Silage Bale Detection for the «Cultivable Area» Update of the Cantonal Agricultural Office, Thurgau

Adrian F. Meyer[1] and Denis Jordan[1]

[1]Institute Geomatics – Fachhochschule Nordwestschweiz FHNW
adrian.meyer@fhnw.ch

Abstract. In Switzerland direct subsidies are paid to farms for sustainable agricultural practice. The cultivable agricultural area layer (German: Landwirtschaftliche Nutzfläche, LN) serves as an annual basis for the calculation of these contributions at the Swiss cantonal agricultural offices. Material deposits like silage bale stacks are usually excluded from the LN. Therefore, the canton of Thurgau could profit from a spatial vector layer indicating locations and area consumption extent of silage bale stacks intersecting with the LN perimeter.

To ease the monitoring process, we propose a Mask-RCNN based prototypical Deep Learning framework which was trained on 10cm SWISSIMAGE orthophoto datasets (swisstopo, Bern). Embedded in an efficient python-based geodata workflow the model boasts a high F1-Score of 92% on evaluation data. This approach allows robust and accurate inference detections over the whole cantonal area. Having the silage bale stack detections at hand reduces the manual workload of the responsible official by directing the eyes to the relevant hotspots.

Keywords: Agriculture; Object Detection; Monitoring; Administration; Subsidy Payments; Cadastral; Mask-RCNN; Aerial Imagery; Remote Sensing

1 Introduction

Switzerland's direct payment system is the basis for sustainable, market-oriented agriculture. The federal government supports local farms in the form of various types of subsidies such as biodiversity contributions, landscape quality contributions, or food supply security contributions.

Subsidies are often calculated by area and the agricultural offices of the respective cantonal administration are responsible for monitoring agricultural areas in order to approve the requested amounts. Only certain land usage profiles are eligible for subsidies payment. The cultivable agricultural area layer (German: Landwirtschaftliche Nutzfläche, LN) is a GIS product maintained by the cantonal agricultural offices and serves as the key calculation index for the receipt of contributions.

Major adjustments of the LN are part of the periodic update (German: Periodische Nachführung, PNF) which is carried out within the framework of the official cadastral survey (German: Amtliche Vermessung, AV) [1][2], while smaller updates are performed annually. Its correct determination is of immense importance, because if the LN vector polygons derived from the cadastral survey data deviate largely from the actual conditions on site, the monitoring effort during the annual farm structure data survey process (German: Betriebsstrukturdatenerhebung) [10][11] increases.

Farm areas that are not usable for effective productive agriculture are to be excluded from the LN. This includes material deposits such as silage hay bales storage plots which are constantly changing due to the high degree of mechanization in agriculture and can sometimes fall within the perimeter of the registered LN. The tracking of these areas with conventional surveying such as repeated field visits or the visual interpretation of current aerial imagery proves to be very time-consuming and costly. Therefore we propose an automatized workflow to predict areas currently in use for silage bale stack deposits.

Artificial convolutional neural networks (CNN) based on deep learning (DL) have been used for automated detection and classification of image features for quite some time. Reliable detection from aerial imagery using applications of DL would enable cost-effective detection of these storage areas and provide added value to agricultural office of the Canton of Thurgau (German: Landwirtschaftsamt, LWA) but also in other cantons.

In the context of the publicly financed project “Swiss Territorial Data Lab” the applicability of CNNs to generate a localized silage bale stack inventory was investigated. The delivered dataset should consist of vector polygons which are compatible with the LWA’s webGIS workflow and should be made available together with

new acquisitions of aerial imaging campaigns. This project therefore aims at the development of an efficient and flexible algorithm which offers a highly accurate performance and can be quickly deployed over the complete cantonal area of Thurgau (approx. 992 km²). For the LWA it is important that the detections are precise, relevant in size, and do not contain a large number of false positives.

2 Method

2.1 Overview

Silage bale stacks as target objects are clearly visible on the newest 2019 RGB layer of the 10cm SWISSIMAGE dataset [15]. A few hundred of each of these objects were manually digitized as vector polygons ("annotations") with QGIS [14] in separate workflows using a semi-automatic approach.

