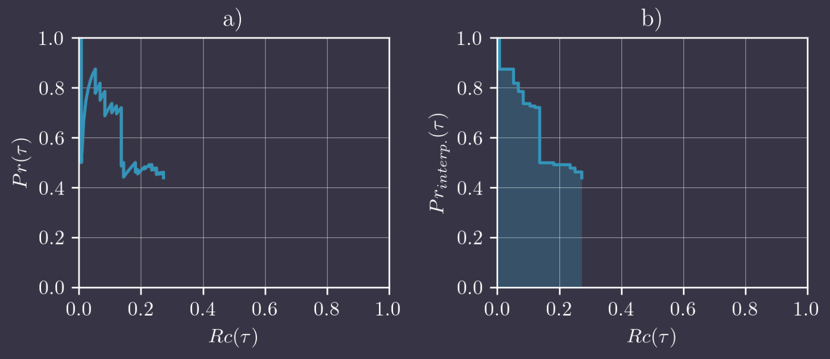
Results II
Model 2 – Dataset with Size Classes

- Confidence threshold: 20%
- Metrics (mAP: 19.44%)
  - Senecio_l
    - Precision: 27.46%
    - Recall: 60.23%
    - AP: 38.63%
  - Senecio_m
    - Precision: 43.90%
    - Recall: 27.27%
    - AP: 17.69%
  - Senecio_s
    - Precision: 34.21%
    - Recall: 3.78%
    - AP: 2.00%

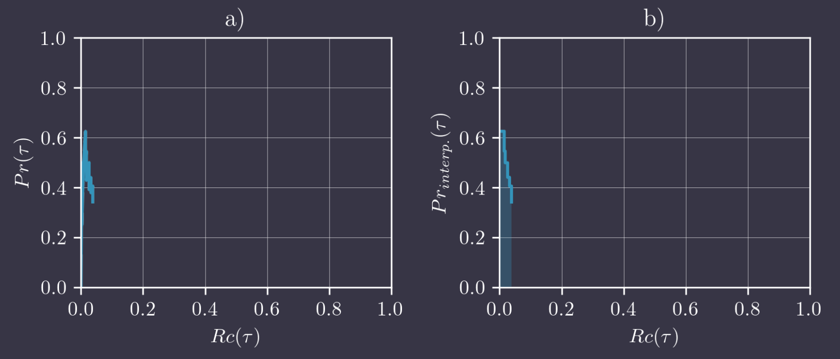
(Interpolated) Precision × Recall Curve: Model 2 - Senecio_l



(Interpolated) Precision × Recall Curve: Model 2 - Senecio_m

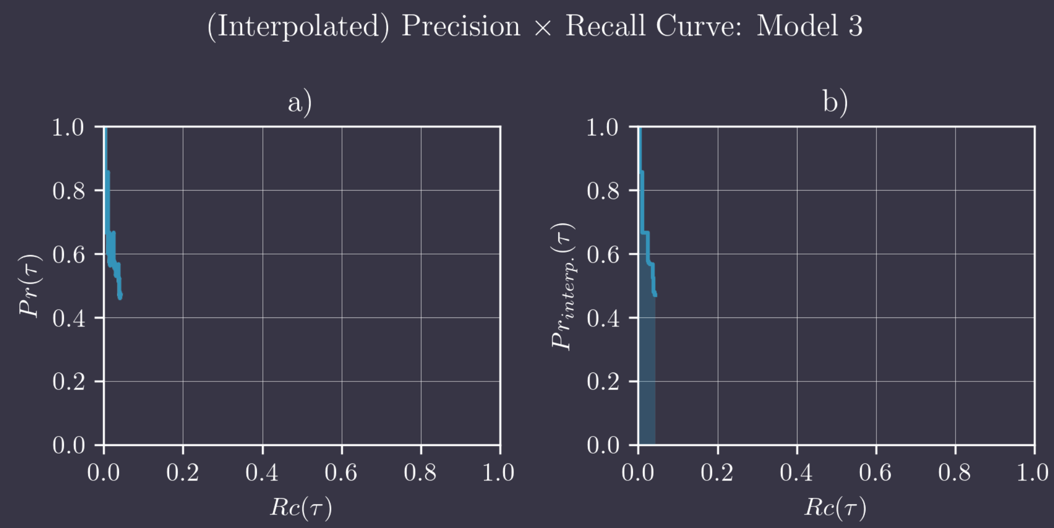


(Interpolated) Precision × Recall Curve: Model 2 - Senecio_s



Results III
Model 3 – Dataset with 50% Training Set

- Confidence threshold: 30%
- Metrics
  - Precision: 47.06%
  - Recall: 4.26%
  - AP/mAP: 2.88%
- Average Bounding Box Area
  - Ground Truth: 23823.78 pixel²
  - Predictions: 42084.12 pixel²



