DGPF

Check for updates

ORIGINAL ARTICLE

Comparison of an Optimised Multiresolution Segmentation Approach with Deep Neural Networks for Delineating Agricultural Fields from Sentinel-2 Images

Gideon Okpoti Tetteh[1] · Marcel Schwieder[1,2] · Stefan Erasmi[1] · Christopher Conrad[3] · Alexander Gocht[1]

Abstract

Effective monitoring of agricultural lands requires accurate spatial information about the locations and boundaries of agricultural fields. Through satellite imagery, such information can be mapped on a large scale at a high temporal frequency. Various methods exist in the literature for segmenting agricultural fields from satellite images. Edge-based, region-based, or hybrid segmentation methods are traditional methods that have widely been used for segmenting agricultural fields. Lately, the use of deep neural networks (DNNs) for various tasks in remote sensing has been gaining traction. Therefore, to identify the optimal method for segmenting agricultural fields from satellite images, we evaluated three state-of-the-art DNNs, namely Mask R-CNN, U-Net, and FracTAL ResUNet against the multiresolution segmentation (MRS) algorithm, which is a region-based and a more traditional segmentation method. Given that the DNNs are supervised methods, we used an optimised version of the MRS algorithm based on supervised Bayesian optimisation. Monotemporal Sentinel-2 (S2) images acquired in Lower Saxony, Germany were used in this study. Based on the agricultural parcels declared by farmers within the European Common Agricultural Policy (CAP) framework, the segmentation results of each method were evaluated using the F-score and intersection over union (IoU) metrics. The respective average F-score and IoU obtained by each method are 0.682 and 0.524 for Mask R-CNN, 0.781 and 0.646 for U-Net, 0.808 and 0.683 for FracTAL ResUNet, and 0.805 and 0.678 for the optimised MRS approach. This study shows that DNNs, particularly FracTAL ResUNet, can be effectively used for large-scale segmentation of agricultural fields from satellite images.

Keywords Agriculture · Image segmentation · Deep learning · Semantic segmentation · Instance segmentation

1 Introduction

Agriculture is an important sector of the world economy and provides humanity with essential products like food (Pandey et al. 2022). The increasing global population is accompanied by the growing demand for different varieties of food. To ensure global food security, food production must substantially increase and in parallel, the negative environmental footprint of agriculture ought to be minimised (Foley et al. 2011). This aspiration is aptly captured within the second goal ("End hunger, achieve food security and improved nutrition and promote sustainable agriculture") of the United Nation's Sustainable Development Goals (SDGs) (United Nations 2015). To sustainably achieve food security, policymakers must design agricultural policies that incentivise farmers to use sustainable agricultural practices while ensuring a decent standard of living for the farmers.

To effectively create and monitor sustainable agricultural policies, spatially explicit information about all agricultural lands is required. Further, the dynamic nature of agricultural lands requires them to be monitored in near real-time to sustainably optimise agricultural practices and react to any emerging environmental threats (Weiss et al. 2020). Remote sensing (RS) is primed for near real-time monitoring of agricultural lands (Atzberger 2013; Weiss et al. 2020). The use of RS for mapping agricultural lands has been demonstrated

✉ Gideon Okpoti Tetteh
gideon.tetteh@thuenen.de

1 Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany

2 Geography Department, Humboldt University of Berlin, Unter Den Linden 6, 10099 Berlin, Germany

3 Martin-Luther-University Halle-Wittenberg, Institute of Geosciences and Geography, 06099 Halle, Germany

in the literature at the regional (Sun et al. 2020; You et al. 2021), national (Boryan et al. 2011; Blickensdörfer et al. 2022), and continental (d'Andrimont et al. 2021) scales. The aforementioned studies classified the land-use (LU) types of agricultural lands at the pixel level. Although pixel-based image analysis is computationally efficient, especially for wide-area monitoring of agricultural lands, object-based image analysis (OBIA) generally produces more accurate LU maps as was demonstrated by some studies (Castillejo-González et al. 2009; Gilbertson et al. 2017; Belgiu and Csillik 2018). Using OBIA to map agricultural lands from RS images involves the segmentation of agricultural fields followed by the assignment of a LU type to each segmented field. The object-based crop type maps generated through OBIA enable the effective assessment of the agricultural practices being used at the field level and also allow for the accurate computation of agricultural statistics such as field sizes and shapes.

Image segmentation is the building block of OBIA (Blaschke 2010). There is a direct correlation between image segmentation quality and object-based classification accuracy (Liu and Xia 2010; Gao et al. 2011; Akcay et al. 2018). The traditional approach to segmenting agricultural fields from RS images involves the use of edge-based methods (Ji 1996; Turker and Kok 2013; Graesser and Ramankutty 2017; North et al. 2019; Wagner and Oppelt 2020), region-based methods (Möller et al. 2007; García-Pedrero et al. 2017; Belgiu and Csillik 2018; Nasrallah et al. 2018; Tetteh et al. 2020a; Luo et al. 2021), and hybrid methods (Rydberg and Borgefors 2001; Li and Xiao 2007; Yan and Roy 2014; Watkins and van Niekerk 2019) that combine edge-based and region-based methods. Choosing which method to use for image segmentation largely depends on the application needs of the user. According to Kotaridis and Lazaridou (2021), the region-based methods, particularly the multiresolution segmentation (MRS) (Baatz and Schäpe 2000) algorithm in eCognition (Trimble Germany GmbH 2019), are by far the most widely used for segmentation within the OBIA paradigm. The study of Ma et al. (2017) also revealed the popularity of the MRS algorithm.

Lately, the use of deep neural networks (DNNs) for various RS tasks like image segmentation has been gaining popularity (Ma et al. 2019). The extensive usage of DNNs in RS particularly for supporting the SDGs was recently reviewed by Persello et al. (2022). The popularity of DNNs has been facilitated by several factors including the availability of high-performance graphic cards, cloud computing, increased public availability of annotated data, and the superior performance of DNNs over shallow models (Kattenborn et al. 2021). Compared with other image segmentation methods, Kotaridis and Lazaridou (2021) showed that more peer-reviewed studies in 2020 used DNNs for segmentation. Without any manual feature engineering, DNNs, particularly deep convolutional neural networks, can exploit

hierarchical relationships between high-level and low-level features in an image, thereby making them suitable for delineating agricultural fields (Waldner and Diakogiannis 2020). Different DNNs have been used in the literature to delineate agricultural fields from RS images (García-Pedrero et al. 2019; Persello et al. 2019; Lv et al. 2020; Masoud et al. 2020; Aung et al. 2020; Waldner and Diakogiannis 2020; Meyer et al. 2020; Yang et al. 2020; Taravat et al. 2021; Waldner et al. 2021; Zhang et al. 2021; Wang et al. 2022; Jong et al. 2022; Long et al. 2022). U-Net (Ronneberger et al. 2015) and its various derivatives were the most used DNNs. The U-Net model and its derivatives are geared towards semantic segmentation; hence, they do not differentiate between objects belonging to the same class. This problem can be resolved through instance segmentation. To extract agricultural fields through instance segmentation, the DNN that was mostly utilised was Mask R-CNN (He et al. 2017).

The superiority of DNNs to shallow machine learning models such as support vector machines and random forests for various RS tasks such as land-cover and land-use classification has been highlighted in the literature (Ma et al. 2019; Kattenborn et al. 2021). However, it remains to be seen how different DNNs will compare with each other and also compare with more traditional segmentation methods like MRS for the delineation of agricultural fields from RS images. Both Yang et al. (2020) and Taravat et al. (2021) compared different DNNs for the semantic segmentation of agricultural fields but instance segmentation was not evaluated. Further, both studies did not compare their results to a more traditional segmentation method like MRS. Even though Masoud et al. (2020) compared their DNN with MRS for segmenting agricultural fields, their study had a small geographical scope (only ten tiles) and they did not evaluate any DNN for instance segmentation.

In this study, we present a large-scale comparison of the MRS algorithm with three different DNNs that have already been used in the literature to segment agricultural fields. For MRS, we used the optimised approach that was proposed by Tetteh et al. (2020a) for the segmentation of agricultural fields. Regarding the three DNNs, we selected (1) U-Net for its popularity and widespread usage for semantic segmentation, (2) Mask R-CNN for being the foremost model when it comes to instance segmentation, and (3) FracTAL ResUNet (Diakogiannis et al. 2021) for its recent usage for the effective segmentation of agricultural fields on a large scale as evidenced by these studies (Waldner et al. 2021; Wang et al. 2022).

2 Study Area and Data

As the study area, we chose Lower Saxony (Fig. 1). With about 62% of its total landmass being used as agricultural land (Tetteh et al. 2020a), Lower Saxony plays an important

role in Germany's economy regarding food production. Its agricultural areas are mostly covered by grasslands, cereals, potatoes, winter rapeseed, and sugar beet (Tetteh et al. 2020a). Lower Saxony has the largest acreage of potatoes and sugar beets in Germany, which reemphasises its key contribution to food production in Germany.

Sentinel-2 (S2) images covering Lower Saxony acquired in May of 2018 were used in this study. As suggested by Tetteh et al. (2020a), we selected May because field boundaries become more visible in this month, hence easier to delineate. Similar to Tetteh et al. (2021), the top-of-atmosphere (TOA) S2 images provided by the European Space Agency (ESA) were converted to bottom-of-atmosphere (BOA) images using the FORCE (Framework for Operational Radiometric Correction for Environmental monitoring) (Frantz 2019) processing software. With the red, green, blue, and near-infrared bands having the highest spatial resolution of S2, they were extracted from each BOA image. For each of those four bands, a mean band was created by averaging the spectral values of all pixels over the month. The four mean bands were then stacked together to create a monthly mean composite (MMC) image for May. This MMC was used in subsequent processes.

To limit the segmentation process to only agricultural areas, we masked out all non-agricultural areas from each MMC. In this study, agricultural areas equate to arable lands and grasslands. Following our previous studies (Tetteh et al. 2020a, 2020b, 2021), we extracted polygons belonging to the arable lands and grasslands in Lower Saxony from the digital landscape model (DLM) of the German Official Topographic Cartographic Information System (ATKIS). The DLM is a spatial database containing the land cover of Germany. All pixels spatially falling outside the arable lands and grasslands were removed from the MMC images.

