# Agricultural Field Delineation Using Satellite Imagery

Matej Batič, Jan Geršak, Matic Lubej, Žiga Lukšič, Nika Oman Kadunc, Devis Peressutti, Nejc Vesel, Sara Verbič

Sinergise, Ljubljana, Slovenia
matej.batic@sinergise.com, jan.gersak@sinergise.com, ziga.luksic@sinergise.com, matic.lubej@sinergise.com, nika.oman-kadunc@sinergise.com, devis.peressutti@sinergise.com, nejc.vesel@sinergise.com, sara.verbic@sinergise.com

**Abstract** Defining the borders of agricultural fields is fundamental for precision agriculture and one of the key parts of the new European Agricultural Policy. The agricultural fields' boundaries are basic building blocks for monitoring agricultural land in the context of climate change, food production and security. The aim of the field delineation process is to automatically determine the borders of agricultural fields from satellite images. It is based on the similarity of spatial, spectral, and temporal properties of pixels belonging to the same field. The basic method was developed within the NIVA project on data from the Sentinel-2 satellite constellation of the European Space Agency. The u-net based deep neural network predicts three image variables from the satellite image: the segmentation of the field, its boundary, and the distance of the segmented image points to the boundary. From these an image of the boundaries of the fields is constructed, either from a single image or from a time series of images. In the post-processing phase, the image prediction is transformed into vector format, which represents the result of the field delineation process.

**Keywords:**
field delineation, satellite imagery, agriculture, deep learning, U-Net

University of Maribor Press

DOI https://doi.org/10.18690/um.feri.4.2023.4
ISBN 978-961-286-720-1

# DOLOČEVANJE MEJA KMETIJSKIH POLJIN Z UPORABO SATELITSKIH POSNETKOV

MATEJ BATIČ, JAN GERŠAK, MATIC LUBEJ, ŽIGA LUKŠIČ, NIKA OMAN KADUNC, DEVIS PERESSUTTI, NEJC VESEL, SARA VERBIČ

Sinergise, Ljubljana, Slovenija
matej.batic@sinergise.com, jan.gersak@sinergise.com, ziga.luksic@sinergise.com, matic.lubej@sinergise.com, nika.oman-kadunc@sinergise.com, devis.peressutti@sinergise.com, nejc.vesel@sinergise.com, sara.verbic@sinergise.com

**Sinopsis** Določitev meja kmetijskih poljin je osnovni proces na področju preciznega kmetijstva ter eden ključnih členov nove Evropske kmetijske politike. Prav tako so meje kmetijskih poljin osnovni gradnik za spremljanje kmetijskih zemljišč v okviru klimatskih sprememb ter prehranske varnosti. Cilj procesa je avtomatska določitev meja kmetijskih poljin iz satelitskih posnetkov. Temelji na podobnosti prostorskih, spektralnih in časovnih lastnostih slikovnih pik, ki pripadajo isti poljini. Osnovno metodo smo razvili v okviru projekta NIVA na podatkih konstelacije satelitov Sentinel-2 Evropske Vesoljske Agencije. Globoka nevronska mreža temelji na u-net arhitekturi in iz satelitskega posnetka napove tri slikovne spremenljivke: segmentacijo poljine, mejo poljine, ter razdaljo segmentiranih slikovnih točk do meje. Iz teh treh napovedi nato sestavimo sliko meja poljin bodisi iz enega posnetka ali pa iz (daljše) časovne vrste posnetkov. V fazi naknadne obdelave slikovno napoved predelamo v vektorski format, ki predstavlja končni rezultat procesa.

**Ključne besede:** določevanje meja kmetijskih poljin, satelitski posnetki, kmetijstvo, globoko učenje, U-Net

https://doi.org/10.18690/um.feri.4.2023.4 DOI
978-961-286-720-1 ISBN

Univerzitetna založba Univerze v Mariboru

## 1 Introduction

Defining the borders of agricultural fields is fundamental for precision agriculture and one of the key parts of the new European Agricultural Policy (CAP), which dictates automatic control of agricultural land. The agricultural fields' boundaries are basic building blocks for monitoring agricultural land in the context of climate change, food production and security.

The aim of the field delineation process is to automatically determine the boundaries of agricultural fields from satellite imagery to update existing but outdated datasets of fields, fill in gaps where such data is non-existent, and finally to get a view of how the agricultural landscapes are evolving through time due to anthropogenic activities, climate changes and agricultural practices.

