
Instance Aware Segmentation of Agricultural Fields using Mask R-CNN

Audric Dongfack K.*

audric.dongfack@centrale-casablanca.ma

May 8, 2020

ABSTRACT

This article aims to delimit agricultural field parcels from satellite images through deep learning for instance segmentation. Manual delineation is precise but time-consuming, and several automated approaches using traditional image segmentation techniques are struggling to capture the variety of possible appearances in the field. Deep learning has proven to be effective in the various computer vision tasks, and could be a good candidate for accurate, efficient and generalizable vision of agricultural fields. Here, we use the best neural network model for instance segmentation to date: the Mask R-CNN that is formed on Sentinel-2 images corresponding to agricultural field polygons in Denmark. Unlike many other approaches, the model works on raw RGB images without pre- and post-processing. The results are generalized across different field sizes, shapes and other properties, but show characteristic problems in some cases, especially when we apply it to portions of images from another geographical area (the Zaer region in our case). Overall, the results appear promising and would therefore prove the validity of the deep learning approach. In addition, the methodology offers many opportunities for improvement. The code will be made available at: <https://github.com/dongack/ForAfric-Agricultural-Fields-Delineation>

Keywords Instance segmentation · Agriculture · Deep learning · Image processing

1 INTRODUCTION

Geodata of field and plot boundaries are essential for many agricultural applications. These include, for example, crop type and yield monitoring, grant management, food security research and decision-making [7, 8]. Initially, these data were mainly used by government agencies. Today, commercial enterprises in the booming agricultural technology sector also need parcels of land in various fields of application such as farm management, yield forecasting and precision agriculture. Initiatives to delimit and map agricultural parcels include the United States Common Land Unit (CLU) and the Land Parcel Identification System (LPIS) [1]. High quality data on agricultural parcels are mainly manually plotted by experts in aerial or satellite imagery. In high-resolution imaging, single-field objects are distinguished from surrounding objects by significant transitions in colour, brightness or texture caused by changes in land use. If available, the task can be supported by the reference property or parcels, additional geospatial data provided by the farmer as well as information on the type of crop, field use status, etc. Although manual segmentation of crop blocks can be very accurate, the task is time-consuming for larger agricultural areas and is subject to intra- and inter-observer variability due to the varying experience, performance and diligence of human operators. The growing demand for consistent and regularly updated parcel data for agricultural regions around the world and particularly in Africa highlights the need for reproducible automation of this task. However, most traditional image segmentation models for automated delimitation of agricultural parcels are designed for small-scale applications and adapted to the specific environmental parameters of the study area [3]. An ideal automatic method would be able to provide satisfactory speed and accuracy for large areas with variable environmental parameters.

* work initiated by the ForAfric project with four colleagues working on different aspects of the subject.

In this article, we propose to adopt the best instance segmentation algorithm to date: the Mask R-CNN [5] to perform this task. We specially use its version implemented on the Tensorflow Framework [6] by Waleed Abdulla [4]. We train it on a training data built on the basis of the field polygons available from the Denmark region. We compare our model formed with the Fully Convolutional Neural Network (FCIS) [2], already used in the literature to perform a similar task. We finally apply the model to our region of interest: the Zaer region.

The rest of the article is structured as follows: in section 2, we present the related work. Section 3 explains the details of our approach and the Background for understanding the model. Furthermore, experimental evaluation and result are reported in section 4. Sections 5 and 6 present the model's comparisons with the FCIS and the prospects for work improvement and possible use. Finally section 7 concludes the paper.