Results IV
Sample Image 1

Image: `kam_210515_n_acb4_0023`

Ground Truth / Prediction

Model 1

Model 2

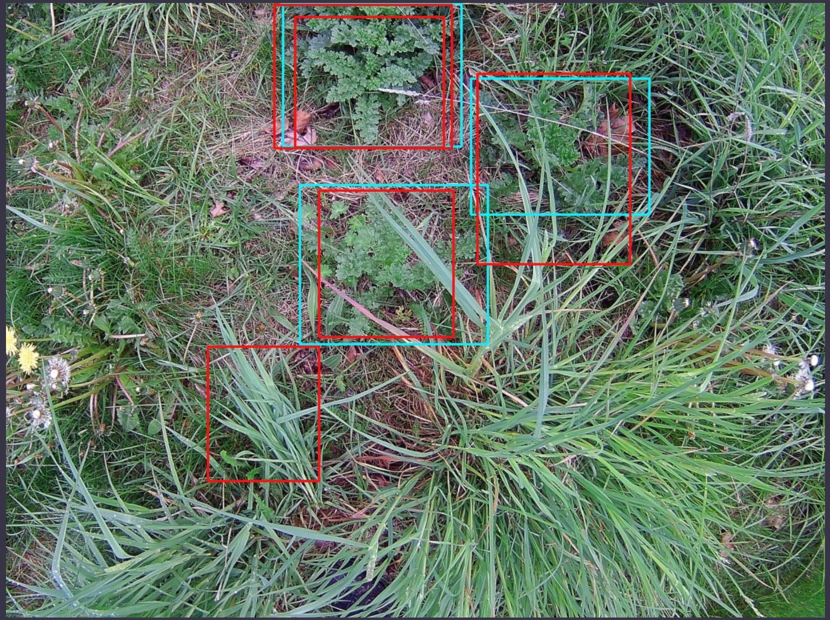
Model 3

Results V
Sample Image 2

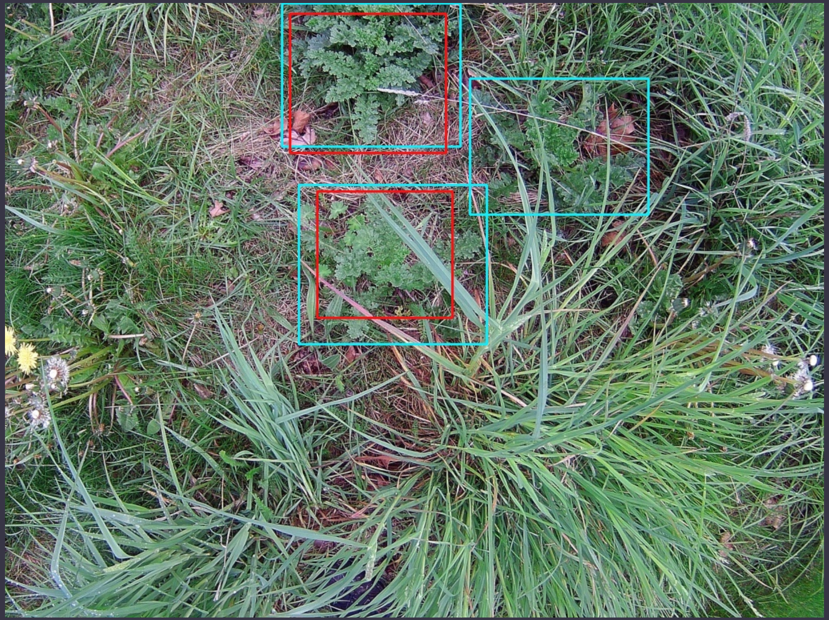
Image: `kam_210515_n_acb4_0039`

Ground Truth / Prediction

Model 1



Model 2



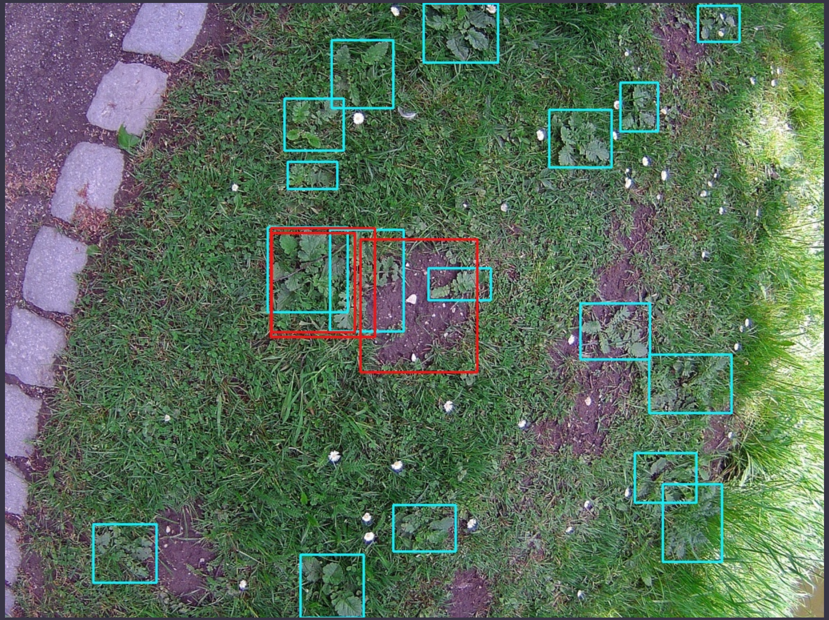
Model 3

Results VI
Sample Image 3

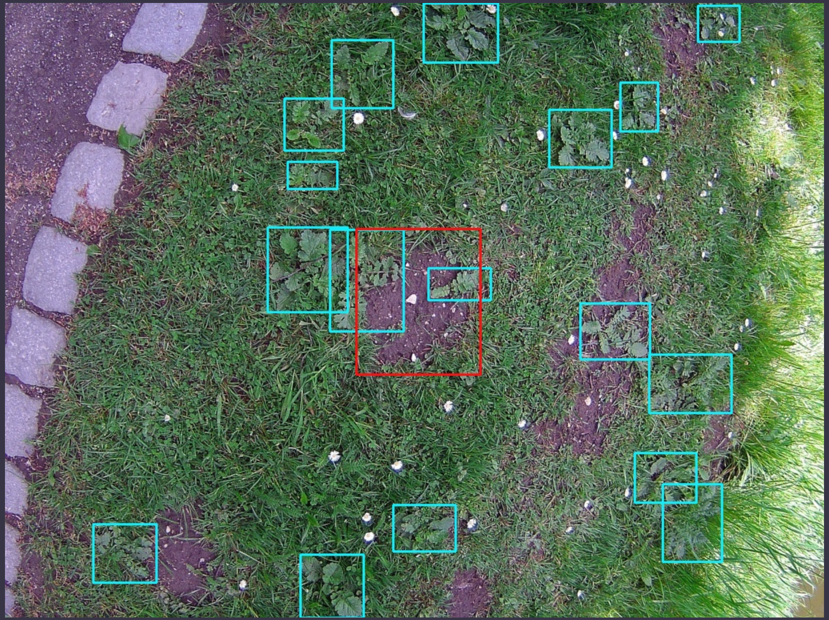
Image: `kle_210516_n_acb4_0184`

Ground Truth / Prediction

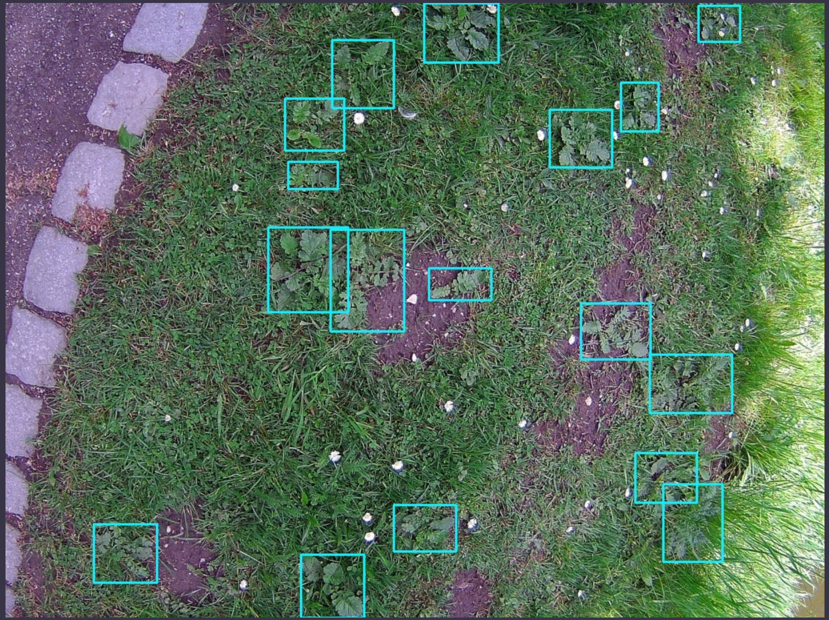
Model 1