Determination of agricultural fields' boundaries is based on the similarity of spatial, spectral, and temporal properties of pixels belonging to the same field (agricultural land with a single crop). The initial method was developed within the NIVA project[1] on data from the Sentinel-2 satellite constellation of the European Space Agency (ESA). We have since improved the methodology, model, and processing chains. We trained a deep neural network based on AI4Boundaries dataset. The u-net architecture uses satellite imagery to predict three outputs: field segmentation, field boundary, and distance of the segmented pixels to the field border. From these three predictions, we then construct an image of the boundaries of the fields, either from a single image or from a (longer) time series of images. In the post-processing phase, the image prediction is transformed into vector format, which represents the result of the field delineation process.

In the following sections, we will dive-in into a more detailed description of each of these steps and share some of the things we have learned. In 2 Data we will describe the satellite data and ground truth dataset used. The last part of the section will present how we normalize the satellite imagery to facilitate generalizability of the model both through time as well as over larger geographical regions. In 3 Model we will outline the model, its architecture and loss functions. 4 Postprocessing (merging / vectorization) will illustrate the postprocessing of predictions, which allow us to produce results over larger areas (e.g., on continental scale). Lastly,

[1] https://www.niva4cap.eu

5 Field delineation as a service will present our field delineation service running on EuroDataCube (EDC), which was used by European Commission Joint Research Centre (JRC) to delineate agricultural fields over the whole Ukraine.

## 2 Data

### 2.1 Satellite data

For the source of satellite imagery, we use the openly available Copernicus Sentinel-2 data, accessed through Sentinel Hub services. Sentinel-2 is a land monitoring constellation of two satellites that provide optical imagery with high spatial resolution and high temporal revisit frequency, providing global coverage of the Earth's land surface every 5 days. For delineating fields, we make use of the Level-1C Top of the atmosphere (TOA) reflectance data for all the bands at 10m per pixel resolution (B02, B03, B04, B08).

### 2.2. Ground truth data

AI4Boundaries (d'Andrimont, 2023), a data set of images and labels readily usable to train and compare field boundary detection models has recently been released by JRC. To train the model, described in next section, we have used the AI4Boundaries ground-truth parcel vectors (2.5 M parcels covering 47105 km$^2$), which have been sourced from openly available Geospatial Aid Application (GSAA) datasets from Austria, Catalonia, France, Luxembourg, the Netherlands, Slovenia, and Sweden for 2019. The data in AI4Boundaries were selected using a stratified random sampling drawn based on two landscape fragmentation metrics, the perimeter/area ratio and the area covered by parcels, thus considering the diversity of the agricultural landscapes across Europe. Training samples of size 256 x 256 were created from Sentinel-2 imagery and ground truth data, as shown in Figure 1.

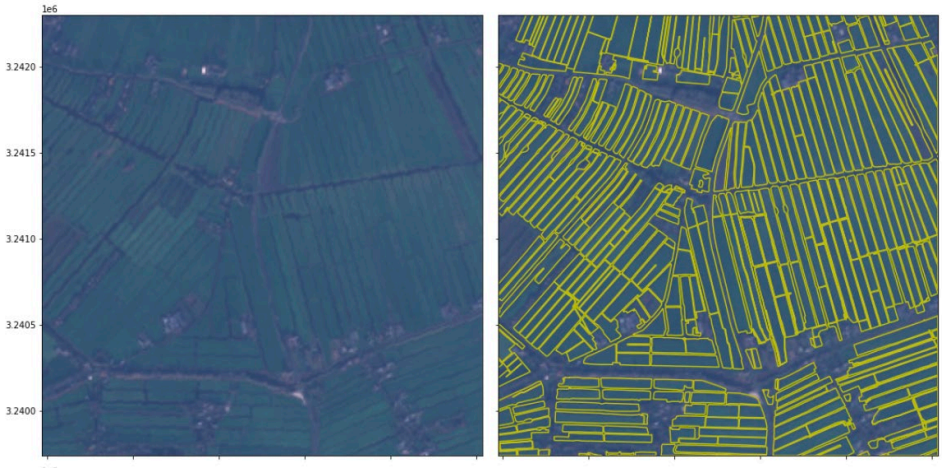



**Figure 1: True color Sentinel-2 image (left), overlaid with vector training data from AI4Boundaries (right).**
Source: own.