2 RELATED WORK

Several automatic and semi-automatic methods for delimiting agricultural parcels from remote sensing images have been proposed in the literature. Most of these traditional methods are defined as image segmentation techniques applied to high-resolution multispectral imaging of different band specifications and geographic regions. This includes image segmentation based on edge detection [12, 13], image value gradients [15] and texture properties [16]. Another approach is based on multitemporal data by combining vegetation indices and edge detection [17]. These studies generally use manually selected characteristics or parameters, which require a priori knowledge of the scale, physical appearance or distribution of the fields in the scene. No training data is required. However, the advent of machine learning algorithms such as Random Forest (RF) and SVM has given rise to many methods using training data. García-Pedrero et al [9] use a machine learning approach to iteratively merge selected adjacent superpixels with the merging criteria determined by a Random Forest classifier supervised via various spectral indices and texture characteristics. The methodology requires training data, manual selection of characteristics and, possibly, further post-processing for the complete delineation of independent parcels of land. Deep learning in its version "Deep learning for computer vision and remote sensing" comes later [19]. Indeed, several algorithms give good accuracy on huge training sets: Notably in the literature AlexNet developed by Krizhevsky et al [21], VGGNet by Simonyan and Zisserman [22], LeNet, ZF Net, GoogLeNet and ResNet [14]. The appropriate computer vision task for the segmentation and classification of agricultural fields through deep learning is instance aware semantic segmentation (IAIS) (or simply instance segmentation). IAIS delimits and classifies image objects while distinguishing instances of objects of the same class. Instance segmentation is at the intersection of object detection (prediction of the delimitation box and the class of a variable number of image objects, but not segmentation) and semantic segmentation (labelling of each image pixel with a semantic category label, but not distinction between object instances of the same category). With semantic segmentation, it would be possible to find unique examples of agricultural fields if they were completely surrounded by areas of different land cover classes. However, agricultural land objects are often directly adjacent to each other and have boundaries that touch each other. In this case, semantic segmentation would produce polygonal objects "in tufts". Instance segmentation is able to distinguish between these connected or overlapping instances of the same class.

Deep learning has not yet been adopted by the remote sensing community as a whole. However, given the rapid increase in volumes of Earth observation imagery and geospatial data, it should be ideally suited for many remote sensing applications. Recent studies have explored the potential of deep learning of satellite and aerial imagery for image recognition [18], semantic segmentation [10] and object detection [11].

Christoph Rieke [3] is to our knowledge the first and only one to date to propose an deep learning instance segmentation approach using the FCIS architecture [2], we propose a second approach using a recent architecture: the Mask R-CNN learder today of instance segmentation.

3 METHODOLOGY

In this section, we present the raw training data, the construction of our training set, the training model and the evaluation methodology that were used in this work. In short, the medium-resolution satellite images (Sentinel-2, RGB) and corresponding georeferenced field boundaries were pre-processed and cut into small image chips to meet the needs of the model and the available computer resources. Then, a recurrent neural network architecture, adapted from[4], was formed and tested in a single configuration. We were not only interested in the topology of the plots, which will allow us to predict the plots of fields from another region different from the construction region of our training data.

3.1 Study region

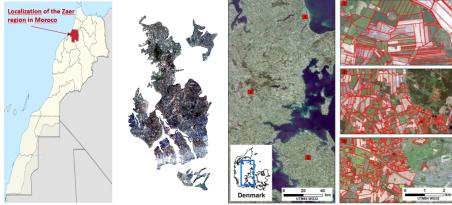


Figure 1: Illustrative images of the zaer region (the first two from the left) and the Danemark region (the first two from the right)

Overall, we can talk about two study areas. The Zaer region in northern Morocco (Africa) and an area in central Denmark (fig. 1). The second one, with the digitized agricultural parcels, was used to form our model and evaluate it. This same model that we will use later to predict agricultural fields in the Zaer region. In the end, this methodology remains all the more justified as we were only interested in the topology of agricultural plots (very varied in the Denmark area). So a priori it could be simple to generalize the model to several other agricultural areas. This same training region is used by Christoph Rieke [3] to train the FCIS. The formation area covers an area of 20900 km² (190 km x 110 km), the region covers the eastern part of continental Denmark, adjacent parts of the Baltic Sea and the second largest Danish island Funen. The region is characterized by a dominant agricultural land use, a temperate climate and generally flat terrain. The mosaic consists of two Sentinel-2 images with a spatial resolution of 10 m, taken at the beginning of the growing season in May 2016. The image tiles come from the same Sentinel-2 image acquisition and partially overlap in a north-south direction (complete image metadata in Table A1). The agricultural field polygons are from the 2016 Danish "Marker" data set, which is part of the European Union Agricultural Parcel Identification System (LPIS) initiative. The data set contains nearly 600 k polygons of field plots for the whole of Denmark, each with a unique identification number, one of the 293 crop type classes and the field area. 249298 plots are contained in the study area, including 159042 in this work. The dataset is available on [23].

