A DEEP LEARNING ARCHITECTURE FOR BATCH-MODE FULLY AUTOMATED FIELD BOUNDARY DETECTION

L. Meyer[1,*], F. Lemarchand[1], P. Sidiropoulos[1,2]

[1]Hummingbird Technologies Ltd - (lucas, francois, panos)@hummingbirdtech.com
[2]University College London

KEY WORDS: field boundary detection, image segmentation, image instantiation, deep learning, earth observation

ABSTRACT:

The accurate split of large areas of land into discrete fields is a crucial step for several agriculture-related remote sensing pipelines. This work aims to fully automate this tedious and resource-demanding process using a state-of-the-art deep learning algorithm with only a single Sentinel-2 image as input. The Mask R-CNN, which has forged its success upon instance segmentation for objects from everyday life, is adapted for the field boundary detection problem. Such model automatically generates closed geometries without any heavy post-processing. When tested with satellite imagery from Denmark, this tailored model correctly predicts field boundaries with an overall accuracy of 0.79. Besides, it demonstrates a robust knowledge generalisation with positive results over different geographies, as it gets an overall accuracy of 0.71 when used over areas in France.

1. INTRODUCTION

An accurate knowledge of field boundaries is a requirement for many actors in agriculture. Amongst many applications, it is a prerequisite input for farmers to on-board fields on farm management software services, it improves the accuracy of crop type classification (Peña-Barragán et al., 2011, De Wit, Clevers, 2004), and it is used from government agencies to monitor subsidies and farming practices.

Typically, the collection of these geographical data is obtained by manual labelling of aerial or satellite imagery. This slow, repetitive, and error-prone acquisition hinders scalability. It prevents the batch-mode boundary delineation in large areas. Consequently, the scientific community has been exploring solutions to accurately and reliably generate field boundaries in a large-scale manner, without intensive user involvement.

The first challenge in the attempt to automate field boundary detection is the inherent subjectivity of their definition. For example, the Land Parcel Identification System (LPIS)[1] lists four different parcel types, with corresponding types of field boundaries. An automated approach is by default exposed to the more or less arbitrary selection of the field boundary definition that it follows.

Despite the definition limitations, field boundary detection has been investigated for several decades. Early automated field boundary detection techniques relying on some form of edge detection through the use of traditional computer vision (Rydberg, Borgefors, 2001, Yan, Roy, 2014). Lately this domain has benefited from the proliferation of deep learning (Persello et al., 2019, Waldner, Diakogiannis, 2019). In spite of the increased accuracy, these techniques still suffer from challenges on generating a single closed polygon for each field (instead of incomplete and noisy curves), large computational cost and lack of generalisation. As a matter of fact, the sparse collections of

---

* Corresponding author
[1] The Land Parcel Identification System, Special report No 25, 2016 `https://op.europa.eu/en/publication-detail/-/ publication/11049e0e-9a82-11e6-9bca-01aa75ed71a1`

automatically-generated field boundary sets are often limited to a single geography, and involve post-processing to remove false omission and commission errors.

This work introduces the first step towards the systematic processing of satellite imagery for batch-mode field boundary detection. This is mainly achieved by transferring the state-of-the-art instance segmentation algorithm Mask R-CNN to this domain of knowledge (He et al., 2017). This task requires the careful tuning of the architecture hyper-parameters as well as adjustments and modifications that increase its accuracy. Additionally, a novel tailored measure for field boundary detection evaluation is suggested. The experimental setup for such an approach includes large volumes of data from multiple geographies. Please note that the Rydberg and Borgerfors's (Rydberg, Borgefors, 2001) field boundary definition is followed in this work. The latter defines field boundaries as changes of crop types or discontinuity of natural features.

The rest of the article is structured as follows. Section 2 gives an overview of the existing approaches to delineate accurately field boundaries. Subsequently, section 3 discusses in detail the suggested architecture used, before experimental results confirm the validity of this approach (section 4). Section 5 concludes this work.