In order to limit computational load, the current LN extent of the canton of Thurgau was defined as Area of Interest (AoI) and tiled into smaller quadratic images (tiles). Those tiles containing an intersecting overlap with a annotation were subsequently presented to a neural object detection network for training in a process known as Transfer Learning. A random portion of the dataset was kept aside from the training process in order to allow an unbiased evaluation of the detector performance.

Multiple iterations were performed in order to find out near-optimal input parameters such as tile size, zoom level, or network- and training-specific variables termed «hyperparameters» for each of the above-mentioned target objects. All detector models were evaluated for their prediction performance on the reserved test dataset. For each target object the best model was chosen by means of its overall performance measured by maximizing the F1-Score [4] on an independent reserved evaluation dataset. This model was used in turn to perform a prediction operation («Inference») on all tiles comprising the AoI – thereby detecting the target objects over the whole canton of Thurgau.

Postprocessing included filtering the resulting polygons by a high confidence score threshold provided by the detector for each detection in order to reduce the risk of false positive results (misidentification of an object as a silage bale stack). Subsequently adjacent polygons on separate tiles were merged by standard vector operations. A spatial intersection with the known LN layer was performed to identify the specific areas occupied by the objects which should not receive contributions but potentially did in last year's rolling payout. Only intersections covering more than 50m² of LN area are considered «relevant» for the final delivery. For completeness, all LN-intersecting polygons of detections covering at least 20m² are included in the final delivery. Filtering can be undertaken easily on the end user side by sorting the features with along a precalculated area column.

2.2 Aerial Imagery

The prototypical implementation uses the publicly available SWISSIMAGE dataset of the Swiss Federal Office of Topography swisstopo [15]. It was last flown for Thurgau in spring 2019 and offers a maximum spatial resolution of 10cm Ground Sampling Distance (GSD) at 3-year intervals. As the direct subsidies are paid out yearly the periodicity of SWISSIMAGE in theory is insufficient for annual use. The challenge of low aerial image frequency remains for manual and automatic methods alike. In this case the high-quality imagery on the one hand can serve as a proof of concept though. On the other hand, the cantons have the option to order own flight campaigns or satellite data to increase the periodicity of available aerial imagery if sufficient need can be shown from several relevant administrative stakeholders.

For our approach aerial images need to be downloaded as small quadratic subsamples of the orthomosaic called tiles to be used in the DL process. The used tiling grid system follows the "Slippy Map" standard [12] with an edge length of 256 pixels and a zoom level system which is derived from a quadratic division tree on a Mercator-projected world map. The whole world equals zoom level = 0 with a GSD at equator ~156 km/px, a zoom level = 18 in this system would approximate to a GSD of ~60 cm/px.

2.3 Dataset: Silage Bale Stacks

Silage hay bales are one of several features of interest specifically excluded from the subsidized cultivable LN area. These bales are processed, and compacted fermenting grass cuttings wrapped in plastic foil. They often roughly measure 1 - 2 cubic meters in volume and are weighed in at around 900kg. They are mainly used as animal food during winter when no fresh hay is available. Farmers are required by regulation to compactly stack them in regular piles at few locations rather than scattered collections consuming large areas.

As no conducive vector dataset for silage bale locations exists in Thurgau, the annotations for this use case had to be created manually. A specific labeling strategy to obtain such a dataset was therefore implemented (see Fig. 1). Using SWISSIMAGE as a WMS bound basemap in QGIS, a few rural areas throughout the canton of Thurgau were selected and initially approximately 200 stacks of silage bales were manually digitized as polygons. Clearly disjunct stacks were digitized as two separate polygons. For partially visible stacks only visible parts were included. Loose collections of bales were connected into one common polygon if the distances between the single bales were not exceeding the diameter of a single bale. Ground imprints where silage bales were previously stored were not included. Also shadows on the ground were not part of the polygon. Plastic membrane rests were not included unless they seemed to cover additional bales. Most bales were of circular shape with an approximate diameter of 1.2 – 1.5 m, but also smaller rectangular ones were common. Colors ranged from mostly white or green tinted over still common dark green or grey to also more exotic variants such as pink, light blue and yellow (the latter three are related to a specific cancer awareness program) [18].