3.2 Construction of training data

The construction of the training set takes place as follows: using Google Earth Engine, the bands (4, 3, 2) of the study area are filtered and then merged to form the RGB image. On the vector file describing the geometries of the fields in the zones, the polygons are reclassified entirely into a single class: the agricultural fields class. Unnecessary information in the vector file is deleted (what is only of interest here is the id and name of the culture in the polygons). The image is cut entirely into 128 * 128 pixel chips (reasonable image size for model training). Field polygons that overlap the boundaries of the grid chips are divided into at least two smaller polygons. All image chips that contain at least one field truth polygon were then randomly divided into 80% training data and 20% validation data. The image chips were then saved in RGB JPEG format without any compression. The polygonal data was saved in the COCO JSON annotation format [24]. This process is inspired by the Preprocessing presented by [3]. Look at the summary of the data construction and some examples of training images (fig.2).

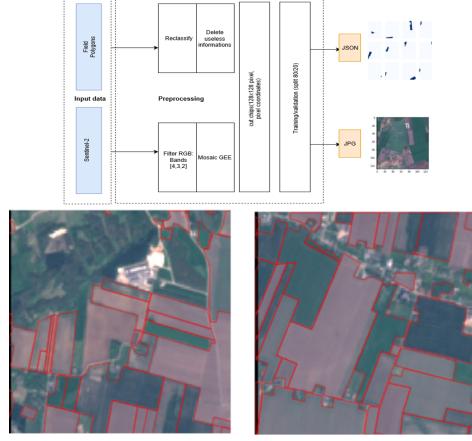


Figure 2: Data construction graph and example of training images

3.3 Mask R-CNN

To fully understand the architecture of the Mask R-CNN it is necessary to talk first of all about Object detection and the Faster R-CNN model.

Object detection: Object detection aims to find the bounding box locations and classes of image objects. The exact amount of objects is unknown. Conceptually, object detection is enabled by proposing a number of rectangular box regions in the input image to see if any of them correspond to actual image objects. This could be done by proposing boxes at every possible location and scale, and then separately applying a CNN on the image content of each box. However, this approach would be extremely performance-intensive. R-CNN (Regional CNN) [25] enables more efficient bounding box object detection with convolutional neural networks. Furthermore, the derivatives Fast-RCNN [26] and Faster-RCNN [27] greatly enhance the model training and testing speed. At this time of writing this paper, Faster-RCNN is the leading framework for object detection, and also serves as the basis for many instance segmentation models. To evaluate models for object detection (as well as instance segmentation), there are two important metrics : The average precision (AP) is the proportion of correctly delineated bounding box or segment instances. The criterion for correct delineation is based on the Intersection-over-Union (IoU) ratio. This is the area of intersection between predicted and ground truth bounding box, divided by the area of union. The end of this part will be dedicated to more explanation of this metric.

R-CNN: R-CNN first searches for a manageable number of class agnostic region proposals, i.e. a set of approximate image regions which are most likely to contain image objects. To find these regions, R-CNN uses the Selective Search algorithm [28], but is also compatible with other proposal methods. Selective Search greedily merges adjacent superpixels that share textures, colors, or intensities at multiple scales (fig. 3). R-CNN considers only the bounding boxes of these proposals.

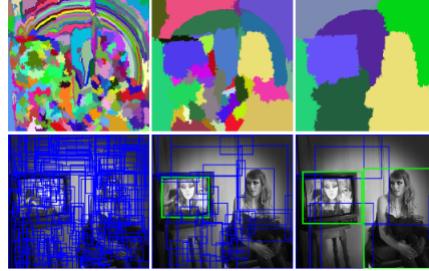


Figure 3: Examples of selective search showing the necessity of different scales [28]

For each region proposal or region of interest (RoI), this yields a class score for each of the $c+1$ predefined classes. The highest class score determines the proposal's class label and is also used as a detection score, indicating the confidence in the detection.

Redundant proposals, for which an overlapping, higher scoring proposal of the same class exists are then eliminated via non-maximum suppression (NMS). This method loops through every region proposal and compares it with all remaining region proposals of the same class. If region proposal A and another proposal B have a significant IoU overlap and the detection score of A is lower than that of B, region proposal A is rejected and removed from the loop. The remaining proposals are considered as the actual predicted image objects. Finally, to reduce localization errors and ensure tighter fitting bounding boxes, a class-specific, simple linear regression is applied on the remaining, classified region proposal. This step mildly corrects their bounding box coordinates.