2. RELATED WORK

The rich literature of field boundary detection algorithms can be mainly categorised between traditional computer vision techniques and machine learning approaches.

The first algorithms investigating field boundary detection were typically build upon some form of edge detection. The main hypothesis is that the transition between fields would be characterised by sharp changes in pixel values (Ji, 1996, Rydberg, Borgefors, 2001), hence, field boundaries would be a subset of image edges. Edge detection commonly involves the computation of multi-directional gradient though kernel convolutions (North et al., 2019, Graesser, Ramankutty, 2017). Identifying

image edges that actually represent field boundaries requires the use of region-based knowledge (Yan, Roy, 2014). For instance, North et al. compute the standard deviation for each pixel within a window moving across the image channel (North et al., 2019). Similarly, Graesser and Ramankutty consider small tiles from a satellite image to normalise the gradients locally. For each tile, an adaptive threshold is set to extract the boundaries (Graesser, Ramankutty, 2017).

Both convolution operations and region-based information, which have been widely used as part of edge detection techniques, are a common feature of deep learning architectures. Deep learning application to various remote sensing topics has already proved to be successful (Zhang et al., 2016, Zhu et al., 2017). Lately, deep learning techniques (or hybrid techniques combining deep learning with edge detection) for field boundary detection have been published. Crommelinck et al. introduced a hybrid method (Crommelinck et al., 2019) in which candidate pixels were identified before a convolution neural network classifies tiles, centered on these candidates, that actually contain boundaries.

Semantic segmentation approaches, in which each pixel is assigned a label depending on whether it belongs or not to to a boundary, offer better granularity and may be directly used without an edge detection stage. (Masoud et al., 2020, Persello et al., 2019). Recently, U-Net (Ronneberger et al., 2015) architectures have become popular in field boundary detection. Gracia-Pedrero et al. employ a U-Net architecture to segment images into three classes: field, buffered boundary, and background. The boundaries are then computed from the contour of the first class (García-Pedrero et al., 2019). Likewise, in (Waldner, Diakogiannis, 2019), Waldner and Diakogiannis adapt a U-Net model to generate not only segmented images, but also predict distances to the field boundaries. In a post-processing step, they use a watershed algorithm to increase the algorithm accuracy (Beucher, Meyer, 1990).

Post-processing semantic segmentation results using computer vision (commonly, some geometric rules or watershed algorithms) is not uncommon in the relevant literature, because most of the introduced architectures generate intermediate results that require merging sub-fields in one field, splitting larger areas into fields, or both. This process hinders the scalability of such a solution, since it often requires tuning ad-hoc parameters in a case-by-case basis. Besides, accurate post-processing is far from being trivial, especially if the intermediate output presents a disconnected set of predicted boundary pixels (i.e. not adjacent pixels).

Moreover, many techniques make use of time-series images (North et al., 2019, Graesser, Ramankutty, 2017). However, working with time-series of satellite imagery introduces the significant challenge of cloud coverage. A fully automated time-series-based pipeline should include a module for identifying and removing clouds, as well as replacing their values commonly through interpolation. This additional complexity has been found to not lead to a substantially increased accuracy (Waldner, Diakogiannis, 2019), while also impeding the solution scalability.

Finally, some works recently make use of very high resolution data acquired from unmanned aerial vehicles (Persello, Bruzzone, 2009, Crommelinck et al., 2019). It is straightforward that a resolution in the order of hundreds finer than the one achieved from satellites would increase the potential of field boundary

detection techniques. However, this comes with a cost in availability and operations, which makes such approaches suitable only for small-scale applications.

In this work we present a technique which is envisaged as the first step towards a systematic field boundary detection pipeline. This technique, designed to remove challenges related to scalability as well as to reduce the ad-hoc parameters that require manual tuning, is described in detail in the next section.