## 2.3 Normalization

As we want our model to perform well over timestamps taken over the whole year, it is important how the data is normalized. Normalization of the band values can have a significant impact on the network performance and the quality of the field delineation results. The input values for satellite bands are zero-bounded digital numbers and the main characteristics of the band histograms are wide value range, long-tail, and the presence of outlier values. When choosing a normalization method most suited to the properties of the satellite imagery, the aim is to center the distributions and reduce the impact of outliers. In addition, for the normalization procedure to be valid across a wide range of use cases, the training dataset must include imagery from a large geographical region and a long time interval (whole year) to capture both geographical and seasonal variability. We performed an investigation and tested several linear and non-linear normalization schemes on our field delineation model (Oman-Kadunc, 2022). A linear transformation with clipping to 1st and 99th percentile performed best. An example of various normalizations is seen in Figure 2. The observed results of our experiments indicate that mapping the main part of histogram data into the interval [0, 1], but moving outlier values out of this interval (using 1st and 99th percentile) has a large positive effect on the network convergence and performance.


**Figure 2: True color Sentinel-2 image before and after various transformations. The last image shows the transformation resulting in best model performance.**
Source: own.

## 3 Model

Reviewing the state-of-the-art in semantic segmentation of temporal images, two main approaches can be considered:

- apply semantic segmentation on each single scene separately and combine predictions temporally at a later stage. A model trained this way should learn to be invariant to the time-period of interest.
- apply semantic segmentation to a temporal stack of images, letting the model extract relevant spatio-temporal features for the task at hand. This approach tends to generate larger and slower models, as the input images contain temporal as well as spatial information (and spectral of course), but implicitly considers temporal dependencies.

The aim of the parcel delineation in CAP practices is generally to monitor agricultural land cover throughout the growing season, but the beginning of the season is of particular interest as it is typically the time when the farmers fill in their applications. A model that can generalize to different time periods seemed therefore useful in this perspective, and that justifies our choice of training a single-scene model and combining temporally the predictions in a subsequent stage. The paper from (Waldner, 2020) represents the state-of-the-art for this approach, and is what we aimed for.

In the initial implementation of the model, we implemented a model architecture as proposed in the above-mentioned paper, which utilizes a u-net backbone (Ronneberger, 2020) with added residual blocks, pyramidal pooling, and conditioned multitasking. While the model performed well on the validation set, we observed

occasional strange behavior of the model when applying it on slightly out-of-distribution data, such as a new region that was not included in the training set. To mitigate such issues, we have decided to revert to a simpler architecture which had only slightly lower scores on the validation set but exhibited a more stable behavior across a variety of real-world datasets.

The architecture that is used in the production model has the u-net backbone with added conditioned multi-tasking outputs as seen on the figure below. Additionally, the max pooling layers within the u-net are replaced with 2D Convolutions with stride 2.

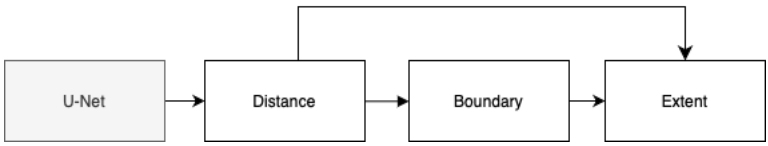



**Figure 3: Architecture of the field delineation model. The model outputs three images, showing distance, boundary, and extent of the fields. The loss is computed for each output separately and averaged out.**
Source: own.

The model is trained to solve for three conditioned tasks, shown in Figure 3. Its three outputs correspond to the boundaries of the fields, the extent of the fields and the distance from each pixel to the border. The Tanimoto loss, introduced in the ResUNet-a paper (Diakogiannis, 2019), is computed for each of the outputs and averaged to get one loss used for updating the model parameters. During the development we have observed that conditioning of the output had a large positive effect on the quality of the predictions when compared to a version of the model where the outputs were not conditioned.

While the distance is in our case not used when converting the predictions into the final output, it still serves to stabilize the outputs of the model and helps with the training. The model was trained using the Adam optimizer with a fixed learning rate across the duration of training.

The base model is trained to predict images of the same resolution as the input. For example, if we input a 256x256xN image, the model will return extent, distance and boundary masks of the same width and height. When constructing the training data,

we rasterize the reference vectors to the same spatial resolution as the satellite imagery. We wanted to see if we could extract sub-pixel information by training the model to extract information of a higher resolution than the input. This is done by adding one or multiple pixel shuffle layers to the model architecture and training the model by rasterizing the reference vectors to the target resolution, shown in Figure 4.