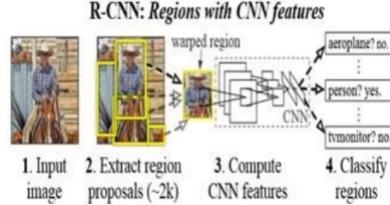


Figure 4: R-CNN [25]

Fast R-CNN: Fast R-CNN in the multi-stage pipeline of R-CNN, the CNN is applied separately on every region proposal. Because many region proposals are partially overlapping, R-CNN performs a lot of redundant calculations. The main achievement of Fast-RCNN [26] is the introduction of Region of Interest Pooling (RoIPool) for more efficient feature extraction. The classification stage of the CNN is removed, and the network is applied only once to the full image. Then, the feature vectors of the region proposals are extracted via RoIPooling directly at the corresponding locations on the output feature maps of the last convolutional layer (fig.5). Subsequently, the network performs classification and bounding box regression in parallel. RoIPool is much quicker than applying the CNN multiple times and on partially overlapping regions. Instead of training and applying three different models separately (CNN feature extraction, SVM classification, bounding box regression), Fast R-CNN combines these tasks into one joint network with shared parameters, making the network much more efficient. The training uses a multi-task loss and updates all network layers jointly.

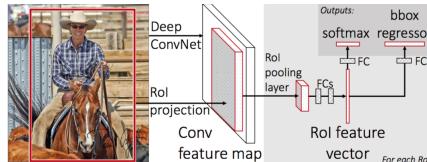


Figure 5: Fast R-CNN architecture [26]

Faster R-CNN: Faster R-CNN (fig. 6) consists of two stages. The first stage, called a Region Proposal Network (RPN), proposes candidate object bounding boxes. The second stage, which is in essence Fast R-CNN, extracts features using RoIPool from each candidate box and performs classification and bounding-box regression. The features used by both stages can be shared for faster inference [5].

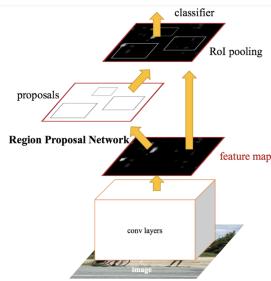


Figure 6: Faster R-CNN architecture [29]

Mask R-CNN: Mask R-CNN [5] is (at this time of writing the paper) the state of the art instance segmentation model and outperforms all other solutions. It takes the object detection architecture of Faster-RCNN (including the region proposal network) and adds a fully convolutional network branch for the pixelwise prediction of class-agnostic, binary segmentation masks (fig.7). The segmentation branch is parallel to the bounding box regression and classification branch of Faster R-CNN. All branches share the convolutional features of the input image and are applied to each ROI directly on the convolutional feature maps, which makes Mask-RCNN very efficient. The (multi-task) loss is the sum of the loss of each of the three network branches.

For each ROI, Mask-RCNN predicts a binary segmentation mask for each class fully convolutionally. The authors found that with conventional ROI Pooling of Faster R-CNN, the regions of interest in the convolutional feature maps do not completely align with the exact location of the respective regions in the original image. For the classification branch of Mask R-CNN (and also for bounding box object detection in general), such small pixel misalignments are mostly irrelevant. However, instance segmentation requires pixel-level accuracy to delineate the exact object instances. Even small translations can have a big impact on the evaluation result. They are caused by rounding issues due to the different scales or pixel dimensions of the feature maps and input image. Therefore, Mask-RCNN replaces the ROI Pool of Faster-RCNN with ROI Align, which uses bilinear interpolation to exactly align the ROIs with the corresponding regions in the original input image. ROI Align greatly improves the mask accuracy by 10 % to 50 % depending on the evaluation metric's overlap threshold [5]. To further improve the network accuracy, the convolutional feature extraction stage of Mask R-CNN uses the Feature Pyramid Network proposed by Lin et al. [2], which enables efficient, in-network processing and exploitation of convolutional feature maps at multiple scales. Compared to FCIS, Mask-RCNN also increases the number of RPN anchors and evaluates the advanced Resnext-101 Xie et al. [30] instead of only Resnet-101.