3. DEEP LEARNING ARCHITECTURE FOR FIELD BOUNDARY DETECTION

Fig. 1 presents the workflow of the introduced technique, which is based on Mask R-CNN (He et al., 2017). Mask R-CNN is a deep learning architecture, which, has recently achieved exceptional performance in several instance segmentation setups. However, Mask R-CNN has been introduced in a totally different context, using images from images that are not relevant to Earth Observation (e.g. COCO dataset (He et al., 2017), which aggregates a large amount of common objects from everyday scenes (Lin et al., 2014)). Our hypothesis is that transferring this technique on field boundary detection requires a number of adjustments in several parts of the pipeline. In the rest of the section, we describe the model architecture that needs to be implemented with bespoke adjustments for the field boundary detection problem.

3.1 Data curation and pre-processing

The large-scale labelled dataset required for the training of our model can be sourced from several agricultural existing parcel registers, which are commonly maintained by governmental agencies in the form of annual records. However, these data have limited accuracy and would have an adverse effect in the algorithm prediction quality if used without denoising. Existing problems include (1) erroneous entries caused by inaccurate semi-automatic approaches used for their creation (2) difference between the field boundary definition used from us and the dataset (3) corrections made along the year to the initial field geometry which cause overlapping field boundaries or duplicate instances.

As a result, the first step of the training is cleaning the dataset. Apart from trimming off parcels, irrelevant geometries are removed using the Schwartzberg compactness score (Schwartzberg, 1965). More specifically, after enforcing non-overlapping fields, we discard any entry that is smaller than 1.5 ha and whose Schwartzberg compactness score is lower than 0.15. Schwartzberg compactness score expresses the ratio between the perimeter P of the field and the circumference of a circle that would have same area $\mathcal{A}$ (Eq. 1).

$$S = 2\frac{\sqrt{\mathcal{A}\pi}}{P} \qquad (1)$$

Subsequently, the labelled dataset is created by matching the ground truth with available satellite imagery. In this work Sentinel-2 is used, but, it should be noted that the proposed method is satellite-agnostic, with the only limitation being that the satellite includes a Near-Infrared (NIR) band. NIR is used to generate a 4-band input. The first three bands are the red, green, and blue of the true color image. The fourth band corresponds to the NDVI computed from the red and near-infrared bands

(a) Raw data

(b) Inputs

(c) Model architecture

(d) Prediction output
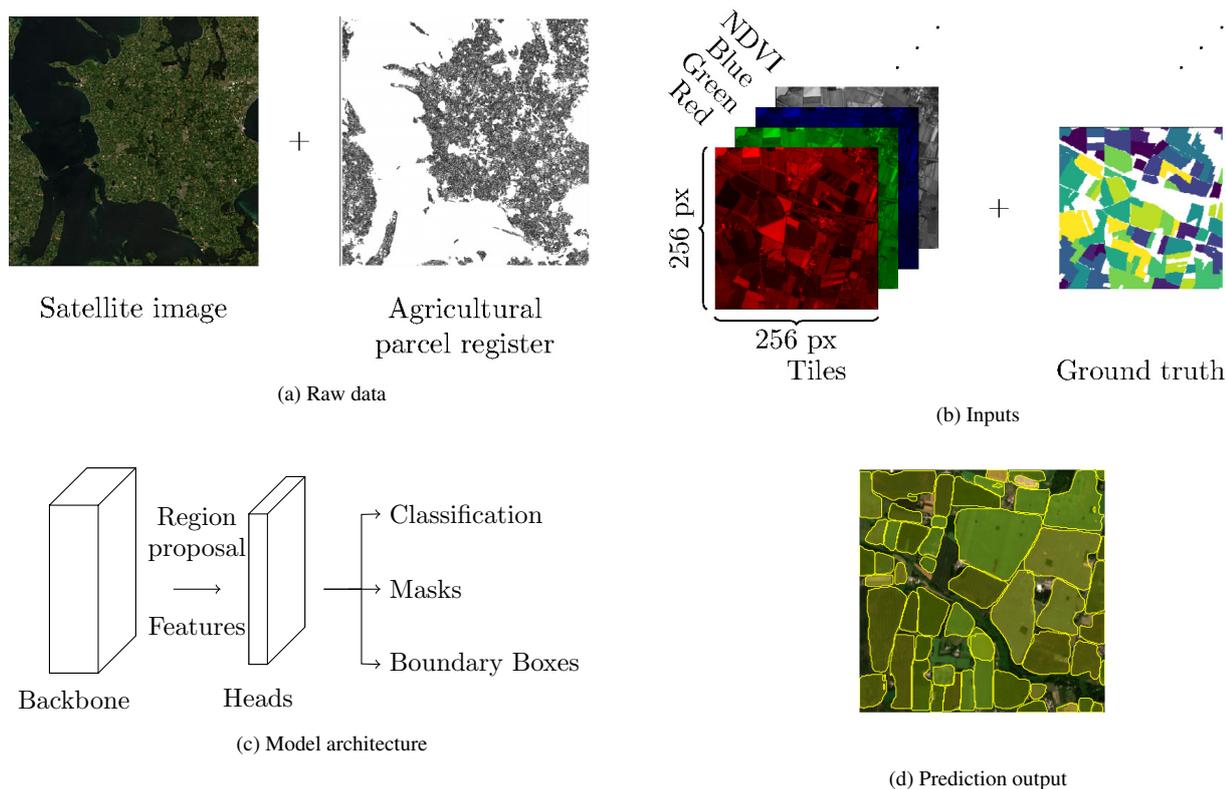
Figure 1. Training of the Mask R-CNN to predict field boundaries. From Sentinel-2 imagery and parcel register 1a, a dataset is generated 1b. The model 1c is then trained to predict field boundaries in new area 1d.