Model 2



Model 3



Results VII
Sample Image 4

Image: `leu_210601_n_acb4_0307`

Ground Truth / Prediction

Model 1

Model 2

Model 3

Results VIII
Sample Image 5

Image: `moe_210527_a_acb4_0102`

Ground Truth / Prediction

Model 1

Model 2

Model 3

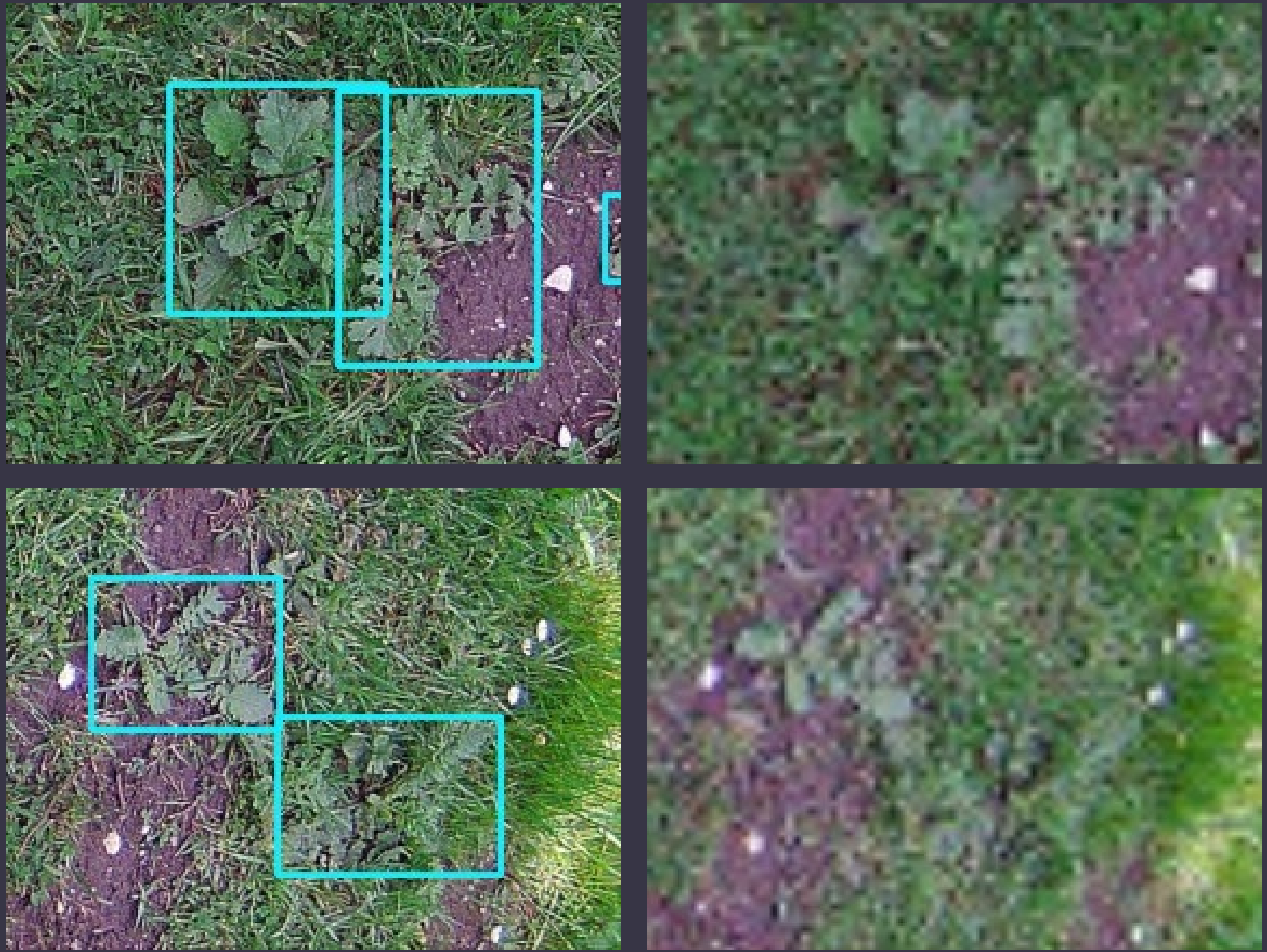
Discussion I
Distorted Bounding Boxes

- In some cases, the predicted bounding boxes were distorted compared to the ground truth labels
- Center point is needed for extracting GPS coordinates