As reference data, we used the agricultural parcels of the Geospatial Aid Application (GSAA). For farmers within the European Union (EU) to access the subsidies of the Common Agricultural Policy (CAP) (European Commission 2017), they declare the boundaries of their agricultural parcels and the corresponding LU types through the GSAA. This declaration is usually done in May of a particular year. We used the GSAA parcels of 2018. The size of the agricultural parcels ranges from 0.1 ha to 155 ha and the average size is 3 ha (Tetteh et al. 2020a).
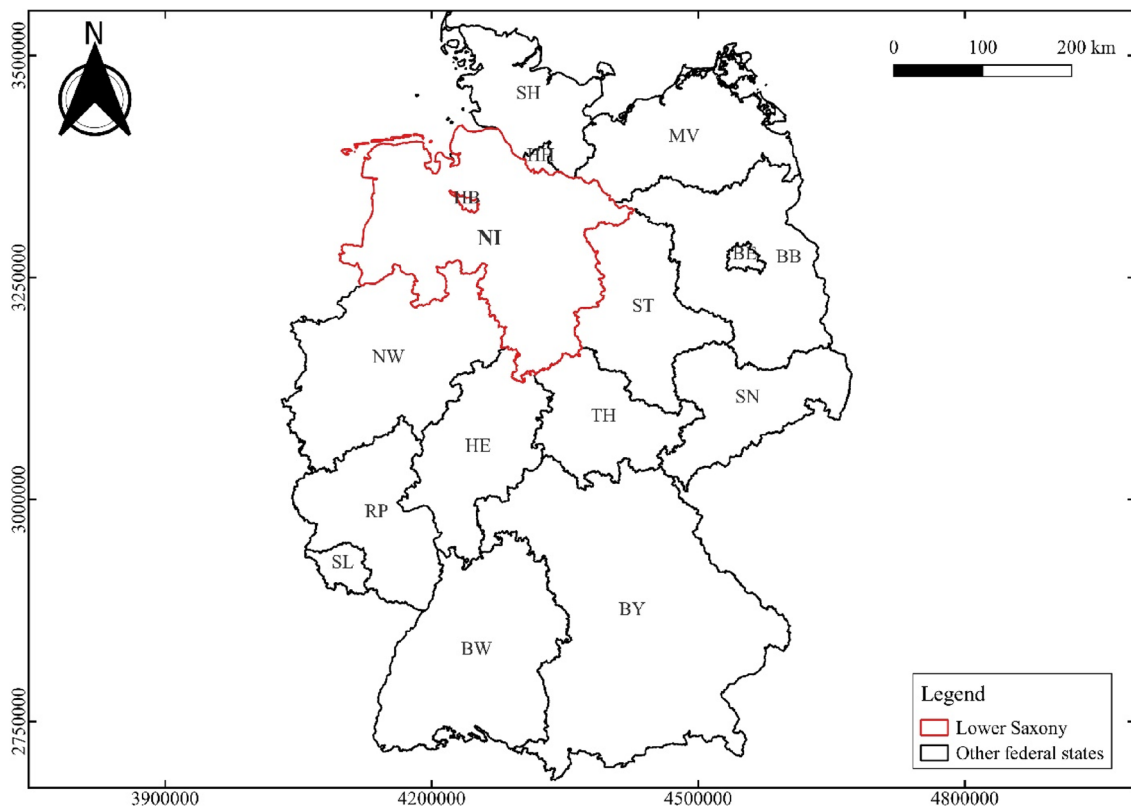
Fig. 1 The geographical location of the study area (Lower Saxony, which is labelled NI on the map above). The coordinates are in EPSG:3035. *BB* Brandenburg, *BE* Berlin, *BW* Baden-Württemberg, *BY* Bavaria, *HB* Bremen, *HE* Hesse, *HH* Hamburg, *MV* Mecklenburg-Western Pomerania, *NW* North Rhine-Westphalia, *RP* Rhineland-Palatinate, *SH* Schleswig-Holstein, *SL* Saarland, *SN* Saxony, *ST* Saxony-Anhalt, *TH* Thuringia

3 Methodology

Figure 2 is the workflow that was employed in this study. The main components of the workflow will be explained in the proceeding subsections.

3.1 Data Preparation

To ensure the efficient segmentation of the agricultural fields by the three DNNs using a graphic processing unit (NVIDIA GRID T4-16Q) with a dedicated memory of 14 GB, Lower Saxony was partitioned into 8417 tiles with each tile being 2.56 km × 2.56 km (256 pixels × 256 pixels). On average, the number of GSAA parcels per tile is 140. To ensure that there are enough GSAA parcels per tile for both the training and testing, we removed the tiles with less than 50 parcels. This brought down the number of tiles to 7169.

From the 7169 tiles, a stratified random sampling approach was used to split the tiles into 70% training tiles (5018) and 30% test tiles (2151). To do the stratification, we used two steps. Following the approach of Tetteh et al. (2021), we first computed the shape factor (SF) per tile as shown in Eq. (1);

$$SF = \frac{1}{n} \sum_{i=1}^{n} \frac{4 \times \pi \times \text{Area}(X_i)}{\left(\text{Perimeter}(X_i)\right)^2} \tag{1}$$

where X is a GSAA parcel and n is the number of GSAA parcels per tile. The SF, which is based on the method of Polsby and Popper (1991), is a measure of the level of compactness per tile. It ranges from 0 (lowest compactness) to 1 (highest compactness). A tile with low compactness indicates that the agricultural fields present at that tile are more elongated and a tile dominated by more circular fields will have high compactness. Second, after some visual analysis,

Fig. 2 Overview of the workflow used in this study. *ATKIS* German Official Topographic Cartographic Information System, *MMCs* monthly mean composites, *GSAA* Geospatial Aid Application, *DNNs* deep neural networks, *MRS* multiresolution segmentation
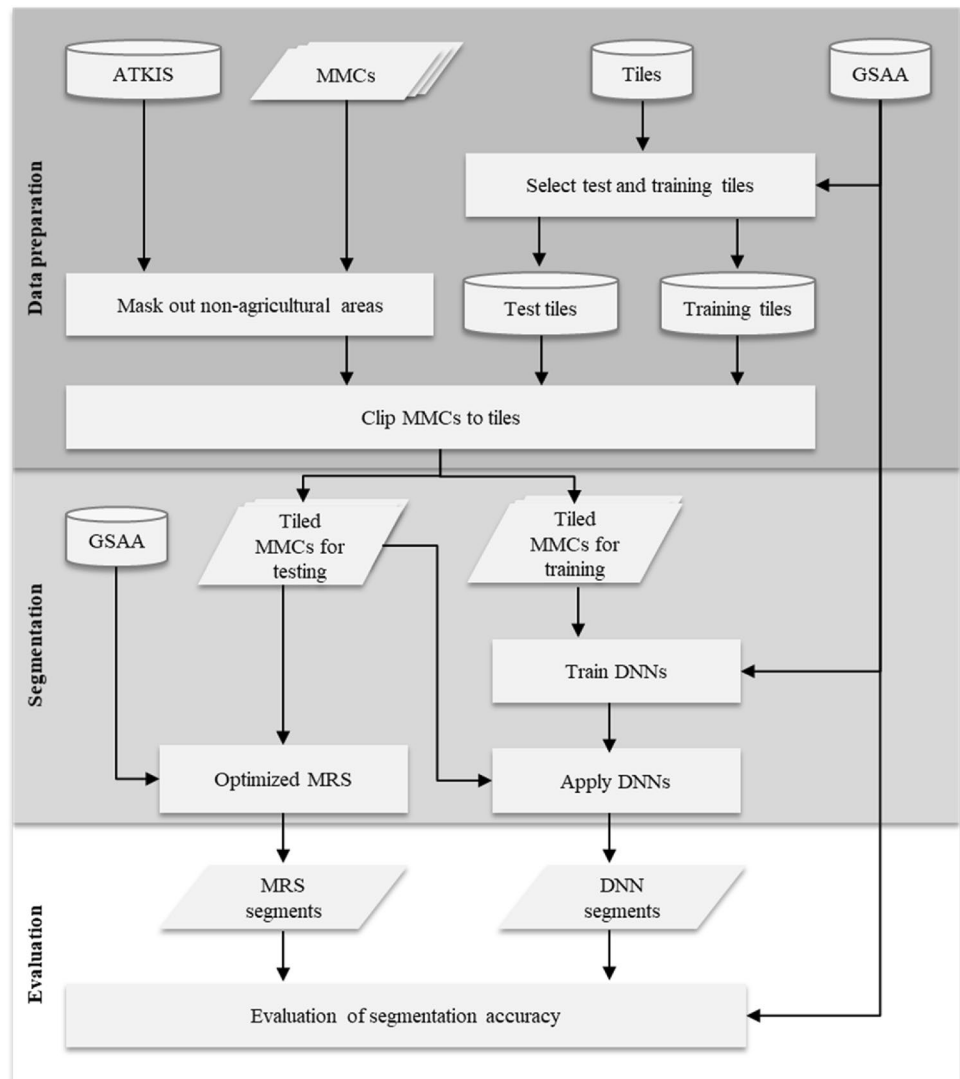
we categorised the SFs of the tiles into three classes namely low compactness $(0.0 < SF \le 0.4)$, medium compactness $(0.4 < SF \le 0.6)$, and high compactness $(0.6 < SF \le 1.0)$. The stratification was done based on this categorization. Figures 3 and 4, respectively, show the training and test tiles, where each tile is coloured by its corresponding SF class.

For each tile, a corresponding image chip was clipped out from the masked MMC images. In all, 5018 training images and 2151 test images were created.

3.2 Segmentation Methods

3.2.1 U-Net

U-Net was initially designed for the semantic labelling of pixels in biomedical images. It is now widely used for the semantic segmentation of different types of images including RS images. Its use for delineating agricultural fields from RS images has been demonstrated in the literature (García-Pedrero et al. 2019; Aung et al. 2020; Yang et al. 2020; Taravat et al. 2021). U-Net has two parts: a contracting path (encoder) for extracting features from an input image and an expansive path (decoder) for precise localization and upsampling of the extracted features to the same dimension as the input image. The contracting path is a typical convolutional network consisting of the repeated application of two convolutions, each followed by a rectified linear unit (ReLU) and max pooling. Every step in the expansive path consists of up-convolution and concatenation followed by the application of two convolutions with a ReLU. As the final layer, a convolution is applied to translate the extracted features to the desired number of classes, and an activation function (softmax in our study) is used to assign class probabilities to each pixel. Further details about U-Net can be found in Ronneberger et al. (2015).