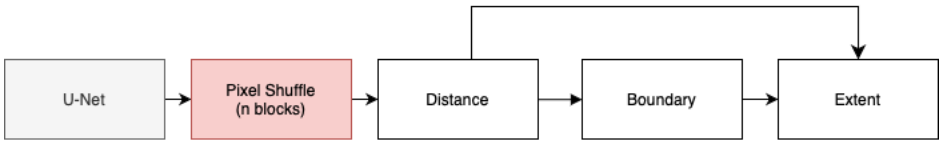



**Figure 4: Architecture of the model when we want the model to upscale the output to a higher resolution. Each pixel shuffle block corresponds to a 2x upscaling.**

Source: own.

While adding the super-resolution blocks helps the output vectors be more aligned with the actual boundaries, seen in dotted green on Figure 5 below, we also increase the number of parameters and thus the training and prediction time.

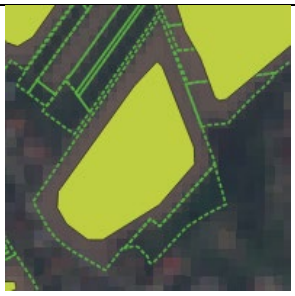
| Prediction of model without pixel shuffle blocks. | Prediction of model trained with 1 pixel shuffle block | Prediction of model trained with 2 pixel shuffle blocks. |
|---|---|---|
|  |  |  |

**Figure 5: Predictions over the same polygon of a model trained without pixel shuffle blocks (left), model trained with 1 pixel shuffle block (center) and model trained with 2 pixel shuffle blocks (right). We observe that the predicted boundaries are closer to the reference boundaries.**

Source: own.

## 4 Postprocessing (merging / vectorization)

The postprocessing of the single-scene model predictions (i.e., output of the final softmax layer or pseudo-probabilities) is split into two main parts:

- Temporal merging of predictions
- Vectorization of predictions and merging of vectors across EOPatches

### 4.1 Temporal merging

The model is applied to each available scene during the period of interest. We have observed that each single observation is subject to some degree of noise due to cloud shadows, atmospheric effects or agricultural activity that distorts the real boundaries of the fields. In addition, when running over large areas, it is not possible to choose a single cloudless timestamp that covers the whole area. We tackle these issues by temporally merging the predictions across multiple timestamps. The problem is that the fields themselves are not static through time and can undergo significant changes during only a short period of time, thus the choice of temporal merging method can have a big influence on the results, as can be seen in Figure 6.

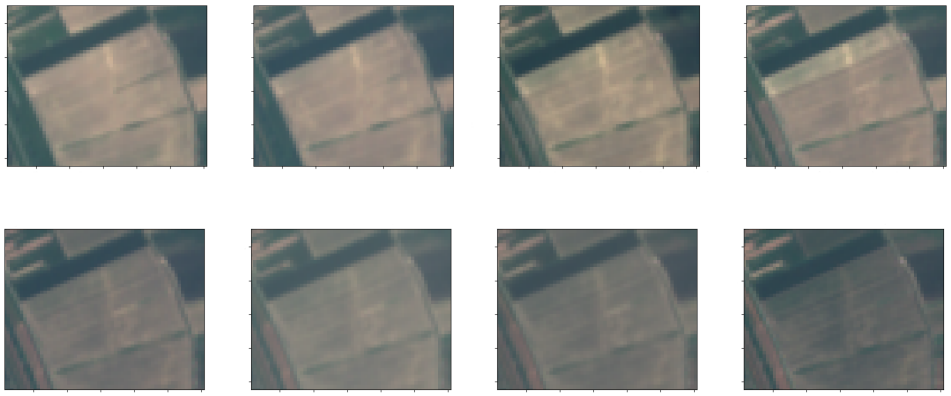


**Figure 6: Temporal changes of an agricultural parcel observed for one month (June)**
Source: own.

We temporally merge the extent and boundary predictions which we later combine into the final raster mask. The temporal merging is done on pixel level, where the pixels at the same position in each temporal prediction are merged using percentile

statistics. As the optimal temporal merging depends on the use-case at hand, we parametrize the percentile value for both the extent and boundary. For example, for use-cases where the goal is to detect the most representative state over a period, we can choose the 50$^{th}$ percentile (median) for both the extent and boundary. If the goal is to detect the most fields possible (i.e., if a certain field is split in two for only a small amount of time, this split should be detected), we can choose a high percentile value for the boundary and a low percentile value for the extent, as illustrated in Figure 7.