(Eq. 2). NDVI is used because of its high information value in agriculture applications, as well as its non-linearity. Being a non-linear combination of two spectral bands, it brings additional information that a deep learning network could struggle to learn from the bands separately.

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (2)$$

In a second step, each band is standardised by subtracting the mean value and dividing by the standard deviation, before the satellite imagery is split in 256 x 256 pixels tiles. Each tile is matched to the ground truth while the pre-processing ends with the generation of corresponding binary masks. Figure 1b illustrates the outcome of these pre-processing steps.

3.2 Mask R-CNN

This section summarises the architecture of Mask R-CNN. For more information, the reader is referred to the original publication (He et al., 2017).

In general, Mask R-CNN is a model that is designed to identify and classify areas of interest that belong to one or more object classes (Fig. 1c).

In Mask R-CNN, the backbone, which is based on traditional convolutional neural networks like ResNet (He et al., 2016), generates features from the image. These features, computed through convolutional operations, may be understood as primitive representations of visual concepts, like shapes or edges. At the end of the backbone lie two parallel branches. The first

branch, called region proposal network (RPN), draws areas of interest that may contain a relevant object, in our case, a field. These areas of interest take rectangular shapes whose dimensions are set as parameters of the model. The second branch, extract the previous features within these candidate areas. The selected features are then passed to the heads stage.

The heads stage, which consist of fully convolutional networks smaller than the backbone, refine and classify each area of interest. This stage generates binary masks for each possible object class. In the case of field boundary detection only one object class is examined (fields), therefore this single segmentation mask defines the estimated field boundaries.

All parts of the network are trained together through backpropagation (He et al., 2017). The loss of the model is a linear combination of 3 losses. These quantify (a) classification accuracy, i.e. assigning the correct class to an object. In the class set the null class (corresponding to the background) is also included. (b) segmentation accuracy, i.e. identifying the correct contour of an object and (c) instantiation accuracy, i.e. estimating the correct boundary box framing each object in an image.

In this work, we use the Matterport Mask R-CNN implementation (Abdulla, 2017), with a ResNet 101 as backbone.