3.2.2 FracTAL ResUNet

Following the encoder–decoder style of U-Net, Diakogiannis et al. (2020) proposed ResUNet-a, a novel network for semantic segmentation. The encoder and decoder blocks of ResUNet-a are composed of residual blocks of convolutional layers (He et al. 2016) followed by pyramid scene parsing pooling (Zhao et al. 2017). In each residual block, multiple parallel atrous convolutions (Chen et al. 2017a, 2017b) with different dilation rates were used. A more detailed explanation of ResUNet-a can be found in Diakogiannis et al. (2020). In a change detection study, Diakogiannis et al. (2021) defined a new model by introducing a self-attention mechanism to the ResUNet-a architecture.
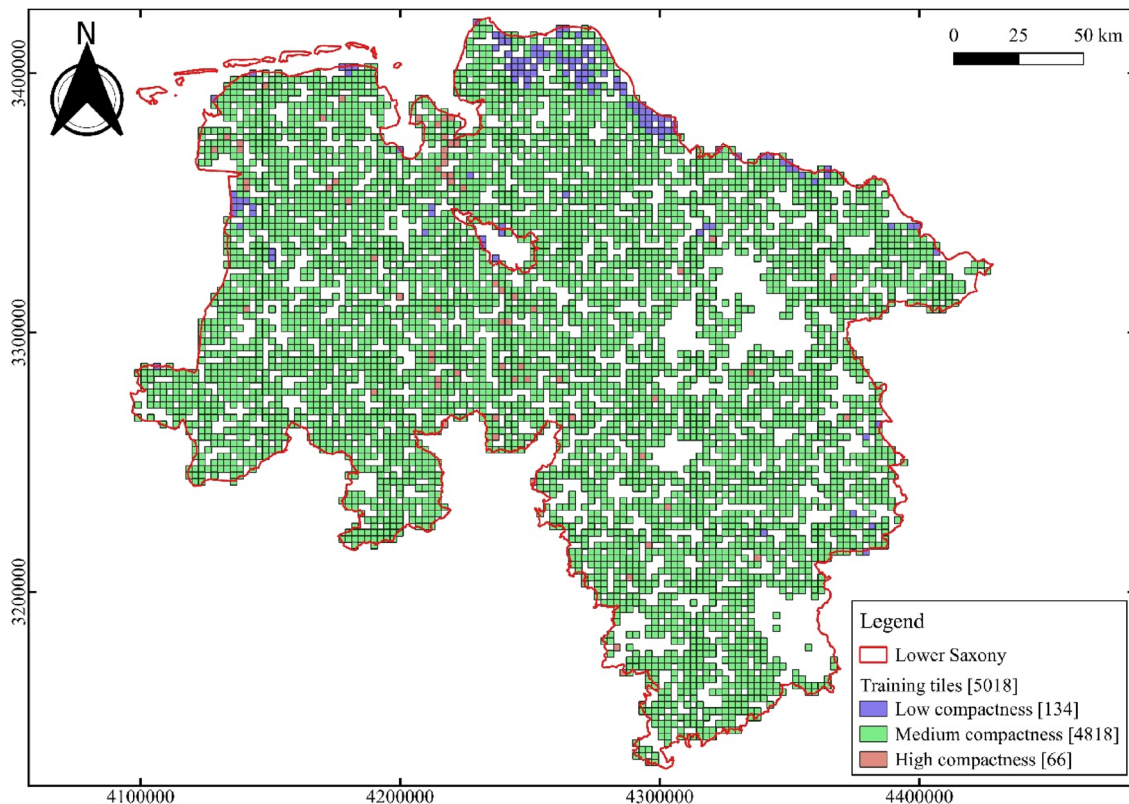
Fig. 3 The tiles that were selected for training. Each tile is coloured by its shape factor (SF) class
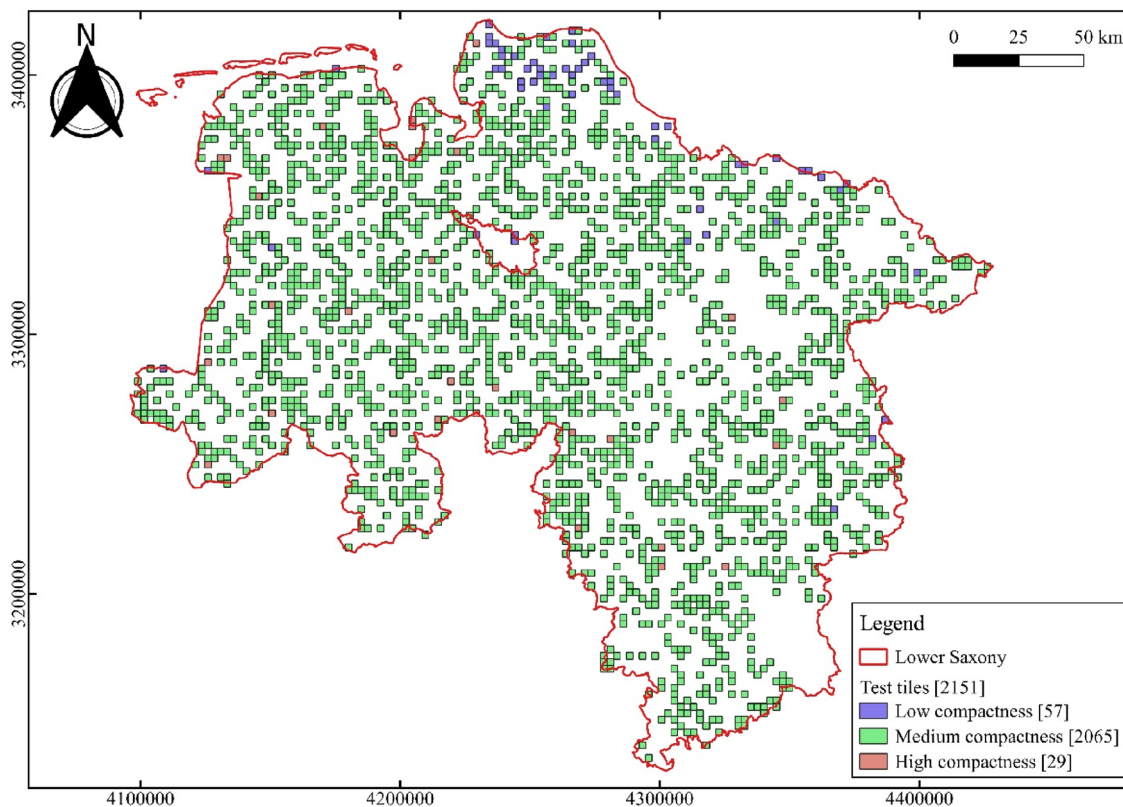
Fig. 4 The tiles that were selected for testing. Each tile is coloured by its shape factor (SF) class

Each residual block of ResUNet-a with the atrous convolutions was replaced by a residual block with a Fractal Tanimoto Attention Layer (FracTAL). The authors consequently named this new network FracTAL ResUNet. This network was used by Waldner et al. (2021) and Wang et al. (2022) for agricultural field delineation from satellite images.

3.2.3 Mask R-CNN

Mask R-CNN is an extension of Faster R-CNN (Ren et al. 2016). It maintains the bounding box recognition and classification branches of Faster R-CNN and in parallel adds a branch for predicting binary segmentation masks on each Region of Interest (RoI) (He et al. 2017). Therefore, Mask R-CNN is meant for instance segmentation (semantic segmentation and object detection). The Mask R-CNN architecture has two components: a backbone and a head. The backbone uses a convolutional neural network (CNN) typically Resnet-101 (He et al. 2015) and the Feature Pyramid Network (FPN) (Lin et al. 2017) to extract feature maps from the input image. The head section uses a Region Proposal Network (RPN) for extracting the RoIs, an RoI alignment layer for aligning the RoIs with the corresponding regions in the input image, fully connected layers for bounding box regression and softmax classification, and a fully convolutional

network (FCN) for generating a binary segmentation mask for each user-defined class. More details about Mask R-CNN can be found in He et al. (2017). Some researchers (Lv et al. 2020; Meyer et al. 2020) have used Mask R-CNN for segmenting agricultural fields.

3.2.4 Optimised MRS

Unlike the DNNs, the MRS algorithm does not require training. It can simply be applied to any image of interest to generate corresponding segments. The outcome of the algorithm is controlled by three main parameters namely scale, shape, and compactness. With each parameter taking varying input values, an endless number of parameter combinations could be generated. Determining the optimal parameter combination to use for the segmentation of each test image could be done through supervised or unsupervised optimisation. Supervised optimisation involves the use of reference data while unsupervised optimisation involves the direct use of the image content to identify the optimal combination. We demonstrated in our previous study (Tetteh et al. 2020b) that optimising the MRS parameters in an unsupervised manner produces significantly lower segmentation accuracies when compared to supervised optimisation. Therefore, in this study, we used the supervised segmentation optimisation (SSO) approach proposed by

Tetteh et al. (2020a). The core of that SSO approach is the use of the MRS algorithm in eCognition, Bayesian optimisation, and supervised segmentation evaluation. To use Bayesian optimisation, one would have to define an objective function to optimise (maximise or minimise). An objective function is a function that takes some input (here, a combination of scale, shape, and compactness) and then returns a metric. In the SSO approach, this metric was computed through supervised segmentation evaluation, which involves the geometric comparison of segments created with the MRS algorithm with their corresponding GSAA parcels. The specific metric that we computed was the area-weighted average of the Jaccard index (Jaccard 1901). The Jaccard index is popularly known as intersection over union (IoU). The parameter combination with the highest area-weighted IoU value is considered the best combination and the corresponding segmentation result is returned by the SSO.

3.3 Segmentation Experiments

To train the DNNs, we generated two classes namely field (class label = 1) and boundary (class label = 2) from the GSAA parcels per training tile. For each GSAA parcel, we applied an inward buffer of 5 m. The inwardly buffered polygons represented the field layer. The geometric difference between the GSAA parcels and the field layer constituted the boundary layer. Those two layers were subsequently rasterised to create a reference image per training tile. Using all four bands, each training and test image had a size of $256 \times 256 \times 4$. For all DNNs, the number of training epochs was set at 50.

Following the approach of García-Pedrero et al. (2019), we compiled U-Net using the Adam optimiser (Kingma and Ba 2017) with a learning rate of 0.0001. For the loss function, we used categorical cross-entropy as is usually done in the literature when it comes to multiclass classification with DNNs. The U-Net model was trained in TensorFlow (Abadi et al. 2016) using a batch size of 20. The trained U-Net model, when applied to any test image, returns a pixel-wise probability image in which each pixel is allocated the probabilities of the field and boundary classes. The actual class label per pixel is then determined as the arg max of the probability image. The outcome of this arg max is an image in which each pixel is either assigned to a field or a boundary.

Regarding FracTAL ResUNet, we used the model and corresponding hyperparameters that were defined in Waldner et al. (2021). To train FracTAL ResUNet, three reference images must be generated for each training image. The three reference images are the extent mask, boundary mask, and distance image. The extent mask is a binary image, where all field pixels (class label 1) are one and other pixels are zero. The boundary mask is also a binary image with boundary pixels (class label 2) being one and other pixels

being zero. The distance image is created by applying a distance transform to the extent mask and then normalising the resultant image between zero and one. The training of the model was done with the MXNet (Chen et al. 2015) deep learning library. Here, the batch size was reduced to four to enable MXNet to run without raising memory errors. When the trained FracTAL ResUNet model is applied to any test image, it generates three output layers namely an extent (field) probability image, a boundary probability image, and a distance image. To delineate the agricultural fields, Waldner et al. (2021) used the extent and boundary probability images as inputs to hierarchical watershed segmentation. The quality of the delineated agricultural fields depends on the specific dynamics threshold (t_b) applied to the edge-weighted graph generated from the boundary probability image and the extent threshold (t_e) applied to the extent probability image. Just like Waldner et al. (2021), we set t_b to 0.2 and t_e to 0.4. The outcome of the heirachical watershed segmentation is an image in which a unique number is assigned to all pixels belonging to each detected field instance.