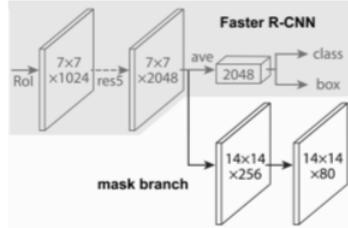


Figure 7: Mask-RCNN model with Faster-RCNN architecture (top) and parallel instance mask branch (bottom) [5]

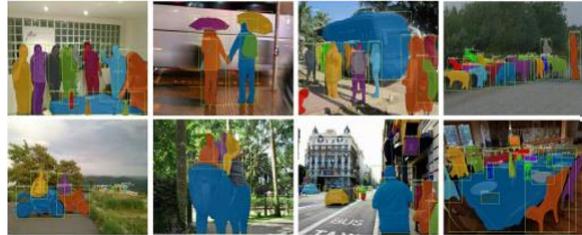


Figure 8: Mask-RCNN results on the COCO test dataset [24]. Displayed are object instances, bounding boxes and object class labels (He et al.[5])

3.4 Metric for the model evaluation

For object detection and instance segmentation models, both the training and the evaluation stage require the metrics Intersection-over-Union (IoU) and average precision (AP): When comparing the position and appearance of a predicted object bounding box (object detection) or object segment mask (instance segmentation) with the respective ground truth instance, one needs to differentiate between a correctly predicted object (true positive or TP) and a falsely predicted object that does not exist in reality (false positive or FP). This differentiation is based on the IoU score of the two polygons. Commonly, an IoU score of 0.5 or above is considered a true positive. IoU (fig.9), also called Jaccard Index, evaluates the similarity and dissimilarity of two polygons A and B. IoU is calculated by dividing the area of overlap between both objects by their area of union. For an IoU score of 1 the regions are exact matches, for a score of 0 they do not overlap at all. IoU is scale invariant.

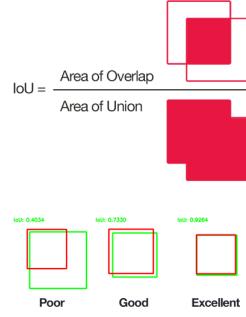


Figure 9: Jaccard index

The instance segmentation method predicts a number of instances, each with a confidence value, the detection score. For each class independently, the instances are then ranked by their detection score. Each instance is compared with each ground truth object. Based on an IoU threshold (e.g. 0.5), each predicted instance is labeled as a TP or FP. For each ground-truth object, only the predicted instance with the highest detection score is labeled as the true positive: All other predicted instances that fulfill the IoU criterium but have lower detection scores are labeled as false detections. If a true positive is found, both the ground truth and predicted object are removed from the loop. If a ground truth object is not matched by any predicted instance, it can be considered a false negative (FN). The number of false negatives results from the number of ground truth objects minus the number of true positives. From the TP/FP assignments, the precision and recall for each instance in the ranked list can be calculated. Precision (P) is the percentage of all predicted instances that match a ground truth object. Recall (R) is the percentage of ground truth objects that were correctly predicted. Each calculation takes into account the number of TP/FP/FN up to this point in the list, thus also includes the assignments from all prediction with higher detection scores. Per image chip, a maximum of 100 instances with the highest detection scores are considered.

$$\text{Precision}(P) = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall}(R) = \frac{TP}{TP + FN} \quad (2)$$

precision values of all instances are averaged within 101 bins of 0.01 recall. When the resulting precision and respective recall values are plotted, the area (equation 3) under the resulting precision-recall curve gives the average precision (AP) for class.

$$AP = \int_0^1 P(r) dr \quad (3)$$

4 Results

Our first experiment was done on a small data (300ko) containing 7 images for training and 2 images for validation. We trained the network for 8 hours on 160 epochs (in three workshops) using the Google Colab GPU. See the results in Figure (11). Despite a good network training on the training data, it can only detect barely 1 instance. By calculating the average precision on each image, we obtained an overall average of 0.025 very low so that we can assume the efficient network for an instance segmentation task on any area.

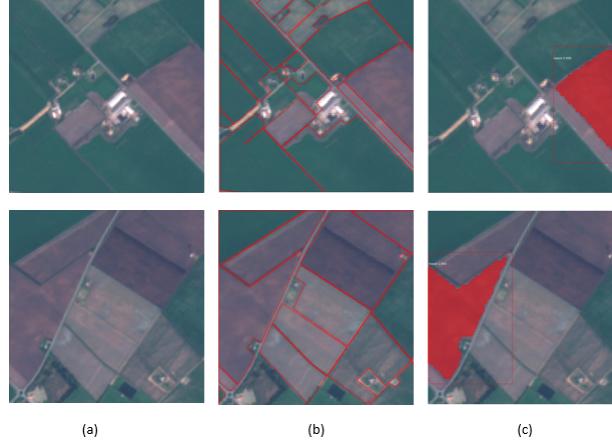


Figure 10: Result of our first experience: (a) represent the original images, (b) the ground truths and (c) the prediction image .mAP=0.025

In the second experiment, we build a larger dataset (125MB) that we use to form the network over 9*3 hours of training in 100 epochs on the Google Colab GPU. we get the following results (fig.12) .