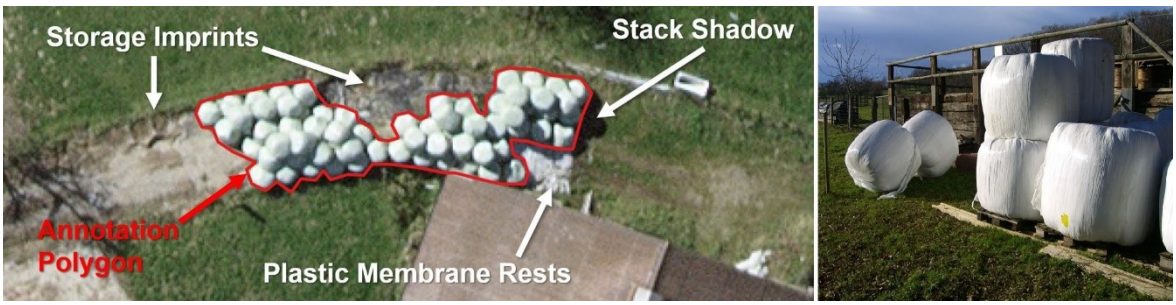

Fig. 1. Example of the annotation rules (left), example photo of Swiss silage bales (right)[16].

With these initial 200 annotations a preliminary detector was trained on a relatively high zoom level (18, 60cm GSD, tiling grid at about 150m) and predictions were generated over the whole cantonal area (See section «Training» for details). Subsequently, the 300 highest scoring new predictions (all above 99.5%) were checked visually in QGIS, precisely corrected, and then transferred into the training dataset. All tiles containing labels were checked visually again at full zoom and missing labels were created manually. The resulting annotation dataset consists of approximately 700 silage bale stacks.

2.4 Deep Learning

DL was performed with the Swiss Territorial Data Lab's Object Detection Framework [3]. The technology is based on a Mask RCNN architecture [6], an extension of Fast R-CNN [5], implemented with the High-Level API Detectron2 [17] leveraging the Deep Learning framework PyTorch [13]. Parallelization is achieved with CUDA-enabled GPUs on the High-Performance Computing cluster at the FHNW server facility in Muttenz. The Mask RCNN Backbone is formed by a 50 layer deep residual neural network (ResNet-50) [7] implementation and is accompanied by a Feature Pyramid Network (FPN) [8]. This combination of code elements results in a neural network leveraging more than 40 Mio. parameters. The model weights are obtained pretrained on the COCO dataset [9] and are modified through transfer learning. The model accepts three channel images and feature regions represented by pixel masks superimposing the imagery in the shape of the target object vector polygons.

Training is performed iteratively by presenting subsets of the tiled dataset to modify the edge weights in the network graph. Input images are not augmented as RGB aerial imagery relies on consistent northing and shadow angles. Progress is measured step-by-step through statistically minimizing the loss functions. The process is aborted if validation loss is not decreasing further after each 250 step iterations. Typically, less than 10 000 step iterations are sufficient to reach this point. Only tiles containing masks (labels) can be trained. Two

smaller subsets of all labeled tiles are reserved from the training set (TRN), so a total of 70% of the trainable tiles are presented to the network for loss minimization. The validation set (VAL, 15%) and the test set (TST, 15%) are pseudo-randomly distributed and statistically independent from the TRN set. The VAL set is used to perform recurrent evaluation during training. Training can be stopped if the loss function on the validation set has reached a minimum since after that point further training would push the model into an overfitting scenario. The TST set serves as an unbiased reserve to evaluate the detector performance on previously unseen data. Tiles not containing a label yet were classified into a separate class called "other" (OTH, see Fig. 2). This dataset was only used for generating predictions (inference).