When it comes to Mask R-CNN, we used the TensorFlow implementation of Abdulla (2017). To enable Mask R-CNN correctly learn the variable field sizes and shapes contained in a satellite image, we followed the approach of Meyer et al. (2020) by changing the RPN anchor scales from (32, 64, 128, 256, 512) to (8, 16, 32, 64, 128) and the anchor ratios from (0.5, 1, 2) to (0.1, 0.5, 1, 2, 4). Further, we changed the maximum number of ground truth instances to use per image from 100 to 554 to ensure that all available GSAA parcels per image are used during training. We used 554 because it is the maximum number of GSAA parcels per tile. The number of image channels to use was changed from three to four. We set the number of classes to one corresponding to class label 1, given that we are only interested in field instances. The batch size was set to eight to avoid the memory errors raised by TensorFlow when the batch size was set higher than eight. After applying the trained Mask R-CNN model to any test image, a binary image is created in which pixels belonging to each detected field instance are assigned a value of one and non-field pixels are set to zero.

To effectively use the SSO approach in delineating agricultural fields in any input image, Tetteh et al. (2021) proposed an image masking approach in which the agricultural land-cover polygons extracted from the DLM of ATKIS were first inwardly (negatively) buffered by 5 m to create a separation between adjacent polygons. These inwardly buffered polygons were then used to mask out the non-agricultural areas. This masking process pre-segmented the input image. We adopted this masking approach in this study to mask the test images before applying the SSO to segment the fields.

3.4 Evaluation of Segmentation Accuracy

In semantic and instance segmentation tasks, the accuracy of a segmentation output is usually measured at the pixel level from a confusion matrix. However, we are interested in the geometric accuracy of only the segmented fields; hence, we opted for object-based accuracy assessment (OBAA). Before the OBAA, we first created field polygons by simply vectorising only the field pixels of the segmented output image generated by each DNN. The output of MRS is a vector layer; hence, no vectorization was needed. For each method, we calculated two OBAA metrics commonly used in computer vision tasks to assess the geometric similarity between target objects (vectorised field layers) and their corresponding reference objects (GSAA parcels) per test tile. The first metric was the IoU (Eq. (2)):

$$IoU = \frac{\text{Area}(X \cap Y)}{\text{Area}(X \cup Y)} \qquad (2)$$

where X refers to all GSAA parcels, Y refers to all vectorised fields, $\cap$ is the spatial intersection operator, and $\cup$ represents the spatial union operator. The IoU metric ranges from 0 (no geometric match) to 1 (complete geometric match). Smaller fields are generally more sensitive to the IoU metric than bigger fields, especially where there is a small spatial misalignment between the fields and their corresponding reference objects (Tetteh et al. 2021). Therefore, as a second metric, we computed F-score, which is captured as F_1 in Eq. (3):

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (3)$$

where *Precision* (Eq. (4)) measures the level of under-segmentation in the segmentation output and *Recall* (Eq. (5)) measures the level of over-segmentation:

$$\text{Precision} = \frac{\text{Area}(X \cap Y)}{\text{Area}(Y)} \qquad (4)$$

$$\text{Recall} = \frac{\text{Area}(X \cap Y)}{\text{Area}(X)} \qquad (5)$$

The variables and symbols in Eqs. (4) and (5) have the same meaning as in Eq. (2). The F-score, precision, and recall metrics also range from 0 (worst segmentation) to 1 (perfect segmentation).

4 Results

The performance of each method averaged over the 2151 test tiles is reported in Table 1. The distribution of the precision, recall, F-score, and IoU values can, respectively, be seen in Figs. 10, 11, 12, and 13 of the appendix. From Table 1,

the FracTAL ResUNet method achieved the highest average recall, F-score, and IoU values. The performance of the optimised MRS approach was close to that of FracTAL ResUNet. Except for the precision metric, Mask R-CNN obtained the worst performance in all other metrics. The optimised MRS and FracTAL ResUNet methods obtained the lowest precision with U-Net achieving the highest precision.

Based on the F-score and IoU metrics, we analysed the performance of each method for the three SF classes created at the data preparation stage. Figures 5 and 6 are violin plots, respectively, showing the distribution of the F-score and IoU values per SF class for the four methods. In both figures, the density curves of the methods, particularly for Mask R-CNN, have wider spreads at the low compactness class but narrower spreads at the medium and high compactness classes. Regardless of which method, on average, the lowest F-score and IoU values were obtained by test tiles with low compactness, and the highest F-score and IoU values were obtained by test tiles with medium or high compactness.

A visual inspection (Figs. 7, 8, 9) of the segmentation results of the methods at three tiles, respectively, selected from the three SF classes reaffirms the results shown in Figs. 5 and 6. The segmentation outcome for a tile with low compactness is shown in Fig. 7, the outcome for a tile with medium compactness is captured by Fig. 8, and the outcome for a tile with high compactness is shown in Fig. 9. For each of those three figures, the corresponding F-score and IoU values obtained by each method are also reported in Table 2. As discernible from Table 2, the lowest accuracies were obtained in the low compactness class, and the highest accuracy was achieved in the high compactness class.

5 Discussion

Looking at Tables 1 and 2, a positive correlation can be established between the F-score and IoU metrics. This correlation can be linked to the similar mathematical formulations of those two metrics (Maxwell et al. 2021). The F-score values were higher than the IoU values due to the higher weight put on correctly delineated areas (intersection between the reference and target objects) by F-score. Regardless of which

Table 1 The performance achieved by each method averaged over the 2151 test tiles

| Method	Precision	Recall	F-score	IoU
Mask R-CNN	0.726	0.654	0.682	0.524
U-Net	**0.733**	0.843	0.781	0.646
MRS	0.715	0.928	0.805	0.678
FracTAL ResUNet	0.714	**0.940**	**0.808**	**0.683**

The best performances are in bold font
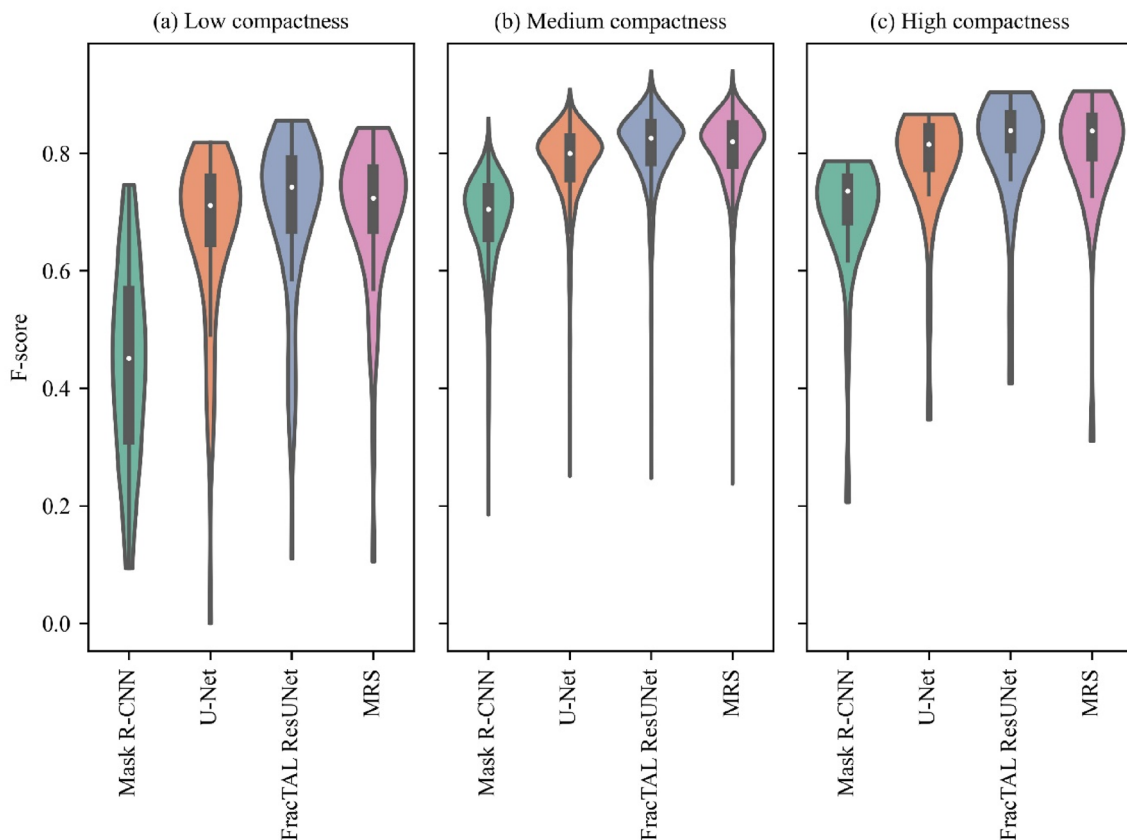
Fig. 5 Violin plots showing the distribution of the F-score values for each method for the test tiles with **a** low compactness, **b** medium compactness, and **c** high compactness

of those two metrics is opted for, FracTAL ResUNet proved to be the clear-cut winner among the DNNs and ultimately the best method as it also outperformed the optimised MRS. Although the respective differences in the F-Score and IoU values between the FracTAL ResUNet and optimised MRS as captured in Table 1 were the smallest, a paired t-test, respectively, performed with all the F-score and IoU values revealed that the differences were statistically significant (p value < 0.006 for F-score and p value < 0.001 for IoU). Overall, Mask R-CNN had the worst performance. Just like the segmentation results generated for France and Denmark by Meyer et al. (2020) from S2 images, the segments created with Mask R-CNN in this study as captured in Figs. 7c, 8c, and 9c were often wobbly and did not properly capture the spatial boundaries (edges) of the agricultural fields. Consequently, Mask R-CNN generally produced the most over-segmented results as can be observed in Table 1, where it had the worst average recall. With very similar precision values (Table 1), all four methods generated segmentation results with acceptable under-segmentation rates.