3.3 Model Adjustment for Field Boundary

The default implementation of Mask R-CNN has been developed for use cases that significantly differs from field boundary detection using satellite imagery. In order to adjust Mask R-CNN for field boundary detection we have carefully re-examined the tuning of its hyperparameters. Two main issues have been found with the default Mask R-CNN.

Firstly, the number of areas of interest generated from the RPN network is too low for field boundary detection. The default value of 100 areas is smaller than many images in field boundary detection datasets. Therefore, we have increased this value to 200 areas of interest.

Secondly, and perhaps most importantly, fields exhibit a large variation of sizes and shapes. For example, pedestrians in a surveillance setup (a typical use of Mask R-CNN) are expected to have medium variation in their size and even smaller in their shape. On the other hand, (a) satellite images may include fields that vary from 1.5 ha to hundreds of ha (b) the range of field shapes is even larger since they include very elongated rectangles, square fields, circular fields, multi-line polygons, etc. Therefore, (a) we have modified the possible side size of candidate regions to 8, 16, 32, 64, or 128 pixels (default set is $\{32, 64, 128, 256, 512\}$) and (b) we have augmented the ratio between width and height of the bounding boxes to $\{0.1, 0.5, 1, 2, 4\}$ (default set is $\{0.5, 1, 2\}$).

The training is achieved with batches of 4 images using an NVIDIA TITAN RTX GPU with 24GB of dedicated memory.

3.4 Post-processing

As already mentioned, one the benefits of this architecture is that it does only require a trivial post-processing. By default the architecture produces closed polygon masks within the boundary box of the predictions, which can be straightforwardly used to extract each field boundary by estimating the contour of the output mask. The bottom-right panel of Fig. 1d shows an example of the final output. Vectorised polygons can also be obtained by reprojection of the predicted geometries.

4. EVALUATION

In this section the validity of our main assumptions is tested. Apart from evaluating the adjusted pipeline accuracy, comparing it with the default one, a second goal is to examine its generalisation capability. A technique that aims to systematise field boundary detection should be transferable across different geographies. For that reason, we have included in our dataset two agricultural areas, from Denmark and France, respectively.

After describing the study areas and the associated datasets, we present the different measures to assess the prediction accuracy. This includes adapting the precision and recall measures on this problem. Subsequently we conduct the core evaluation separately for each area. Finally, we examine the generalisation capability of the model, by evaluating over an area the performance of the model trained over a different one.

4.1 Study Area and Materials

Within the LPIS framework, most member states of the European Union makes publicly available datasets of agricultural parcels, which inform on their crop types and geometries. This dataset has been used before in the relevant literature (García-Pedrero et al., 2019). This work also used this source, more specifically, the French and Danish datasets for the year 2018[2] and 2019[3] respectively. In the case of Denmark most of the

dataset was used, while for France the data were reduced to areas with high agricultural production. The total ground truth set consisted of 250,126 fields in Denmark and 395,969 fields in France.

As explained in the previous section, labelling is conducted through Sentinel-2 imagery, which was downloaded from the Copernicus Access Hub[4]. Apart from downloading imagery of the relevant year, a global cloud coverage lower than 1% was imposed to reduce cloud artefacts. Additionally, while we select only one satellite scene to cover one area for the dataset, these span a large period (Fig. 2) in order to provide to the dataset a richer variety of field aspects.

These criteria result in a selection of 4 Sentinel-2 images over Denmark and 3 over France. Each image is 10980 x 10980 pixels of 10 meters resolution. Figure 2 gives the zone and dates of the selected satellite imagery.

(a) Denmark   (b) France
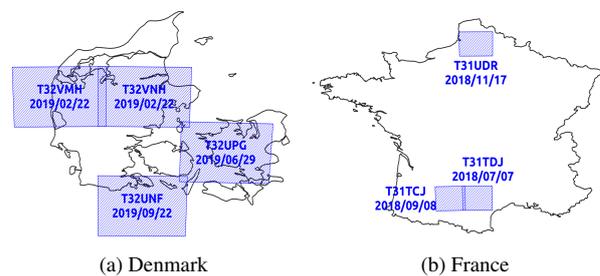
Figure 2. Selected satellite scenes over country borders

Following the pre-processing described in section 3.1, we generate for each country a dataset of images of shape 256 x 256 x 4 and their corresponding ground truth. The data were finally split in training, test, and validation sets accounting for respectively 80%, 10% and 10% of the images. For such configuration, using an NVIDIA TITAN RTX GPU with 24GB of dedicated memory, the training over 10 epochs takes about 3.5 hours, while inference takes less than a few seconds per image.