- If the centers of predicted and ground truth bounding boxes are reasonably close together, it is not a problem

Ground Truth / Prediction
Points are denoting the center of the bounding boxes

Discussion II
Small Bounding Boxes

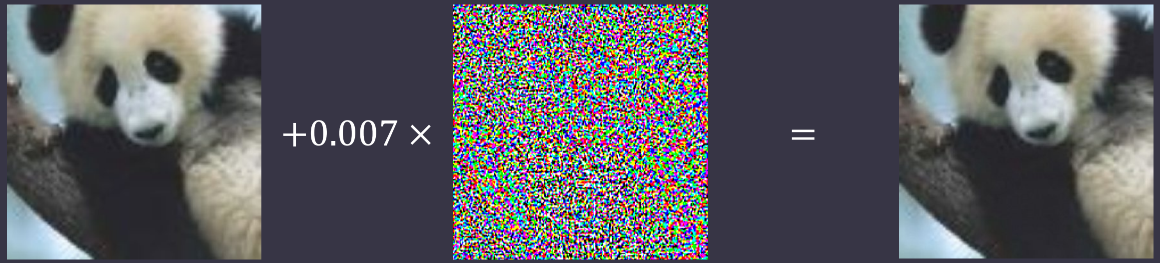
- The models have problems detecting small specimens (<20000 pixel^2)
- Resizing to 300x300 pixels potentially loses features like edges etc.
- Current object detection methods perform significantly worse for small objects since they lack characteristic shapes and textures
- Classification and bounding box regression might be hard or even impossible
- Colour alone is often not enough to distinguish small specimens from the background



Discussion III
False Positive Predictions

- All models misidentified objects as *Senecio jacobaea*
  - What is the model actually looking for?
  - CNNs are considered a black box whose decisions cannot be properly explained
- FPs are not desirable, but would be corrected in the validation step

Adversarial Example (Goodfellow et al. 2015)
Demonstrated with GoogLeNet (ImageNet dataset)