The impact of the size and shape of agricultural fields on the accuracy of the subsequent segmentation process has been well documented in previous studies (Tetteh et al. 2020a, 2020b, 2021). In those previous studies, it was observed that in areas where the agricultural fields were small and/or elongated (i.e. small compactness), the segmentation accuracy was low, and in areas with big and more compact fields (high compactness), the segmentation accuracies were high. This observation is coterminous with the results shown in Figs. 5 and 6, where the F-score and IoU values of all methods increased with increasing compactness. The negative impact of elongated fields on segmentation accuracies was more prominent in the results of the Mask R-CNN method. As visible in Fig. 7c, where the tile had low compactness, Mask R-CNN obtained the worst F-score and IoU values (see Table 2). A look at Fig. 7c clearly shows that Mask R-CNN was unable to detect numerous agricultural fields at that tile, which led to massive over-segmentation. Beyond the low compactness, the agricultural fields at the tile shown in Fig. 7a were mostly dominated by mowing pasture, thereby making it difficult to identify visible boundaries between the individual fields. In discussing their segmentation results, Waldner et al. (2021) noted that in areas where pasture was prevalent, the field delineation was less accurate. Therefore, tiles such as Fig. 7a will pose problems for any segmentation algorithm because the
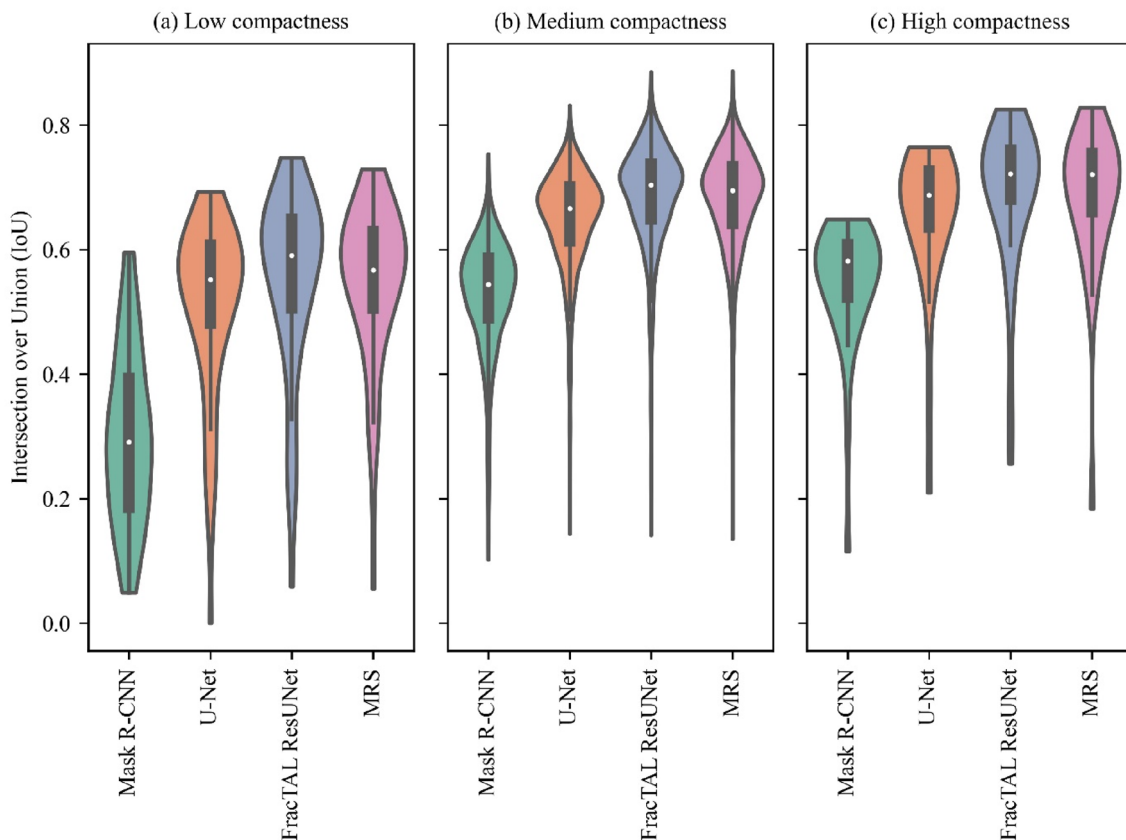
Fig. 6 Violin plots showing the distribution of the intersection over union (IoU) values for each method for the test tiles with **a** low compactness, **b** medium compactness, and **c** high compactness

spatial resolution of S2 does not allow for the proper resolution of the agricultural fields present at such tiles (Tetteh et al. 2020a, 2020b, 2021). In Fig. 7, although the results of U-Net (Fig. 7d), FracTAL ResUNet (Fig. 7e), and the optimised MRS (Fig. 7f) had numerous instances of under-segmentation, those three methods performed fairly well as they correctly delineated most of the agricultural fields, unlike Mask R-CNN. In Figs. 8 and 9, where the agricultural fields were more compact and had different LU types, the corresponding segments generated by all methods had better geometric matches to the GSAA parcels when compared with the segmentation results of Fig. 7 as aptly captured by the F-score and IoU values of Table 2. For both the F-score and IoU metrics, the highest leap in segmentation accuracy was recorded for Mask R-CNN from low compactness to medium compactness (see Table 2). From the low compactness to the medium compactness, the segmentation accuracies of U-Net, FracTAL ResUNet, and the optimised MRS remained fairly stable (see Table 2).

In this study, the two methods that stood out were FracTAL ResUNet and the optimised MRS. The performance of FracTAL ResUNet could be linked to (1) the use of residual

convolution blocks to deal with the problem of vanishing or exploding gradients while training a DNN (Diakogiannis et al. 2020), (2) the use of the self-attention mechanism to emphasise important features in convolution layers (Waldner et al. 2021), and (3) the use of conditioned multitasking whereby a distance image is first predicted, then this information is used to predict boundaries, and finally, both predictions are used as the basis to predict extents. It is important to emphasise here that FracTAL ResUNet as was used by Waldner et al. (2021) could best be described as a feature engineering method to extract features (extent probability and boundary probability images) that will be post-processed to generate the agricultural fields. It remains to be seen how FracTAL ResUNet would perform when it is directly used to extract agricultural fields through pixel-wise semantic labelling without applying any post-processing method like hierarchical watershed segmentation. As implemented by Waldner et al. (2021), the accuracy of the segmented agricultural fields would largely depend on the specific dynamics threshold (t_b) and extent threshold (t_e) passed to the hierarchical watershed segmentation algorithm.

Fig. 7 The segmentation result obtained at a tile with low compactness. **a** The masked MMC image of the tile, **b** the GSAA parcels (cyan outlines) overlaid on the masked MMC image, **c** the segmentation result (yellow outlines) of Mask R-CNN, **d** the segmentation result (orange outlines) of U-Net, **e** the segmentation result (blue outlines) of FracTAL ResUNet, and **f** the segmentation result (red outlines) of the optimised MRS
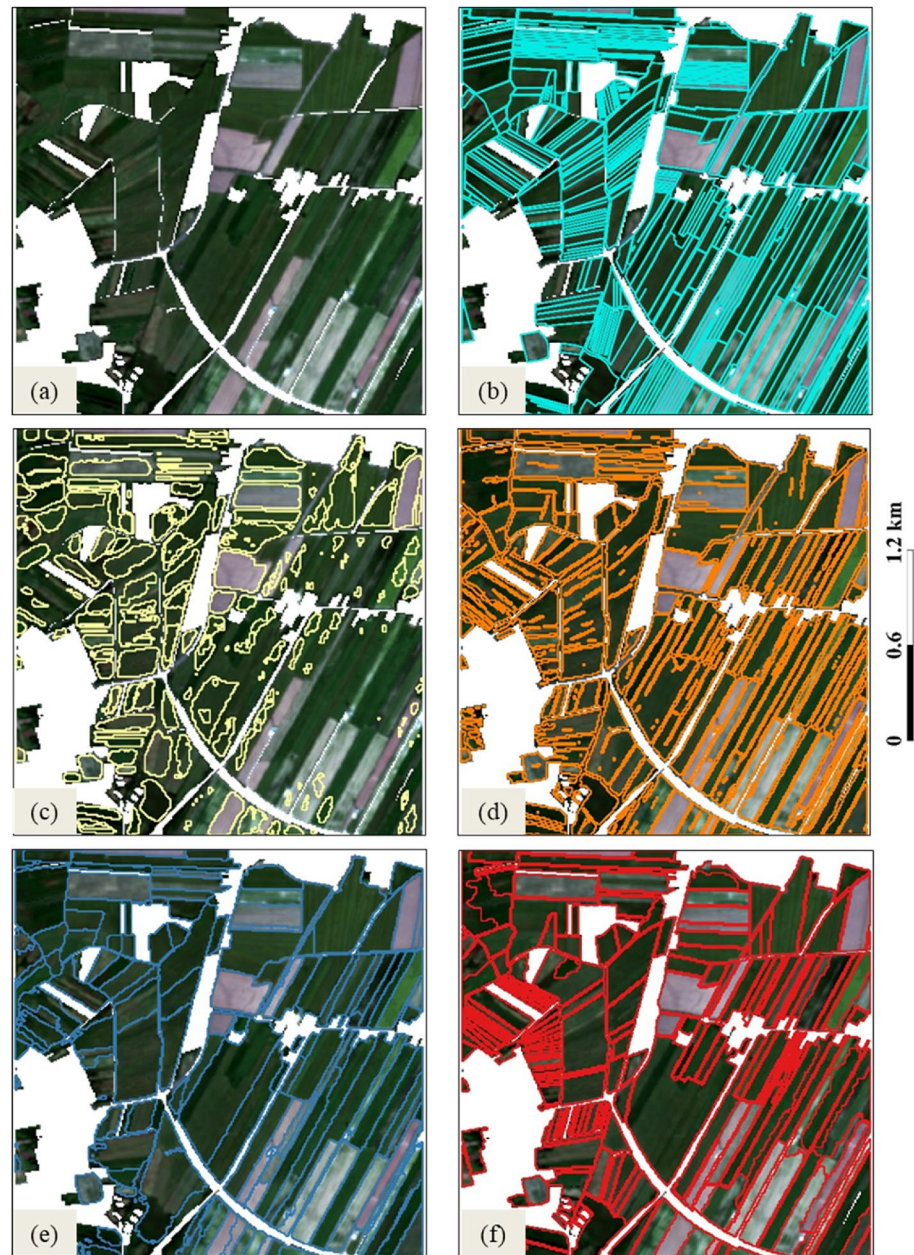

The performance of the MRS algorithm could largely be linked to the direct use of the reference GSAA parcels to guide the segmentation process at each test tile. Another factor that might have helped the MRS approach was the creation of the masks from the inwardly buffered land-cover polygons extracted from ATKIS and the subsequent use of those masks to pre-segment the test images. In the study of Tetteh et al. (2020a), where the size of each tile was 10 km × 10 km, the average segmentation accuracy achieved for Lower Saxony was lower than in this study. Largely, this can be attributed to the smaller sizes (2.56 km × 2.56 km) of tiles used in this study. It

was reported by Drăguţ et al. (2019) that the segmentation accuracy achieved by the MRS algorithm inversely correlates with the spatial extent of the input image.