4.2 Evaluation Measures

The accuracy of the predictions is assessed via several metrics introduced in (Persello, Bruzzone, 2009). The overall accuracy gives an estimation of how well the pixels are classified. It is computed following equation 3, where TP_{px}, TN_{px}, FP_{px}, and FN_{px} are respectively the true positive, true negative, false positive, and false negative rates for the pixel-wise classification (boundary or background).

$$OA = \frac{TP_{px} + TN_{px}}{TP_{px} + FP_{px} + TN_{px} + FN_{px}} \quad (3)$$

It is important to highlight that this measurement does not convey how well the boundary is outlined, since (a) it punishes equally a false positive close to the boundary with a false positive in the center of the field (b) it punishes equally a false negative in a small field, which may reduce how distinct the field is, with a false negative in a larger field with small effect to the result quality.

For this reason, we are suggesting a new measure, which redefines the concept of true or false positives and negatives. In

---

[2] https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#rpg

[3] https://kortdata.fvm.dk/download/Markblokke_Marker?page=MarkblokkeGaeldende

[4] https://scihub.copernicus.eu/dhus/#/home

this measure, for each field of the ground truth F_{gt}, the predicted field F_p that has the biggest overlap is estimated. The pixels issued from this intersection are considered as true positive. The remaining pixels of F_{gt} are counted as false negative, because they have not been detected as belonging to F_p (even if they overlap with another field F'_p in the predicted mask). The remaining pixels of F_p are counted as false positives, because they do not fall within F_{gt}. If the list of ground truth fields is exhausted, all remaining pixels in the predicted fields are also counted as false positives. Conversely, if the list of predicted fields is exhausted first, all remaining pixels in the ground truth are then considered as false negative.

This algorithm estimates the TP_f, FP_f and FN_f rates, from which recall, precision and f1-score can be defined. The f1-score is defined by the equation 4.

$$f1 = \frac{2\,TP_f}{2\,TP_f + FP_f + FN_f} \tag{4}$$

In general, this definition is more strict than the commonly used accuracy measures, since it requires an one-to-one correspondence. To have insight of the distinct type of field-specific errors (over-segmentation and under-segmentation errors), we are also computing the fragmentation error (e_{fg}) and the under-segmentation error (e_{us}) defined respectively by equations 5 and 6. $\mathcal{GT}$ and $\mathcal{P}$ represent the sets of ground truth and predicted masks respectively . $|\cdot|$ denotes the cardinality of a set, while $\mathcal{A}$ the area in pixel of a given mask. F_{p*} corresponds to the prediction mask that has the largest overlap with a given ground truth. The fragmentation error is preferred over the over-segmentation error defined in (Persello, Bruzzone, 2009) as it accounts for all overlapping fields.

$$e_{\text{fg}} = \frac{1}{|\mathcal{GT}|} \sum_{F_{gt} \in \mathcal{GT}} \frac{|\{F_p \in \mathcal{P}, F_p \cap F_{gt} \neq \varnothing\}| - 1}{\mathcal{A}_{F_{gt}} - 1} \tag{5}$$

$$e_{\text{us}} = 1 - \frac{1}{|\mathcal{GT}|} \sum_{F_{gt} \in \mathcal{GT}} \frac{\mathcal{A}_{F_{gt} \cap F_{p*}}}{\mathcal{A}_{F_{p*}}} \tag{6}$$

4.3 Validation of the Tailored Configuration of the Mask R-CNN

In order to verify that the tailored adjustments made to the default implementation (section 3.3) actually improve the predictions of field boundaries, we compare the predictions of the two different configurations on the Danish and French datasets, respectively. Figure 3 shows the evolution of the loss for the training and test datasets for both geographies. The test loss seems to converge with no evidence of over-fitting. The lower values of the loss indicate that the adjustments suggested in this work improve the accuracy of Mask R-CNN in a field boundary detection context.