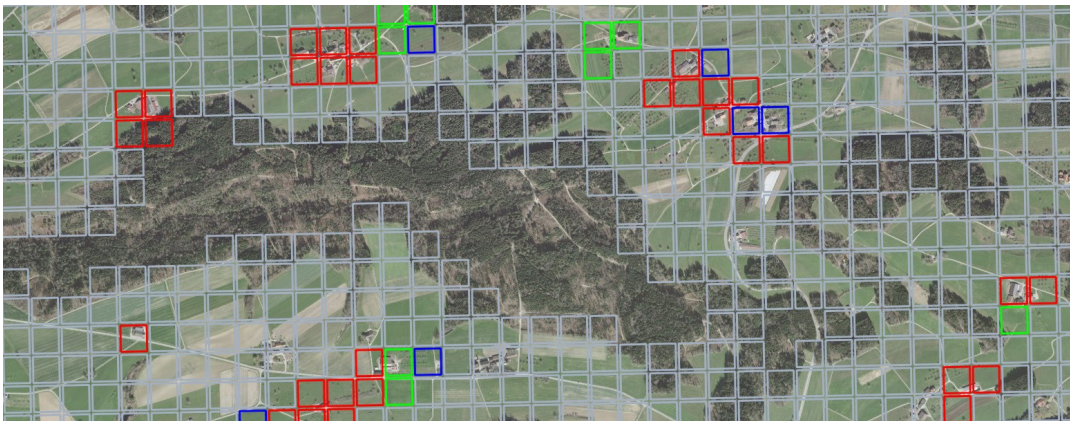

Fig. 2. Dataset Split – Grey tiles are only used in prediction (OTH); they do not contain any labels during training. The colourful tiles contain labels, but are scattered relatively sparsely. Red tiles are used for training the model weights (TRN); green tiles validate the learning progress during training to avoid overfitting (VAL) and blue tiles are reserved for unbiased post-training evaluation (TST).

Multiple training runs were performed separately to manually optimize the network-specific hyperparameters such as batch size, learning rate or momentum. Also, multiple zoom levels (spatial resolution, quadratic subdivision of tiles, see Fig. 2) were tested as a hyper-parameter variable in this manner. Learning rate was scheduled over the iterations using the "WarmupMultiStepLR" system.

2.5 Prediction and Assessment

For the TRN, VAL and TST subset, confusion matrix counts, and classification metrics calculations can be performed since they offer a comparison with the digitized «ground truth» reference. For all subsets (including the rest of the cantonal LN as OTH), predictions are generated as vectors covering those areas of a tile that the detector algorithm identifies as target objects and therefore a confidence score is attributed. In case of the tiles containing annotation polygons, the overlap between the predictions and the labels can be checked. Is any overlap found between a label and a prediction this detection is considered a true positive (TP). If the detector missed a label entirely this label can be considered as false negative (FN). Did the detector predict a target object that was not present in the labelled data it is considered false positive (FP). On the unlabeled OTH tiles, inference predictions cannot be checked against reference data.

The counting of TPs, FPs and FNs on the TST subset allows the calculation of standard metrics such as precision (user accuracy), recall (producer accuracy) and F1 score (as a common overall performance metric calculated as the harmonic mean of precision and recall) [4]. The counts, as well as the metrics can be plotted as function of the minimum confidence score threshold which can be set to an acceptable filter percentage for a certain detection task. A low threshold should generally yield fewer FN errors, while a high threshold should yield fewer FP detections. The best performing model by means of maximum F1 score was used to perform a prediction run over all tiles intersecting with the cantonal LN surface area.

2.6 Post-Processing

In order to obtain a consistent result dataset, detections need to be postprocessed. Firstly, the confidence score threshold operation is applied. Here, a comparatively high threshold can be used for this operation. «Missing» a detection of a target object (FN) is not as costly for the analysis of the resulting dataset at the agricultural office as analyzing large numbers of FP detections would be. Also missing single individual small target objects is much less problematic than missing whole large areas. These larger areas are typically attributed with higher confidence scores though and are therefore less likely to be missed.