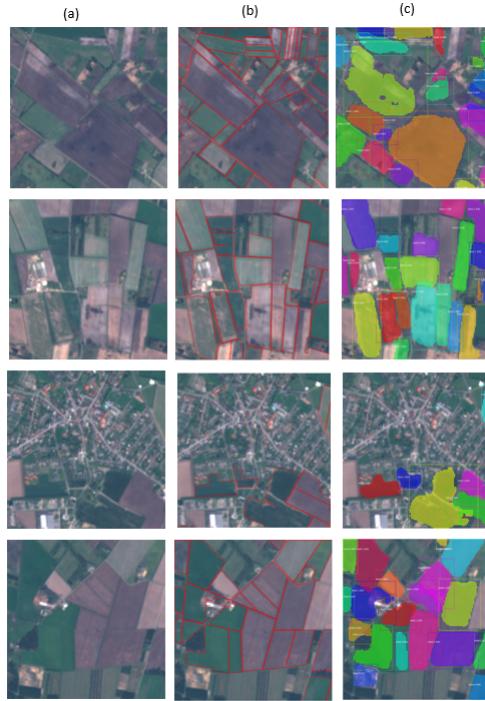


Figure 11: Result of our second experience: (a) represent the original images, (b) the ground truths and (c) the prediction images .mAP=0.7

Despite the very long training time we can notice that the model has greatly improved and detects several instances (agricultural fields) for each image. An evaluation of the latter gives us a detection score average greater than 0.8 and a mAP of 0.71 (max) on all image groups and mAP of 0.5 (min). This can testify to the performance of the training model.we could better illustrate the model's performance on a random image by construct its matrix of confusion and its precision-recall curve.The chosen image is represented in the figure 13 and its confusion matrix and precision curve -recall on the figure 14.

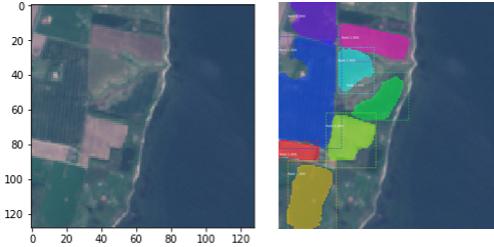


Figure 12: The image and its prediction

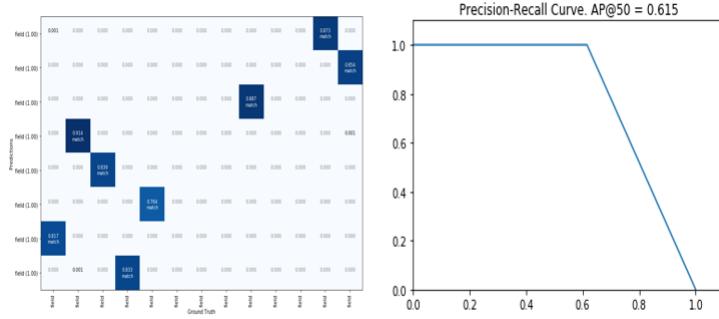


Figure 13: Confusion matrix and precision-recall curve : AP@0.5=0.615

4.1 Result in some extracted part of Zaer Region

To keep the format of the model, we apply it to a portion of the Zaer area, which gives the following result:



Figure 14: Result on a portion in the center of the Zaer region (Morocco)

the number of instances detected seems very low when looking at the field truth but the average detection score is higher than 0.8. However, the result remains surprising and this could be due to a significant difference between the field topologies of the African and European regions. A hypothesis that may seem clear to formulate in this way, but which is not a priori fair when we visually observe this variety of geometry.

5 COMPARISON WITH THE FCIS MODEL

we benchmark our model with the FCIS model train on the same data by Rieke [3]. We would like to point out that it is still difficult to give an opinion on the performance of the two models in view of this comparison. Indeed, the conditions under which the FCIS model was trained are entirely different from ours, especially since our training was not always

carried out properly since we still have a fairly large overall loss margin (loss : 0.5). After all, it may seem obvious that the Mask R-CNN is very promising.