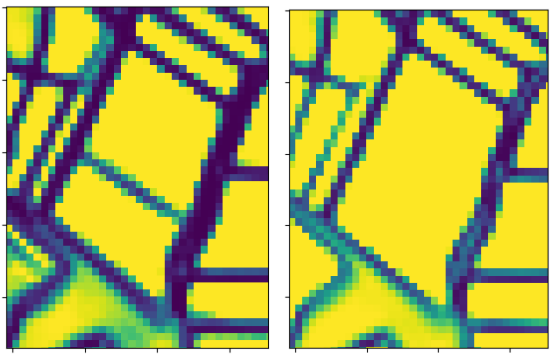


**Figure 7: The combined prediction when the percentile value is high for boundary and low for extent (left) and when both percentiles are around 50 (right)**

Source: own.

We have also explored alternative merging methods, such as max difference merging which uses the assumption that the timestamps where the difference between extent and boundary are the highest are the ones where the model is the most confidently distinguishing between the extent and the boundary. So given a temporal stack, you take the extent and boundary at the position where there is the biggest difference between the two.

Another method that we developed is merging with a rolling window, which is designed with the goal of detecting all possible stable boundaries within a period. The idea behind this method is to perform temporal merging by utilizing a rolling temporal window to smooth out outliers and to choose a stable period. The procedure computes the mean and standard deviation of extent and boundary for values inside each of the temporal windows. The window providing the best estimate

is chosen to be the one with a high boundary value, low extent value and a low standard deviation (representing a stable state).

### 4.2 Vectorization

From the step above, given a time interval, we can aggregate predictions and obtain a single pseudo-probability image for extent, boundary, and distance. We now combine these and obtain a vector layer for the entire country. To obtain smoother vectors, the pseudo-probabilities are combined into a single image as

$$p = 1 + p_{extent} - p_{boundary}, \quad (1)$$

as we didn't use the distance masks in this iteration. This resulting image has continuous values in the (0, 2) range, and can be treated as a level set functional that can be sectioned to obtain nice and smooth contours. To obtain the contours from the raster image, we used the GDAL contour[2] utility, using parameters that gave best overlaps with the GSAA vectors.

Another very useful feature of GDAL we used is the Virtual Raster Format (VRT), which allowed us to build a virtual raster containing the merged functionals of all EOPatches. This way the predictions could be blended into a smooth functional even at the borders of EOPatches. Using VRT we can run contouring parallelizing over smaller and overlapping areas, generating vector shapes that are matching over the overlapping area. To obtain a single vector layer, the overlapping geometries were merged performing a geometrical union.

## 5 Field delineation as a service

Lastly, we have put all the pieces together using eo-grow, Earth observation framework for scaled-up processing in Python[3]. In a nutshell, the following steps to produce boundaries are performed:

[2] https://gdal.org/programs/gdal_contour.html
[3] http://github.com/sentinel-hub/eo-grow/

- split the area of interest (AOI) into a regular grid to speed up processing through massive parallelization;
- download remote sensing imagery for the time interval of interest using Sentinel Hub batch processing API that outputs the imagery directly to our AWS S3 bucket;
- predict and post-process agricultural parcel boundaries on remote sensing imagery for the time interval of interest, parallelized over the EOPatches in the grid;
- perform vector merging over the whole area into a single result.

eo-grow splits the AOI into a regular grid of EOPatches, like shown in Figure 8 below. Sentinel Hub batch processing API delivers available satellite imagery for each EOPatch into directly into an AWS S3 bucket. In the next step the data is fed to the model to produce predictions, which are then post-processed and temporally merged. Vectorization is performed, and the results finally merged into a single file that can be used in GIS software.

An algorithm using the approach above is available as a service on EuroDataCube[4]. It was used within the EO4UA initiative to delineate agricultural fields over Ukraine for years 2016-2022. The web application showing results can be seen in Figure 9. The dataset facilitated further research into how war is affecting the agricultural landscape in Ukraine, their local food producing capabilities and, consequently, global food supplies.