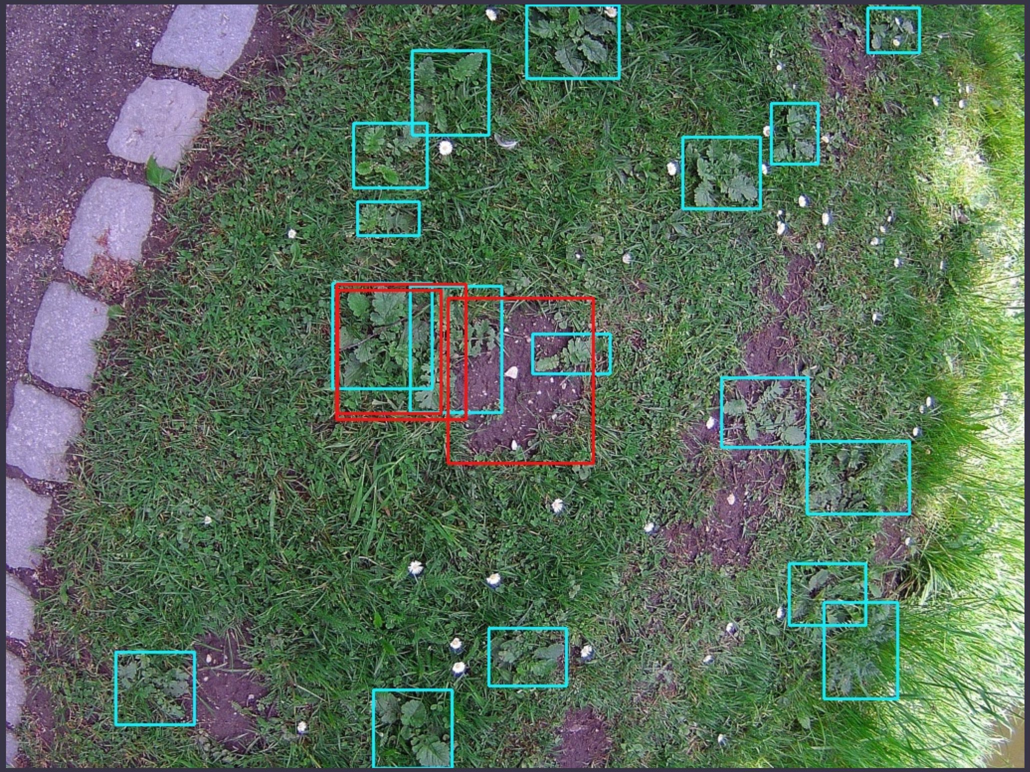
„panda“
57.7% confidence

„nematode“
8.2% confidence

„gibbon“
99.3% confidence

Discussion IV
False Negative Predictions

- Possible Reasons for FNs
  - Small Size
  - Cut off at image edges
  - Covered by other vegetation
  - Bad lighting or contrast
- FNs would not be validated again in the second step since they are not mapped in the first place. So, the error does not get corrected
  - This impairs the quaility of the weed control method
  - It is better to trade off some precision for a higher recall



Discussion V
Duplicate Detections

- Duplicate Detections are problematic since
  - Plants would be treated multiple times with herbicide
  - They inflate the precision and recall above the real value and make evaluation difficult
- Non-Maximum Suppression (NMS)
  - Post-processing algorithm
  - Merges all detections referring to the same ground truth object
  - Compares confidence and IoUs of detections

Discussion VI
Issues With The Dataset

- There might be specimens missing a lable
  - Research has shown that there is a significant performance drop if 50% or more labels are missing
- The distribution of specimens into size classes might not be optimal
  - It would be better to relabel the dataset manually by grouping specimens of similar appearance
- Some field conditions are underrepresented in the dataset
  - Dataset size should be increased and should represent all conditions more equally
- Type of illumination might greatly change the appearance of *Senecio jacobaea* specimens

Discussion VII
Conclusion

- It is possible to detect *Senecio jacobaea* in images at a height of 1 meter with background vegetation of similar apperance
- A toolchain including all steps, from collecting data to evaluating models, was proposed and tested
- It was shown that a larger dataset greatly improves the quality of the model
- The size of the specimens drastically influences the models ability to find it

- Object detection of *Senecio jacobaea* is a solvable albeit challenging problem.