The three DNNs used in this study are supervised methods. The DNNs can be trained on some training images, the trained model can be saved, and then the saved model can be subsequently applied to segment unseen (test) images. This concept does not apply to the MRS algorithm because it is designed for unsupervised segmentation, hence no training is required. Therefore, to ensure a fair comparison between the DNNs and the MRS algorithm, we used the SSO approach to optimise the MRS parameters. The optimal

Fig. 8 The segmentation result obtained at a tile with medium compactness. **a** The masked MMC image of the tile, **b** the GSAA parcels (cyan outlines) overlaid on the masked MMC image, **c** the segmentation result (yellow outlines) of Mask R-CNN, **d** the segmentation result (orange outlines) of U-Net, **e** the segmentation result (blue outlines) of FracTAL ResUNet, and **f** the segmentation result (red outlines) of the optimised MRS
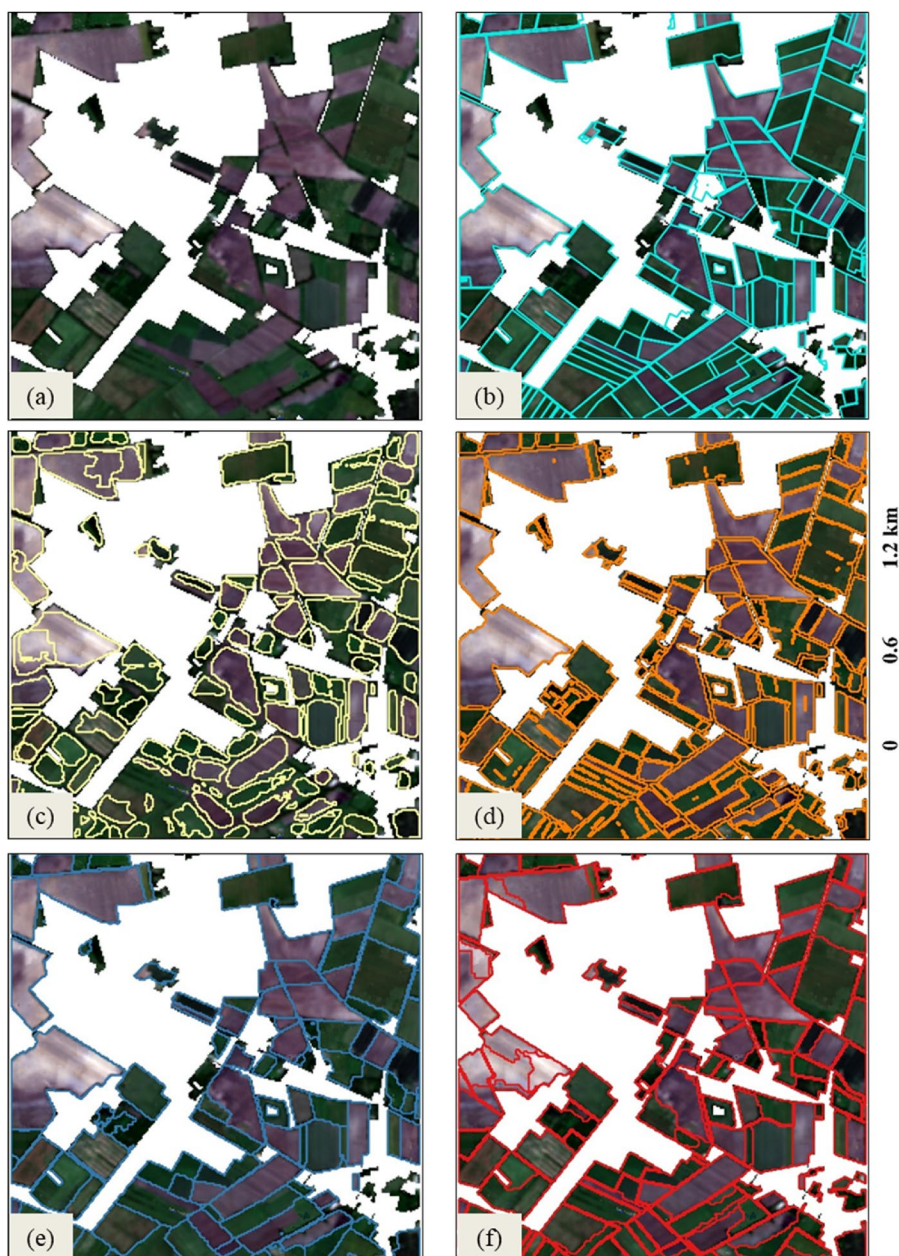
MRS parameter combination established for a tile with the SSO approach is specific to that tile, and hence cannot be transferred in space to a different tile. This informed our decision to optimise the MRS parameters at each test tile. Optimising the MRS parameters at the training tiles and then transferring the parameters to the test tiles would produce very poor segmentation results because in many instances tiles with close spatial proximity or the same SF class have completely different MRS parameters. Unlike the MRS algorithm, a trained DNN model can be transferred in space to segment unseen images.

In their review paper, Persello et al. (2022) highlighted how DNNs and earth observation data can be applied to support the SDGs of the UN. Specific to the second goal of the SDGs, fashioning out policies to sustainably achieve food security will require agricultural lands to be monitored at regional, national, and global scales. The ability of DNNs, particularly FracTAL ResUNet, once trained on sample images to generalise well on unseen images (Waldner et al. 2021; Wang et al. 2022), opens up the possibility to delineate agricultural fields on a large scale even in areas where reference data are unavailable.

Fig. 9 The segmentation result obtained at a tile with high compactness. **a** The masked MMC image of the tile, **b** the GSAA parcels (cyan outlines) overlaid on the masked MMC image, **c** the segmentation result (yellow outlines) of Mask R-CNN, **d** the segmentation result (orange outlines) of U-Net, **e** the segmentation result (blue outlines) of FracTAL ResUNet, and **f** the segmentation result (red outlines) of the optimised MRS
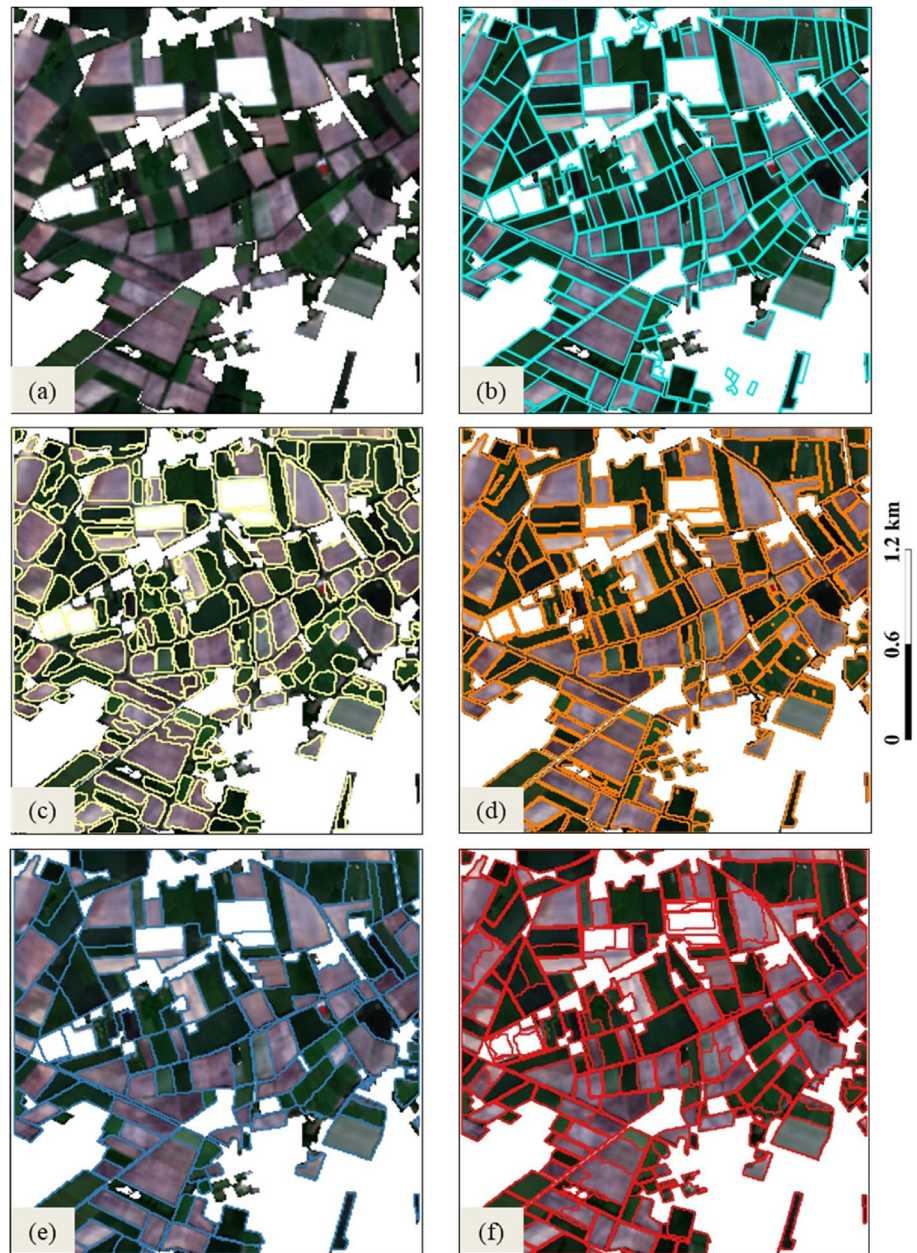
Table 2 The F-score and IoU values obtained by each method at the three test tiles, respectively, shown in Figs. 7, 8, and 9

Figure	Shape factor class	F-score				IoU			
		Mask R-CNN	U-Net	FracTAL ResUNet	MRS	Mask R-CNN	U-Net	FracTAL ResUNet	MRS
Figure 7	Low compactness	0.470	0.758	0.801	0.760	0.307	0.611	0.668	0.612
Figure 8	Medium compactness	0.664	0.791	0.842	0.774	0.498	0.655	0.727	0.632
Figure 9	High compactness	**0.756**	**0.838**	**0.869**	**0.860**	**0.608**	**0.722**	**0.769**	**0.754**

The best performance for each method is highlighted in bold font

6 Conclusions

To determine the optimal method for delineating agricultural fields from Sentinel-2 images acquired in Lower Saxony (Germany), we evaluated three state-of-the-art deep neural networks (DNNs), namely Mask R-CNN, U-Net, and FracTAL ResUNet against an optimised multiresolution segmentation (MRS) approach. Based on the agricultural parcels declared by farmers within the European Common Agricultural Policy (CAP) framework, the segmentation results generated by each method were evaluated using two main metrics namely F-score and intersection over union (IoU). With an average F-score of 0.808 and IoU of 0.683, FracTAL ResUNet combined with a post-processing approach called hierarchical watershed segmentation generated the best segmentation results. FracTAL ResUNet was closely followed by the optimised MRS approach with an average F-score of 0.805 and IoU of 0.678.

For researchers working on the large-scale object-based mapping of agricultural land-use types from satellite images, this study can serve as a guide regarding which segmentation method to use for the delineation of agricultural fields. Based on the outcome of this study, for large-scale segmentation of agricultural fields, we recommend the use of FracTAL ResUNet. Once the FracTAL ResUNet model has been trained, it generalises very well and can be transferred in space to effectively segment unseen images. This is in sharp contrast to the optimised MRS approach, which is not transferable in space. To segment any unseen image with the optimised MRS approach, reference data are always required.