Table 1: FCIS and Mask R-CNN comparison table

Name of Model	Model Size	Training time (But not same GPU)	mAP@0.5	Average of the detection score
FCIS	207 Mo	8 hours	0.75	0.8
Mask R-CNN	244 Mo	9*3 hours	0.71	0.8

6 PERSPECTIVES

The work is far from being completed and there are many areas for improvement. Indeed, the network can be driven from a dataset built on the Zaer area. However, this may seem tedious and may require paid mapping tools to do so and difficult to learn. However, we can safely say that this can work well.

Also, this approach of instance segmentation seems to propose openings. The model could be coupled to improve the accuracy of pixel classifications. Indeed, when we do the classification (per pixel) on a geocomputer software such as QGIS for example of the image with specific classes, it turns out that some pixels are misclassified. That is to say, assuming for example that we have two classes on the image: green for class 1 and orange for class 2, it can happen during the classification that green is found in orange and vice versa (poorly classified pixels).

However, the previous Deep Learning model only manages to regionalize portions of the previous image as belonging to a single instance (a single field). We can therefore consider that in this proportion, the majority pixels describe the class of the image well.

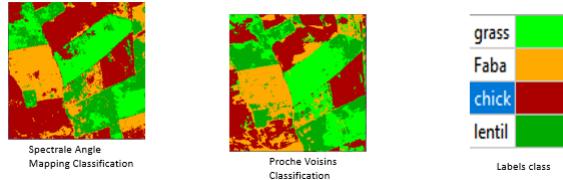


Figure 15: Here is an illustration where we use the QGIS Semi-automatic classification plugin to classify the image portion of the Zaer with two classification algorithms available in QGIS (Near Neighbors and Spectral Angel Mapping, RGB band). Some pixels are misclassified.

7 CONCLUSION

In this article, we propose to test the potential of instance segmentation by using the Mask R-CNN (the best instance segmentation framework in Deep learning at the time of writing this article) for delimiting agricultural fields from satellite images. In view of the results, the model looks promising. The main challenges of this model would be to find a way to generalize it to parcels of fields in different geographical areas. And this could be done by incorporating other geographical data into the model, increasing the model's drive accuracy. The model is still far from equaling the current automatic fields delineation models. However, predictive performance is highly dependent on the accuracy, volume and variability of the samples. Many African countries, particularly Morocco, do not have non-confidential data on their agricultural fields, which slows down research work in these areas. The methodology presented leaves room for many future improvements, for example re-training weights on various areas of the globe (precisely on Moroccan lands). Incorporate other spectral information from the image on the dataset. Finally, another useful feature of this approach would be to improve the accuracy of image classification algorithms.

8 ACKNOWLEDGES

We would like to thank the initiators of the Project: the company Forafric. Because this work has been a very great training for us. We would like to thank in particular our supervisors Mr. Hassan Khalili and Mrs. Bouchra Bensiali Research Teachers at the Ecole Centrale Casablanca for their patience and encouragement throughout the project. A

special dedication to Mr Christophe Rieke Geospatial Data Scientist for the help he provided us during this project and also to Mr Geraldin Nanfack, PhD student in Artificial Intelligence at the University of Namur. Finally, we would like to thank those who have contributed to the project, either on loan or by far.