In some cases, silage bale stacks can cross the tiling grid and are therefore detected on multiple images. This results in edge artefacts along the tile boundaries intersecting detections that should be unified. For this reason, adjacent polygons need to be merged into a single polygon. This is achieved by first buffering all detections with a 1.5m radius (roughly the radius of a single typical bale). Then all touching polygons are dissolved into single features. Afterwards, negative buffering with -1.5m radius is applied to restore the original boundary (see Fig. 3).

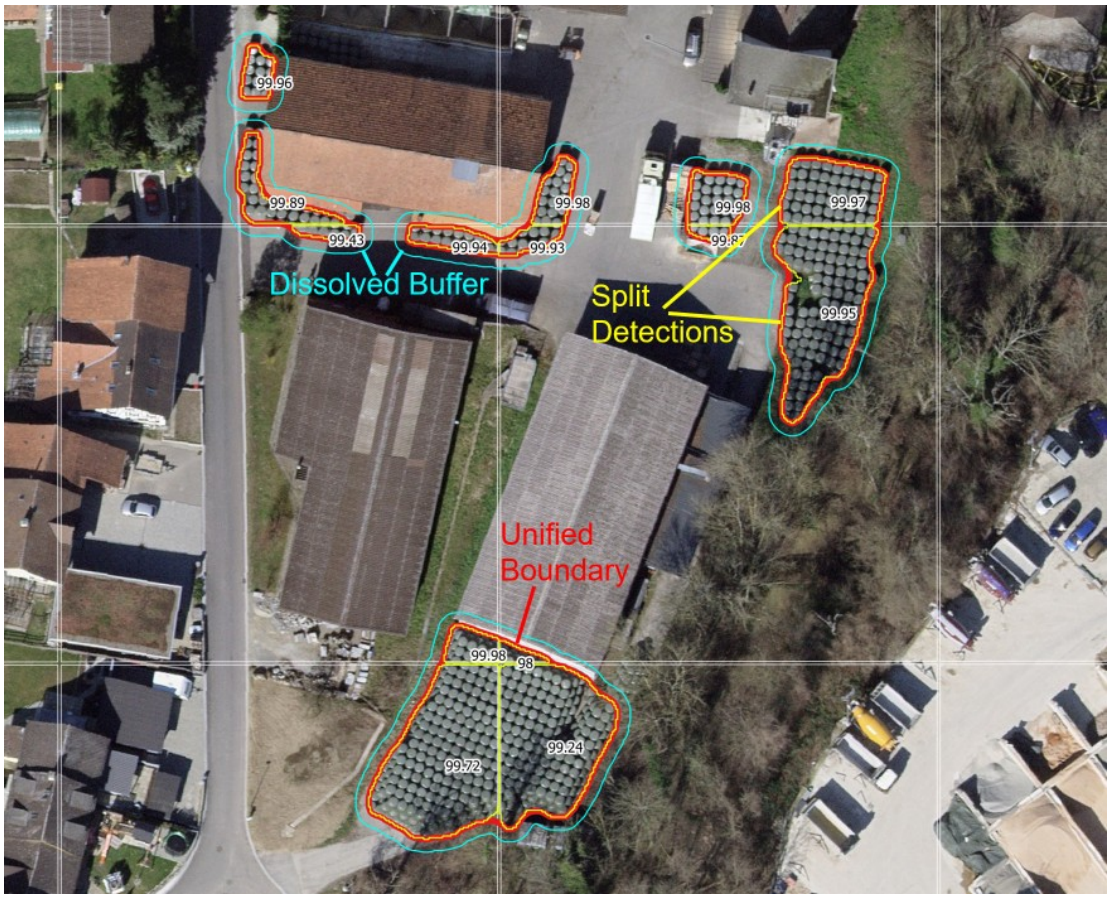

Fig. 3: Example of silage bale detection polygons (red) from raw detections (yellow) dissolved because they are crossing the tile boundary (light blue).

This process also leads to an edge smoothing of the pixel step derived vector boundary into curves containing a high number of vertices. A simplification operation reducing the number of vertices can be performed without the loss of relevant spatial accuracy. For all remaining detection polygons, the confidence score is reattributed as a merged area-weighted average of the input values. With a threshold operation on the resulting area all target objects with an area cover below 20 m^2 are filtered out of the dataset to provide only economically relevant detections.