Future work would focus on: (1) combining FracTAL ResUNet and the hierarchical watershed segmentation algorithm to delineate all agricultural fields in Germany based on multitemporal Sentinel-2 images, (2) the use of Bayesian optimisation to optimise the hyperparameters of FracTAL ResUNet and the hierarchical watershed segmentation algorithm, and (3) testing the temporal transferability of FracTAL ResUNet from one year to another year of interest.
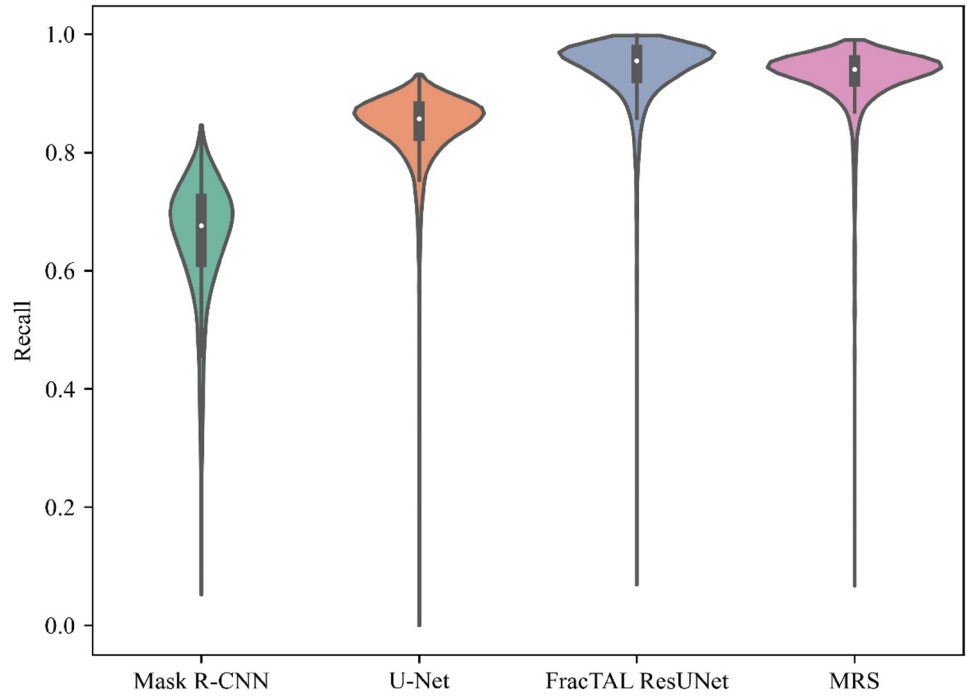
Appendix

See Figs. 10, 11, 12 and 13.

Fig. 10 Violin plots showing the distribution of the precision values for each method for all test tiles
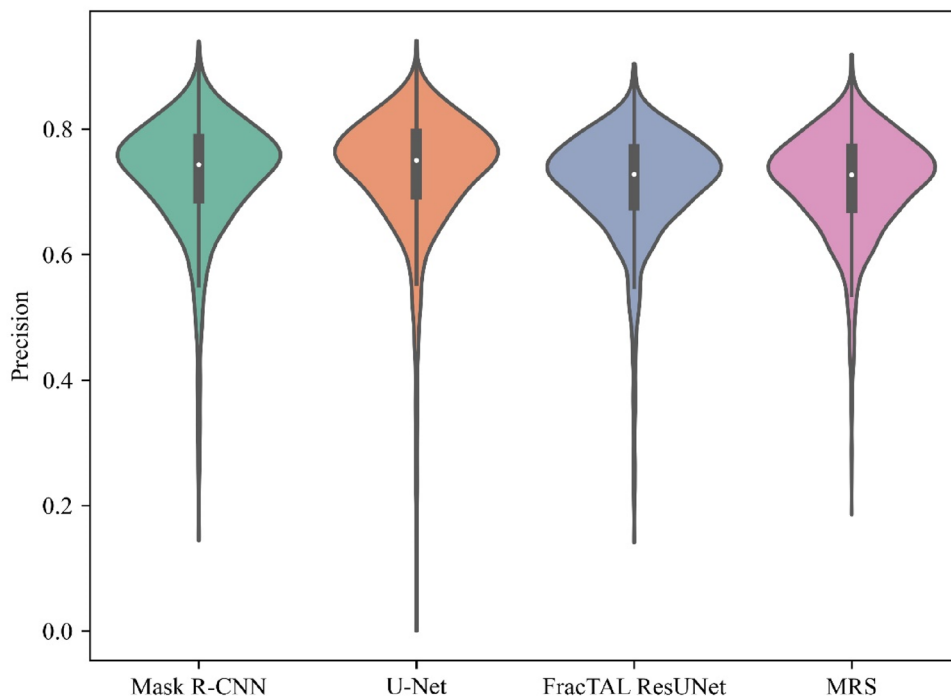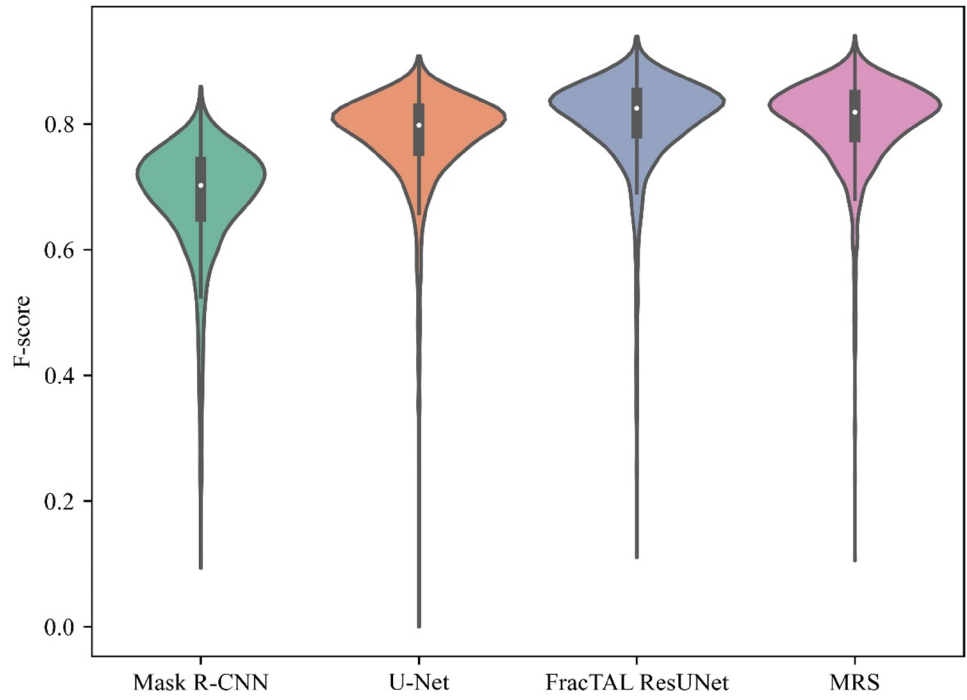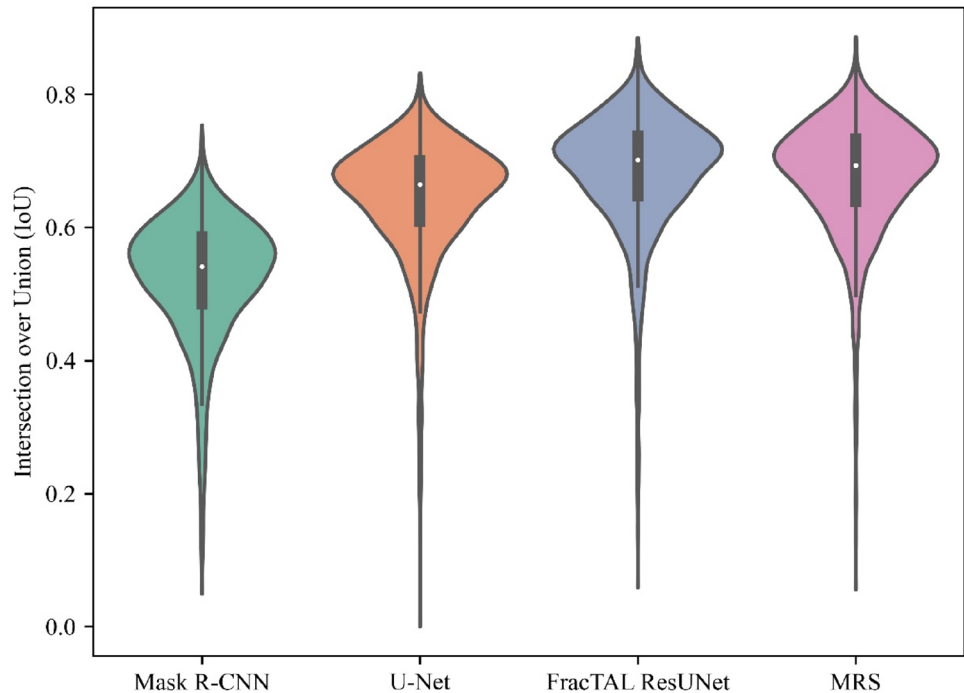
Fig. 11 Violin plots showing the distribution of the recall values for each method for all test tiles

Fig. 12 Violin plots showing the distribution of the F-score values for each method for all test tiles

Fig. 13 Violin plots showing the distribution of the intersection over union (IoU) values for each method for all test tiles

Data availability The Sentinel-2 images and the GSAA dataset of Lower Saxony are publicly and freely available. The ATKIS dataset is proprietary, hence cannot be shared by the authors.

Declarations

Conflict of Interest The authors declare no conflicts of interest.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X (2016) TensorFlow: a system for large-scale machine learning. USENIX

Abdulla W (2017) Mask R-CNN for object detection and segmentation. GitHub

Akcay O, Avsar EO, Inalpulat M, Genc L, Cam A (2018) Assessment of segmentation parameters for object-based land cover classification using color-infrared imagery. ISPRS Int J Geo-Inf 7(11):424. https://doi.org/10.3390/ijgi7110424

Atzberger C (2013) Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens 5(12):949–981. https://doi.org/10.3390/rs5020949

Aung HL, Uzkent B, Burke M, Lobell D, Ermon S (2020) Farm parcel delineation using spatio-temporal convolutional networks. 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). IEEE, Seattle, WA, USA, pp 340–349

Baatz M, Schäpe A (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl J, Blaschke T, Griesebner G (eds) Angewandte geographische informations-verarbeitung XII. Wichmann Verlag, Karlsruhe, pp 12–23

Belgiu M, Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens Environ 204:509–523. https://doi.org/10.1016/j.rse.2017.10.005

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65(1):2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blickensdörfer L, Schwieder M, Pflugmacher D, Nendel C, Erasmi S, Hostert P (2022) Mapping of crop types and crop sequences

with combined time series of sentinel-1, sentinel-2 and landsat 8 data for Germany. Remote Sens Environ 269:112831. https://doi.org/10.1016/j.rse.2021.112831

Boryan C, Yang Z, Mueller R, Craig M (2011) Monitoring US agriculture: the US department of agriculture, national agricultural statistics service. Cropland Data Layer Program Geocarto Int 26(5):341–358. https://doi.org/10.1080/10106049.2011.562309

Castillejo-González IL, López-Granados F, García-Ferrer A, Peña-Barragán JM, Jurado-Expósito M, de la Orden MS, González-Audicana M (2009) Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. Comput Electron Agric 68(2):207–215. https://doi.org/10.1016/j.compag.2009.06.004

Chen T, Li M, Li Y, Lin M, Wang N, Wang M, Xiao T, Xu B, Zhang C, Zhang Z (2015) MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. Carnegie Mellon University

Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017a) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40:834–848

Chen L-C, Papandreou G, Schroff F, Adam H (2017b) Rethinking atrous convolution for semantic image segmentation. Harvard University