References

- [1] Valentia Sagris, Wim Devos. LPIS core conceptual Model: methodology for feature catalogue ans application scheme, JRC Scientific and Technical Reports, 2018.
- [2] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji and Yichen Wei. Fully Convolutional Instance-aware Semantic Segmentation. conference of CVPR, 2018.
- [3] Christoph Rickre. Deep Learning for Instance Segmentation of Agricultural Fields. Master Thesis in Geoinformatics. Friedrich-Schiller-University,2017
- [4] Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. Waleed Abdulla, 2017 Github, GitHub repository: https://github.com/matterport/Mask_RCNN
- [5] Kaiming He, Georgia Gkioxari, +1 author Ross B. GirshickPublished. Mask R-CNN.IEEE International Conference on Computer Vision. DOI:10.1109/ICCV.2017.322, 2017.
- [6] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from <https://www.tensorflow.org/>
- [7] Le Gouvernement Du Grand-Duché De Luxembourg, Ministère de l'agriculture, de la viticulture et de la protection de la consommation. Rapport d'activité, 2017.
- [8] Marianne Cerf et Jean-Marc Meynard . Les outils de pilotage des cultures : diversité de leurs usages et enseignements pour leur conception. Dans Natures Sciences Sociétés 2006/1 (Vol. 14), pages 19 à 29
- [9] García-Pedrero, A. & Gonzalo-Martín, C. & Lillo-Saavedra, M. (2017). A machine learning approach for agricultural parcel delineation through agglomerative segmentation. International Journal of remote sensing, 2017. VOL. 38, NO. 7, 1809 – 1819. <https://dx.doi.org/10.1080/01431161.2016.1278312>
- [10] Saito, S. and Yamashita, T. and Aoki, Y. (2016). Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. Journal of Imaging Science and Technology R60(1). 2016
- [11] Vakalopoulou, M.& Karantzalos, K. & Komodakis, N. & Paragios, N. (2015). Building detection in very high resolution multispectral data with Deep Learning features. IGARSS 2015 - 2015 IEEE International Geoscience and Remote Sensing Symposium
- [12] Heather C. North , Member, IEEE, David Pairman , Member, IEEE, and Stella E. Belliss. Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery
- [13] Rydberg, A. & Borgefors, G. (2001). Integrated method for boundary delineation of agricultural fields in multispectral satellite images. IEEE Transactions: Geoscience and Remote Sensing 39 2514–20
- [14] Li, F.-F., Johnson, J., & Yeung, S. (2017). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved 4 November 2017, from <https://cs231n.github.io/convolutional-networks>
- [15] Butenuth, M. & Straub, M. & Heipke, C. (2004). Automatic extraction of field boundaries from aerial imagery KDNet Symposium on Knowledge-Based Services for the Public Sector. p3-4. 2004
- [16] Yalcin, H. & Günay, B. (2016). Delineation of parcels in agricultural lands in remote sensing.2016 Fifth International Conference on Agro-Geoinformatics. Tianjin, China. <https://doi.org/10.1109/Agro-Geoinformatics.2016.7577601>
- [17] Yan, L. & Roy, D. (2014). Automated crop field extraction from multi-temporal Web Enabled Landsat Data Agro-Geoinformatics. Remote Sensing of Environment 14442–64.
- [18] Zhong, Y. & Fei, F. & Liu, Y. & Zhao, B & Jiao, H. & Zhang, L. (2017) SatCNN. satellite image dataset classification using agile convolutional neural networks. Remote Sensing of Environment 14442–64. Remote Sensing Letters, 8:2, 136-145, DOI: <https://doi.org/10.1080/2150704X.2016.1235299>

- [19] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, Friedrich Fraundorfer. Deep Learning in Remote Sensing: A Review. IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE, IN PRESS.arXiv:1710.03959v1 [cs.CV] , 11 Oct 2017
- [20] Gideon M. MUSYOKA. Automatic Delination Of Small Holder agricultural field boundaries using fully convolutional networks. Phd Thesis. Enschede, The Netherlands, March, 2018
- [21] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances In Neural Information Processing Systems (pp. 1–9). Phd Thesis. Enschede, The Netherlands, March, 2018
- [22] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14. <https://doi.org/10.1016/j.infsof.2008.09.005>
- [23] Denmark LPIS agricultural field boundaries (Denmark Department for Agriculture): 293 crop/vegetation catgeories, 600k parcels, yearly dataset for 2008-2018. https://kortdata.fvm.dk/download/Markblokke_Marker?page=MarkerHistoriske
- [24] Coco dataset and coco dataset format <http://cocodataset.org/#home>
- [25] Girshick, R. & Donahue, J. & T. Darrell & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/1311.2524Girshicketal.2014>
- [26] Girshick, R. (2015). Fast R-CNN. <https://arxiv.org/abs/1504.08083>
- [27] Ren, S. & He, K. & Girshick, R. & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. <https://arxiv.org/abs/1506.01497>
- [28] Uijlings, J. & van de Sande, K. & Gevers, T. & Smeulders, A. (2012). Selective Search for Object Recognition.
- [29] <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d5> consult the 10/06/2019 ,15h
- [30] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, Kaiming He, Aggregated Residual Transformations for Deep Neural Networks arXiv:1611.05431v2 [cs.CV] 11 Apr 2017: <https://arxiv.org/pdf/1611.05431.pdf>