Moreover, table 1 confirms the superiority of the suggested configuration in the prediction of field boundaries. For the Danish dataset all accuracy measurements are improved with the bespoke configuration. f1-score is raised by 7.3 percent. For the French dataset, there is a decrease in the precision value with a corresponding improvement of the recall which results in a 5.2 percent increase of the f1-score.
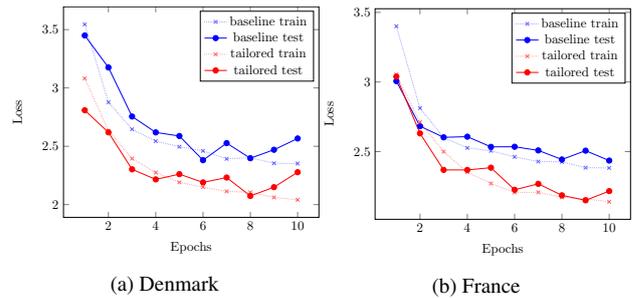
Figure 3. Convergence of the loss for the train and test datasets of each country during the training of the Mask R-CNN with the default and tailored configurations.

It is worth to note that for both Mask R-CNN variations the fragmentation error remains almost zero. This may be a result of the Mask R-CNN architecture, which typically merges adjacent areas of interest that belong to the same class. Moreover, over-segmentation only occurs for ground truth fields of significantly large area. For such fields, the model might predict few instances whose number is however negligible compared to the size of the reference field, hence a steadily low fragmentation error.

4.4 Cross Area Prediction

The bespoke Mask R-CNN has showed evidence of improved predictions in the same geography where it has been trained. Table 2 shows additionally that the bespoke model can perform well on areas it has never seen before. The model trained with Denmark dataset performs better at predicting French fields than the default model trained directly with the French dataset. This supports the concept of training the model over a geography with a reliable dataset to predict field boundaries over other geographies with similar farming practices. On the other hand, the model trained with the French dataset does not perform as well on Danish dataset. This may indicate a difference in the numbers of errors in the two ground truth datatets.

Further evidence for the last point is given from the surprising result of increased overall accuracy of the model transferred from Denmark to France in comparison to the model which was trained with the same dataset. Also, this counter-intuitive result may be partially explained from the fact that French dataset was noisier than the Danish one. Fig. 4j shows few parcels from the France dataset that clearly appear to be fields (correctly predicted by the model), while they are not labelled in the ground truth. This is promising for the additional use of such a technique for identifying errors in commonly available large-scale datasets. Besides, the French dataset contain much more fields than the Danish one. This greater number of fields may present a diversity that makes it more difficult for the model to learn. A longer number of epochs might be necessary for the model to train on the French dataset in order to achieve similar performances with the Danish dataset, for which generalisation seems easier.

Finally, figure 4 provides prediction examples made with the bespoke Mask R-CNN trained with the Danish dataset. It shows that most of the fields are being detected and fairly outlined. Importantly, urban areas, water bodies, and to a less extent, forests are correctly ignored by the model. Hence, one can imagine predicting a whole satellite image with the model without any processing required to remove non-agricultural areas.

(a) Ground truth for Danish
validation area 1

(b) Predictions for Danish
validation area 1

(c) Ground truth for Danish
validation area 2

(d) Predictions for Danish
validation area 2

(e) Ground truth for Danish
validation area 3

(f) Predictions for Danish validation
area 3

(g) Ground truth for French
validation area 1

(h) Predictions for French validation
area 1

(i) Ground truth for French
validation area 2

(j) Predictions for French validation
area 2

Figure 4. Example of predictions of the tailored configuration of the Mask R-CNN trained with Denmark dataset over few validation
areas from France and Denmark. Field boundaries are highlighted in yellow.