References I

- Bochkovskiy, A., Wang, C., and Liao, H. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *CoRR*, abs/2004.10934.
- Bosquet, B., Mucientes, M., and Brea, V. M. (2018). STDnet: A ConvNet for Small Target Detection. *Proceedings of the British Machine Vision Conference (BMVC)*.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Deutscher Verband für Landschaftspflege e.V. (2017). Kreuzkräuter und Naturschutz. In *Tagungsband der internationalen Kreuzkraut-Fachtagung in Göttingen 2017*, Nr. 23 der DVL-Schriftenreihe Landschaft als Lebensraum.
- Dumoulin, V. and Visin, F. (2018). A guide to convolution arithmetic for deep learning. *arXiv e-prints*.
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338.
- Franklin, D., Yato, C., Kislan, T., Nguyen, D., Darbha, R., Tran, B., and Linderoth, M. (2016). Hello AI World. https://github.com/dusty-nv/jetson-inference.
- Goodfellow, I. J., Shlens, J., and Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. *arXiv e-prints*.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Hosang, J. H., Benenson, R., and Schiele, B. (2017). Learning non-maximum suppression. *CoRR*, abs/1705.02950.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv e-prints*.
- Kieloch, R. and Domaradzki, K. (2011). The role of the growth stage of weeds in their response to reduced herbicide doses. *Acta Agrobotanica*, 64:259–266.
- Lam, O. H. Y., Dogotari, M., Prüm, M., Vithlani, H. N., Roers, C., Melville, B., Zimmer, F., and Becker, R. (2021). An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using Rumex obtusifolius as a case study. *European Journal of Remote Sensing*, 54(sup1):71–88.
- Lampen, A. (2017). Risikobewertung: Wie hoch ist die Gefährdung duch Pyrrolizidin-Alkaloide? In *Tagungsband der internationalen KreuzkrautFachtagung in Göttingen 2017*, Nr. 23 der DVL-Schriftenreihe Landschaft als Lebensraum, pages 25–34.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proc. of the IEEE*.
- Lin, T.-L., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft COCO: Common Objects in Context. *arXiv e-prints*.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *arXiv e-prints*.
- Mogili, U. R. and Deepak, B. B. V. L. (2018). Review on Application of Drone Systems in Precision Agriculture. In *International Conference on Robotics and Smart Manufacturing (RoSMa2018)*, 133, pages 502–509.
- Muthukumar, V., Narang, A., Subramanian, V., Belkin, M., Hsu, D. J., and Sahai, A. (2020). Classification vs regression in overparameterized regimes: Does the loss function matter? *CoRR*, abs/2005.08054.

References II

- Neumann, H. and Huckauf, A. (2015). Jakobs-Kreuzkraut (Senecio jacobaea): eine Ursache für Pyrrolizidin-Alkaloide im Sommerhonig? *Journal für Verbraucherschutz und Lebensmittelsicherheit*, 11:105–115
- NVIDIA (2021). Jetson Xavier NX Developer Kit. `https://developer.nvidia. com/embedded/jetson-xavier-nx-devkit`.
- Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., and da Silva, E. A. B. (2021). A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*, 10(3).
- pandas development team, T. (2020). pandas-dev/pandas: Pandas.
- Prashanth, D. S., Mehta, R. V. K., and Sharma, N. (2020). Classification of Handwritten Devanagari Number – An analysis of Pattern Recognition Tool using Neural Network and CNN. Procedia Computer Science, 167:2445–2457. *International Conference on Computational Intelligence and Data Science*.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *arXiv e-prints*.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., Berg, A. C., and Li, F. (2014). ImageNet Large Scale Visual Recognition Challenge. *CoRR*, abs/1409.0575
- Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Johannes222, Chenuet, M., a andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607, Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., vugia truong, zliang7, lizhming, and Truong, T. (2020). opencv/cvat: v1.1.0.
- Stojnić, V., Risojević, V., Muštra, M., Jovanović, V., Filipi, J., Kezić, N., and Babić, Z. (2021). A Method for Detection of Small Moving Objects in UAV Videos. *Remote Sensing*, 13(4).
- Sumit, S., Watada, J., Roy, A., and Rambli, D. (2020). In object detection deep learning methods, YOLO shows supremum to Mask R-CNN. *Journal of Physics: Conference Series*, 1529:042086.
- Suter, M. and Lüscher, A. (2017). Habitatpräferenzen von Jakobs- und WasserKreuzkraut und Risikofaktoren für deren Auftreten. In *Tagungsband der internationalen Kreuzkraut-Fachtagung in Göttingen 2017*, Nr. 23 der DVL-Schriftenreihe Landschaft als Lebensraum, pages 9–17.
- Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.
- Xu, M., Bai, Y., and Ghanem, B. (2019). Missing Labels in Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9:611–629.
- Zehm, A. (2017). Auf welchen Flächen mit Relevanz für den Naturschutz sollen welche Kreuzkräuter reguliert werden? In *Tagungsband der internationalen Kreuzkraut-Fachtagung in Göttingen 2017*, Nr. 23 der DVL-Schriftenreihe Landschaft als Lebensraum, pages 19–22.
- Zeiler, M. D. and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham. Springer International Publishing.
- Zender, J. (2021). Object Detection based on Convolutional Neural Networks of Senecio jacobaea for Weed Control. `https://git.hsrw.eu/21125/thesis_doc`.