3 Results

Silage Bale Stacks as a target object generally resulted in successful and robust models achieving high prediction performance. The detections were considered deliverable to the LWA.

Tab 1: Performance of silage bale detector models at several zoom levels evaluated by maximum F1-Score.

	Zoom Level 16	Zoom Level 17	Zoom Level 18	Zoom Level 19	Zoom Level 20
GSD	~ 240 cm/px	~ 120 cm/px	~ 60 cm/px	~ 30 cm/px	~ 15 cm/px
# Tiles Trained	~ 600	~ 1 000	~ 1 600	~ 3 000	~ 5 000
# Tiles Inference	~ 8 000	~ 25 000	~ 84 000	~ 310 000	~ 1 310 000
Duration of Run	~ 0.6 h	~ 2 h	~ 4 h	~ 15 h	~ 100 h
TST Max F1 RGB	52.5 %	74.7 %	87.2 %	92.3 %	90.9 %

The model trained with tiles at zoom level 19 (every pixel approx. 30cm GSD) showed the highest performance with a maximum F1 Score of 92.3% (see Tab. 1). Increasing the resolution even further by using 15 cm/px GSD did not result in a gain in overall detection performance while drastically increasing storage needs and computational load. The detector model at zoom level 19 is performing very well on the independent TST dataset detecting the largest portion of silage bale stacks at any given confidence threshold. The number of FP reaches very low counts towards the higher end of the threshold percentage, increasing precision while decreasing recall (see Fig. 4).

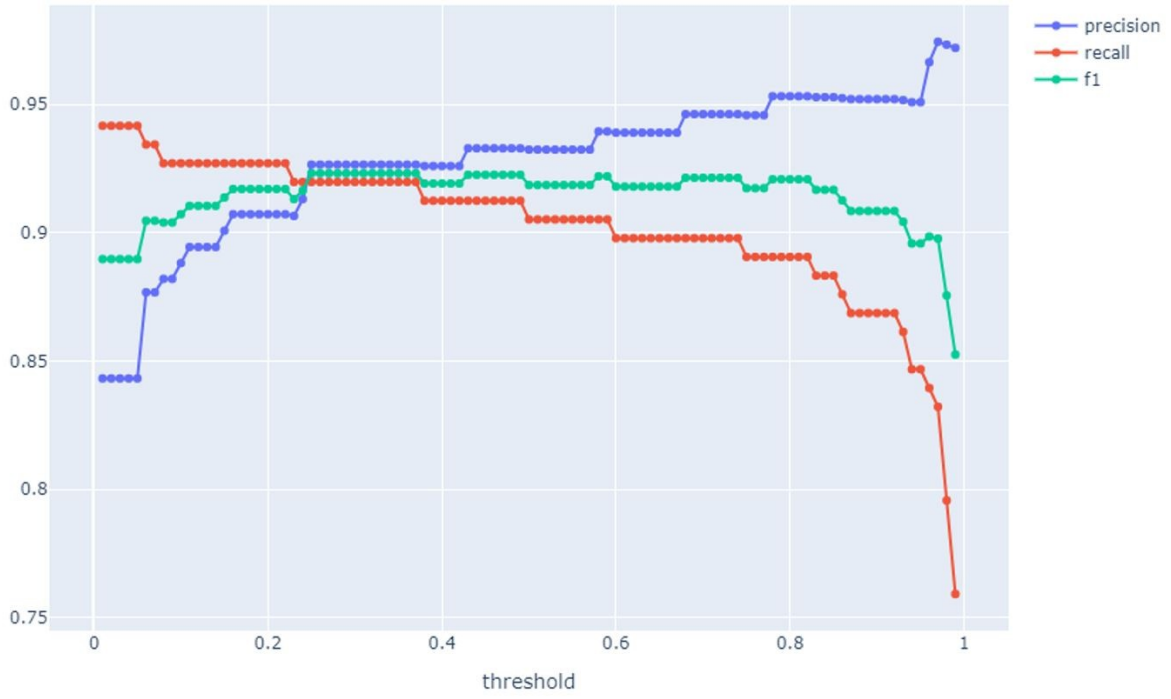

Fig. 4: Performance metrics of the Zoom Level 19 model on the TST dataset as a function of the minimum confidence score threshold.