Commission E (2017) CAP explained: direct payments for farmers 2015–2020. Publications Office, LU

d'Andrimont R, Verhegghen A, Lemoine G, Kempeneers P, Meroni M, van der Velde M (2021) From parcel to continental scale—a first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. Remote Sens Environ 266:112708. https://doi.org/10.1016/j.rse.2021.112708

Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

Diakogiannis FI, Waldner F, Caccetta P (2021) Looking for change? Roll the dice and demand attention. Remote Sens 13(18):3707. https://doi.org/10.3390/rs13183707

Drăguţ L, Belgiu M, Popescu G, Bandura P (2019) Sensitivity of multiresolution segmentation to spatial extent. Int J Appl Earth Obs Geoinformation 81:146–153. https://doi.org/10.1016/j.jag.2019.05.002

Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, Mueller ND, O'Connell C, Ray DK, West PC, Balzer C, Bennett EM, Carpenter SR, Hill J, Monfreda C, Polasky S, Rockström J, Sheehan J, Siebert S, Tilman D, Zaks DPM (2011) Solutions for a cultivated planet. Nature 478(7369):337–342. https://doi.org/10.1038/nature10452

Frantz D (2019) FORCE—landsat + sentinel-2 analysis ready data and beyond. Remote Sens 11(9):1124. https://doi.org/10.3390/rs11091124

Gao Y, Mas JF, Kerle N, Pacheco JAN (2011) Optimal region growing segmentation and its effect on classification accuracy. Int J Remote Sens 32(13):3747–3763. https://doi.org/10.1080/01431161003777189

García-Pedrero A, Gonzalo-Martín C, Lillo-Saavedra M (2017) A machine learning approach for agricultural parcel delineation through agglomerative segmentation. Int J Remote Sens 38(7):1809–1819. https://doi.org/10.1080/01431161.2016.1278312

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. https://doi.org/10.1109/ACCESS.2019.2950371

Gilbertson JK, Kemp J, van Niekerk A (2017) Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. Comput Electron Agric 134:151–159. https://doi.org/10.1016/j.compag.2016.12.006

Graesser J, Ramankutty N (2017) Detection of cropland field parcels from Landsat imagery. Remote Sens Environ 201:165–180. https://doi.org/10.1016/j.rse.2017.08.027

He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. IEEE

He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. Springer International Publishing

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. 2017 IEEE international conference on computer vision (ICCV). IEEE, pp 2980–2988

Jaccard P (1901) Etude de la distribution florale dans une portion des Alpes et du Jura. Bull Soc Vaudoise Sci Nat 37:547–579. https://doi.org/10.5169/seals-266450

Ji CY (1996) Delineating agricultural field boundaries from TM imagery using dyadic wavelet transforms. ISPRS J Photogramm Remote Sens 51(6):268–283. https://doi.org/10.1016/0924-2716(95)00017-8

Jong M, Guan K, Wang S, Huang Y, Peng B (2022) Improving field boundary delineation in ResUNets via adversarial deep learning. Int J Appl Earth Obs Geoinf 112:102877. https://doi.org/10.1016/j.jag.2022.102877

Kattenborn T, Leitloff J, Schiefer F, Hinz S (2021) Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J Photogramm Remote Sens 173:24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. Scientific Research Publishing

Kotaridis I, Lazaridou M (2021) Remote sensing image segmentation advances: a meta-analysis. ISPRS J Photogramm Remote Sens 173:309–322. https://doi.org/10.1016/j.isprsjprs.2021.01.020

Li P, Xiao X (2007) Multispectral image segmentation by a multichannel watershed-based approach. Int J Remote Sens 28(19):4429–4452. https://doi.org/10.1080/01431160601034910

Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. IEEE

Liu D, Xia F (2010) Assessing object-based classification: advantages and limitations. Remote Sens Lett 1(4):187–194. https://doi.org/10.1080/01431161003743173

Long J, Li M, Wang X, Stein A (2022) Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images. Int J Appl Earth Obs Geoinformation 112:102871. https://doi.org/10.1016/j.jag.2022.102871

Luo C, Qi B, Liu H, Guo D, Lu L, Fu Q, Shao Y (2021) Using time series sentinel-1 images for object-oriented crop classification in google earth engine. Remote Sens 13(4):561. https://doi.org/10.3390/rs13040561

Lv Y, Zhang C, Yun W, Gao L, Wang H, Ma J, Li H, Zhu D (2020) The delineation and grading of actual crop production units in modern smallholder areas using RS data and mask R-CNN. Remote Sens 12(7):1074. https://doi.org/10.3390/rs12071074

Ma L, Li M, Ma X, Cheng L, Du P, Liu Y (2017) A review of supervised object-based land-cover image classification. ISPRS J Photogramm Remote Sens 130:277–293. https://doi.org/10.1016/j.isprsjprs.2017.06.001

Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA (2019) Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J Photogramm Remote Sens 152:166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Masoud KM, Persello C, Tolpekin VA (2020) Delineation of agricultural field boundaries from sentinel-2 images using a novel super-resolution contour detector based on fully convolutional

networks. Remote Sens 12(1):59. https://doi.org/10.3390/rs120 10059

Maxwell AE, Warner TA, Guillén LA (2021) Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: literature review. Remote Sens 13(13):2450. https://doi.org/10.3390/rs13132450

Meyer L, Lemarchand F, Sidiropoulos P (2020) A deep learning architecture for batch-mode fully automated field boundary detection. Int Arch Photogramm Remote Sens Spat Inf Sci XLIII-B3-2020:1009–1016. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1009-2020

Möller M, Lymburner L, Volk M (2007) The comparison index: a tool for assessing the accuracy of image segmentation. Int J Appl Earth Obs Geoinformation 9(3):311–321. https://doi.org/10.1016/j.jag.2006.10.002

Nasrallah A, Baghdadi N, Mhawej M, Faour G, Darwish T, Belhouchette H, Darwich S (2018) A novel approach for mapping wheat areas using high resolution sentinel-2 images. Sensors 18(7):2089. https://doi.org/10.3390/s18072089

North HC, Pairman D, Belliss SE (2019) Boundary delineation of agricultural fields in multitemporal satellite imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 12(1):237–251. https://doi.org/10.1109/JSTARS.2018.2884513

Pandey C, Sethy PK, Behera SK, Vishwakarma J, Tande V (2022) Chapter 1—Smart agriculture: technological advancements on agriculture—a systematical review. In: Poonia RC, Singh V, Nayak SR (eds) Deep learning for sustainable agriculture. Academic Press, pp 1–56

Persello C, Tolpekin VA, Bergado JR, de By RA (2019) Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. Remote Sens Environ 231:111253. https://doi.org/10.1016/j.rse.2019.111253

Persello C, Wegner JD, Hänsch R, Tuia D, Ghamisi P, Koeva M, Camps-Valls G (2022) Deep learning and earth observation to support the sustainable development goals: current approaches, open challenges, and future opportunities. IEEE Geosci Remote Sens Mag 10(2):172–200. https://doi.org/10.1109/MGRS.2021.3136100

Polsby DD, Popper R (1991) The third criterion: compactness as a procedural safeguard against partisan gerrymandering. Social Science Research Network, Rochester, NY

Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical image computing and computer-assisted intervention—MICCAI 2015. Springer International Publishing, Cham, pp 234–241

Rydberg A, Borgefors G (2001) Integrated method for boundary delineation of agricultural fields in multispectral satellite images. IEEE Trans Geosci Remote Sens 39(11):7

Sun Z, Di L, Fang H, Burgess A (2020) Deep learning classification for crop types in North Dakota. IEEE J Sel Top Appl Earth Obs Remote Sens 13:2200–2213. https://doi.org/10.1109/JSTARS.2020.2990104

Taravat A, Wagner MP, Bonifacio R, Petit D (2021) Advanced fully convolutional networks for agricultural field boundary detection. Remote Sens 13(4):722. https://doi.org/10.3390/rs13040722

Tetteh GO, Gocht A, Conrad C (2020a) Optimal parameters for delineating agricultural parcels from satellite images based on supervised Bayesian optimization. Comput Electron Agric 178:105696. https://doi.org/10.1016/j.compag.2020.105696

Tetteh GO, Gocht A, Schwieder M, Erasmi S, Conrad C (2020b) Unsupervised parameterization for optimal segmentation of agricultural parcels from satellite images in different agricultural landscapes. Remote Sens 12(18):3096. https://doi.org/10.3390/rs12183096

Tetteh GO, Gocht A, Erasmi S, Schwieder M, Conrad C (2021) Evaluation of sentinel-1 and sentinel-2 feature sets for delineating agricultural fields in heterogeneous landscapes. IEEE Access 9:116702–116719. https://doi.org/10.1109/ACCESS.2021.3105903

Trimble Germany GmbH (2019) eCognition developer 9.5.0 reference book. Trimble Germany GmbH, Germany

Turker M, Kok EH (2013) Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. ISPRS J Photogramm Remote Sens 79:106–121. https://doi.org/10.1016/j.isprsjprs.2013.02.009

United Nations (2015) Transforming our world: the 2030 agenda for sustainable development. United Nations, New York, NY

Wagner MP, Oppelt N (2020) Extracting agricultural fields from remote sensing imagery using graph-based growing contours. Remote Sens 12(7):1205. https://doi.org/10.3390/rs12071205

Waldner F, Diakogiannis FI (2020) Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sens Environ 245:111741. https://doi.org/10.1016/j.rse.2020.111741

Waldner F, Diakogiannis FI, Batchelor K, Ciccotosto-Camp M, Cooper-Williams E, Herrmann C, Mata G, Toovey A (2021) Detect, consolidate, delineate: scalable mapping of field boundaries using satellite images. Remote Sens 13(11):2197. https://doi.org/10.3390/rs13112197

Wang S, Waldner F, Lobell DB (2022) Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision. Remote Sens 14(22):5738. https://doi.org/10.3390/rs14225738

Watkins B, van Niekerk A (2019) A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery. Comput Electron Agric 158:294–302. https://doi.org/10.1016/j.compag.2019.02.009

Weiss M, Jacob F, Duveiller G (2020) Remote sensing for agricultural applications: a meta-review. Remote Sens Environ 236:111402. https://doi.org/10.1016/j.rse.2019.111402

Yan L, Roy DP (2014) Automated crop field extraction from multi-temporal web enabled landsat data. Remote Sens Environ 144:42–64. https://doi.org/10.1016/j.rse.2014.01.006

Yang R, Ahmed ZU, Schulthess UC, Kamal M, Rai R (2020) Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: a case study from Bangladesh. Remote Sens Appl Soc Environ 20:100413. https://doi.org/10.1016/j.rsase.2020.100413

You N, Dong J, Huang J, Du G, Zhang G, He Y, Yang T, Di Y, Xiao X (2021) The 10-m crop type maps in Northeast China during 2017–2019. Sci Data 8(1):41. https://doi.org/10.1038/s41597-021-00827-9

Zhang H, Liu M, Wang Y, Shang J, Liu X, Li B, Song A, Li Q (2021) Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. Int J Appl Earth Obs Geoinf 105:102557. https://doi.org/10.1016/j.jag.2021.102557

Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. IEEE