[4] https://collections.eurodatacube.com/field-delineation/

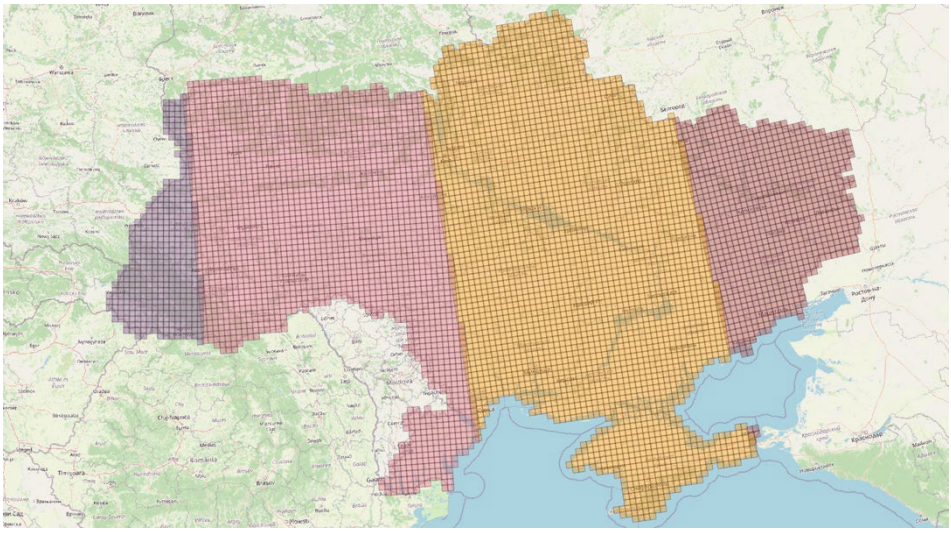


**Figure 8: Ukraine split into tiles of 10km x 10km, tiles are in their own UTM zone.**
Source: own.

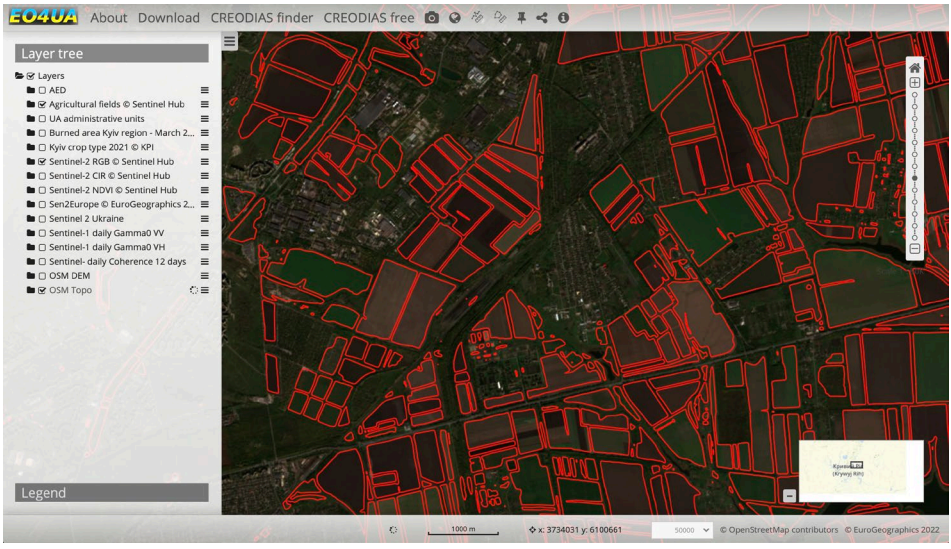



**Figure 9: Web viewer of the EO4UA initiative[5], showing Sentinel-2 true color imagery overlaid with delineated agricultural fields.**
Source: own.

[5] https://www.eo4ua.org/

**References**

d'Andrimond, R. (2023). AI4Boundaries: an open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. *Earth Syst. Sci. Data, 15,* 317–329. doi:10.5194/essd-15-317-2023

Diakogiannis, F. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing, 162*, 94-114. doi:10.1016/j.isprsjprs.2020.01.013

Oman-Kadunc, N. (2022). How To Normalize Satellite Images for Deep Learning [blog post]. *Retrieved from https://medium.com/sentinel-hub/how-to-normalize-satellite-images-for-deep-learning-d5b668c885af*

Waldner, F. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote sensing of environment, 245*. doi:10.5194/essd-15-317-2023

Ronneberger, O. (2020). U-net: Convolutional networks for biomedical image segmentation. *In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.* doi:10.1007/978-3-319-24574-4_28