For delivery of the dataset a detector was subsequently used at a threshold of 96% minimizing FP errors resulting in a conscious bias on precision, see Fig. 4 and 5. At this value 809 silage bale stacks were rediscovered in the TRN, TST and VAL subset. Just 10 FP detections were found in these subsets. 97 silage bale stacks were not rediscovered (FN). The model precision (user accuracy) on the TST set was found to reach approx. 98% and the recall (hit rate, producer accuracy) was acceptable at approx. 85%.

In the applied inference run the model detected a total of 2 473 additional silage bale stacks after post-processing over the rest of the LN area of the canton of Thurgau (OTH subset), of which 288 stacks cover more than 20 m^2 and were prepared for delivery. The relevant total intersection area of the final vector polygon dataset with the LN layer amounts to approx. 8 000 m^2.

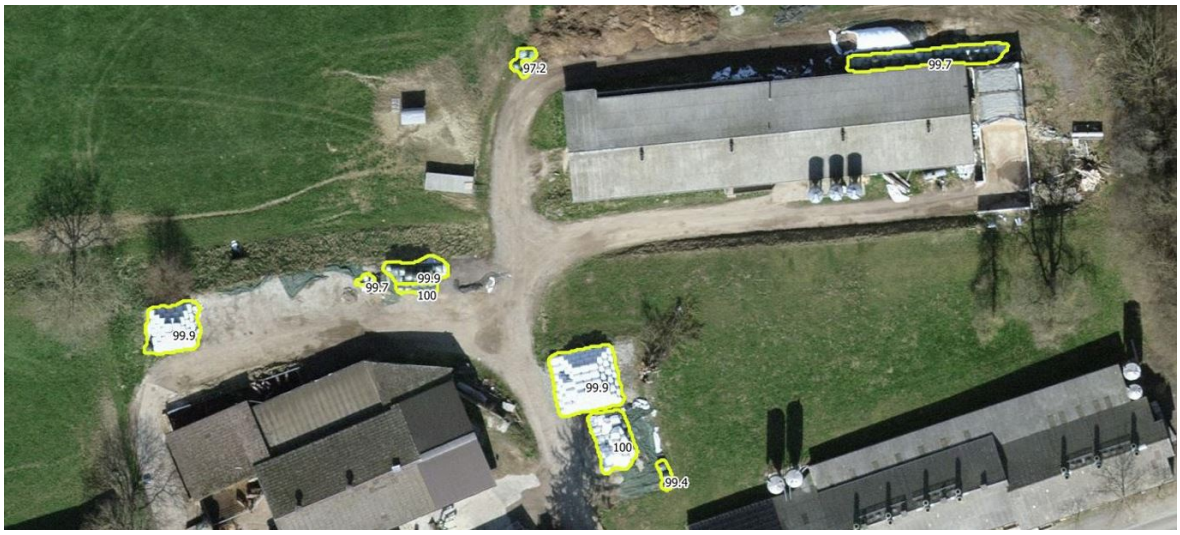

Fig. 5: Raw inference detections (yellow) of silage bale stacks displaying very high confidence scores outside of the TRN/VAL/TST subsets.

4 Conclusion

The agricultural office describes the detections of silage bales as very accurate with only a small percentage of actual FP detections. Clearly delineated objects such as silage bales are generally less demanding to detect than more complex target objects. The high F1 score surpassing 90% suggests a productively usable result. Especially larger stacks are detected with very high confidence scores and can be targeted first by area-filtering in the monitoring process. The Mask-RCNN approach proved to be a viable deep learning kernel architecture.