Dataset	Configuration	Precision	Recall	f1-score	OA	e_{fg}	e_{us}
Denmark	baseline	0.676	0.544	0.595	0.759	0.002	0.331
	bespoke	0.704	0.594	0.639	0.793	0.002	0.314
France	baseline	0.726	0.430	0.534	0.681	0.002	0.279
	bespoke	0.687	0.485	0.562	0.707	0.002	0.329

Table 1. Prediction accuracy assessment of the Mask R-CNN trained with different configurations on the two datasets.

Training area	Validation area	Precision	Recall	f1-score	OA	e_{fg}	e_{us}
Denmark	France	0.633	0.490	0.544	0.714	0.002	0.363
France	Denmark	0.688	0.472	0.551	0.715	0.001	0.352

Table 2. Prediction accuracy assessment of Mask R-CNN for one geography when the model is trained with data from another area.

5. CONCLUSION AND FUTURE WORK

The major contribution of the present article is the introduction of a new pipeline based on Mask R-CNN for the delineation of field boundaries over large areas. A tailored version of this instance segmentation model has shown good accuracy over Danish and French regions. Trained with a larger and richer dataset, it could help the full automation of agricultural parcel delineation for further application such as crop type classification. More modifications to the core architecture as well as the pre-processing stage could further improve the pipeline performance.

REFERENCES

Abdulla, W., 2017. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. `https://github.com/matterport/Mask_RCNN`.

Beucher, S., Meyer, F., 1990. *Morphological segmentation.* Academic Press.

Crommelinck, Koeva, Yang, Vosselman, 2019. Application of Deep Learning for Delineation of Visible Cadastral Boundaries From Remote Sensing Imagery. *Remote Sensing*, 11(21), 2505. https://doi.org/10.3390/rs11212505.

De Wit, A., Clevers, J., 2004. Efficiency and accuracy of per-field classification for operational crop mapping. *International journal of remote sensing*, 25(20), 4091–4112.

García-Pedrero, A., Lillo-Saavedra, M., Rodríguez-Esparragón, D., Gonzalo-Martín, C., 2019. Deep Learning for Automatic Outlining Agricultural Parcels: Exploiting the Land Parcel Identification System. *IEEE Access*, 7, 158223–158236.

Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sensing of Environment*, 201, 165–180.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Ji, C., 1996. Delineating agricultural field boundaries from TM imagery using dyadic wavelet transforms. *ISPRS journal of photogrammetry and remote sensing*, 51(6), 268–283.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. *European conference on computer vision*, Springer, 740–755.

Masoud, K. M., Persello, C., Tolpekin, V. A., 2020. Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using a Novel Super-Resolution Contour Detector Based on Fully Convolutional Networks. *Remote Sensing*, 12(1), 59.

North, H. C., Pairman, D., Belliss, S. E., 2019. Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1), 237-251. https://doi.org/10.1109/jstars.2018.2884513.

Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6), 1301–1316.

Persello, C., Bruzzone, L., 2009. A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1232–1244.

Persello, C., Tolpekin, V., Bergado, J., de By, R., 2019. Delineation of Agricultural Fields in Smallholder Farms From Satellite Images Using Fully Convolutional Networks and Combinatorial Grouping. *Remote Sensing of Environment*, 231(nil), 111253. https://doi.org/10.1016/j.rse.2019.111253.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Rydberg, A., Borgefors, G., 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), 2514–2520.

Schwartzberg, J. E., 1965. Reapportionment, gerrymanders, and the notion of compactness. *Minn. L. Rev.*, 50, 443.

Waldner, F., Diakogiannis, F., 2019. Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network. *arXiv preprint arXiv:1910.12023*.

Yan, L., Roy, D., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sensing of Environment*, 144, 42–64.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.