The highest zoom level 20 (15cm GSD) requires enormous computational resources especially for the prediction run while performing suboptimal on evaluation metrics. Hence, RGB winter imagery resampled from SWISSIMAGE at a resolution of 30cm GSD proved to be sufficient in resolution and quality while still maintaining a reasonable effort on computational resources for inference runs over the complete cantonal agricultural surface.

Very few false positive samples such as animal shelters, material deposits or white-colored vehicles remained in the final prediction dataset. Options to automatically tackle this challenge in the future include new models distinguishing multiple classes, the choice of larger (higher parametric) model architectures, larger training datasets or a revised and improved post-processing workflow.

On an economical scale the extra effort for the LWA resulting from misplaced silage bale stacks in the LN areas is not negligible but also not extremely critical. In the scope of this study, silage bale stacks did serve as an accessible initial proof of concept regarding the usability of the detector. The new detections allow the professionals at the agricultural office to direct their eyes more quickly at relevant hotspots and spare them some aspects of the long and tedious manual search on aerial imagery that was performed in the past.

For the future, extending the range of target objects to larger and more complex areas such as complete farm yards or land usage patterns such as grazed pastures on steep slopes could provide strong additional benefits for the monitoring process at the agricultural office.

5 Acknowledgments

We want to thank the cantonal agricultural office of Thurgau and the team at the Swiss Territorial Data Lab for the detailed review of our results. Especially we would like to show our immense gratitude to Alessandro Cerioni at the Système d'Information du Territoire à Genève (SITG) who contributed large efficient and powerful sections of code to the object detection framework used in this study. Furthermore, we are very grateful to Pascal Salathé (FHNW) for his input on the Swiss direct subsidy system and to Natalie Lack (FHNW) for her internal review of the article.

References

1. Amt für Geoinformation des Kanton Thurgau: Handbuch Amtliche Vermessung. Kanton Thurgau (2022) 25-26, 158-171, 185-188
2. Bundesamt für Landwirtschaft BLW: Direktzahlungen an Schweizer Ganzjahresbetriebe. BLW, Bern (2021)
3. Cerioni, A. & Meyer, A.: Object Detection Framework. Swiss Territorial Data Lab (2021)
4. Chinchor, N.: MUC-4 Evaluation Metrics, Proceedings of the Fourth Message Understanding Conference (1992) 22-29
5. Girshick, R.: Fast r-cnn. Proceedings of the IEEE international conference on computer vision (2015) 1440-1448
6. He, K., Gkioxari, G., Dollár, P., & Girshick, R.: Mask r-cnn. Proceedings of the IEEE international conference on computer vision (2017) 2961-2969
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas (2016) 770-778
8. Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection. Computer Vision and Pattern Recognition (2016)
9. Lin, T., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick. C.L., Dollár, P.: Microsoft COCO: Common Objects in Context, Computer Vision and Pattern Recognition (2014)
10. Meier, T. & Menzel, S.: Bundesamt für Landwirtschaft BLW, Agrarbericht 2020 – Politik, Einführung (2020) 2-4
11. Meyer, D.: Bundesamt für Landwirtschaft BLW, Agrarbericht 2020 – Politik, Direktzahlungen (2020) 11-12
12. OpenStreetMap Foundation: Slippy Map. https://wiki.openstreetmap.org/wiki/Slippy_Map (2021)
13. Paszke, A., Gross, S., Chintala, S., Chanan, G.: PyTorch. Facebook AI Research / Meta AI (2016)
14. QGIS Association: QGIS Geographic Information System. https://qgis.org/en/site/ (2021)
15. Swisstopo: «SWISSIMAGE 10 cm», The Digital Color Orthophotomosaic of Switzerland. Federal Office of Topography swisstopo (2021)
16. Tschudin, P.: Photo Silageballen; https://www.flickr.com/photos/35637563@N00/80239710/ (2006)
17. Wu, Y., Kirillov, A., Massa, F., Lo, W. Y., & Girshick, R.: Detectron2. Facebook AI Research / Meta AI (2019)
18. Zindel, N.: Siloballen-Aktion für den guten Zweck. Pink Ribbon Schweiz (2018)