LOGIT-BASED SUPERPIXEL SEMANTIC SEGMENTATION OF IMAGES FOR PRECISION AGRICULTURE

Corneliu FLOREA[1], Alexandru BURGHIU[2], Mihai IVANOVICI[3]

In this work we approach the problem of remote sensing image segmentation using a classical approach: the image is first segmented and, subsequently, each segment is labeled using a classifier. For segmentation, we rely on a superpixel framework and several methods are evaluated. For the classifier, again, several state-of-the-art algorithms are tested and performances are compared. The best performing method is obtained by a modified SEED superpixel algorithm with boosted trees for classification. The evaluation is carried out on the Agriculture-Vision database and the results are encouraging.

Keywords: Superpixel, Logit, Boosted trees, Remote Sensing, Semantic segmentation, Precision Agriculture.

1. Introduction

Rapid global population growth and increasing demand for food production require effective monitoring and management of agricultural resources. Aerial imagery plays an essential role in providing valuable information for various agricultural applications such as crop yield estimation, precision agriculture and land use planning. Accurate segmentation and classification of agricultural regions in aerial imagery is important for these applications as it offers a cost-effective alternative to manual inspection, which requires experts to travel in person. However, the complexity and variability of agricultural landscapes pose significant challenges in developing robust and efficient region segmentation and classification methods. For instance, when attempting to delineate weed distributions in aerial images of farmland, the algorithm needs to accurately recognize and differentiate between sparse weed clusters that vary greatly in shape and size. We are using the Agriculture Vision dataset, which compared to previous agricultural image collection originates in larger resolution images, that record details up to 10 cm per pixel (cm/px) and precise annotations from professional agronomists with a strict quality assurance process.

[1] Professor, National University of Science and Technology POLITEHNICA of Bucharest, Romania, e-mail: corneliu.florea@upb.ro
[2] Student, National University of Science and Technology POLITEHNICA of Bucharest, Romania, e-mail: alexandru.burghiu@stud.etti.upb.ro;
[3] Professor., Transilvania University of Brasov, Romania, e-mail: mihai.ivanovici@unitbv.ro

In this paper we approach the problem of semantic segmentation of remotely-sensed images with application in precision agriculture using a "rather classical" solution, in the sense that we avoid the heavy computational deep learning models, which have dominated the solutions in the recent period. Given an image acquired from a camera mounted on a aerial drone, recording a specific location and land cover, the purpose is to label each pixel into a relevant class. An overview of the method may be seen in Figure **1**. The method is based upon a classical approach where, first, the image is segmented, and each segment is further classified with the predictor trained on the appropriate database. For the segmentation, we actually over-segment the image based on the superpixel approach. For classification we rely on a boosted ensemble of trees.

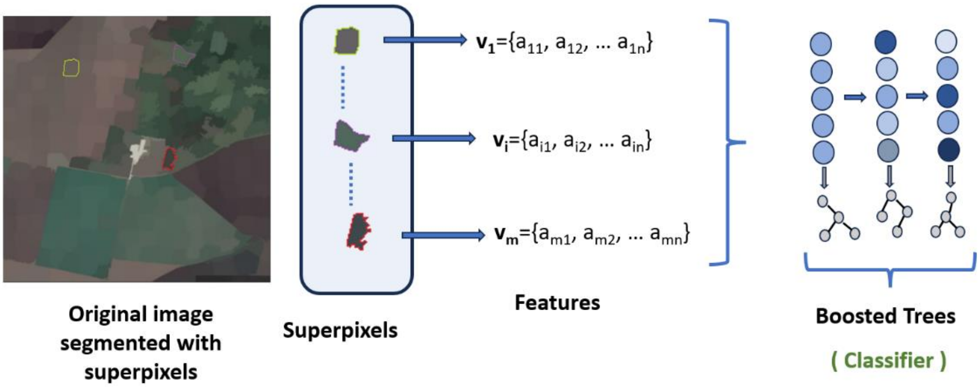


Fig 1. A schematic overview of the proposed method.

Overall, the contribution of this paper is at the method level, where we propose a combination of features (derived from superpixels) and classifier that has not been used for semantic segmentation in general, and for agriculture aerial images in particular. The remainder of the paper is organized as follows: since the main parts of the method are based on segmentation and the overall framework, in the next section we review the most relevant works into these directions. Section 3 is dedicated for presenting relevant aspects of the method, while section 4 is about implementation and results. The paper ends with conclusions.

2. Previous works

The embraced method uses superpixels and addresses the problem of semantic segmentation of remote sensing images with application to precision agriculture, more specifically for the cartography of the agricultural crop. The existing approaches are discussed in following paragraphs.

Superpixel. The concept of superpixel has been introduced by Ren and Malik in 2003, [1] and represents an over-segmentation of an image formed by gathering perceptually similar pixels. The superpixel is a validated low-level representation of image as it groups pixels that share similar colors or other low-level properties such texture.

Many notable implementations of the superpixel concept exist, starting with Normalized Cuts [1], Felzenszwalb and Huttenlocher (FH) [2] and the entropy rate superpixels (ERS) [3]. In summary, they are clustering methods and the difference lays in the clustering objective function, in the data space exploration or in the similarity measure and threshold. Some popular algorithms include SLIC [4], energy driven sampling - SEEDS [5] and Linear Spectral Clustering - (LSC) [6]. SLIC [4] initializes the centers into a regular grid and grows the segments by clustering pixels around the center with a distance-based similarity measure. SEEDS [5] uses an energy-based minimization while LSC [6] assumes spectral decomposition and limitations in variance. Stutz et al. [7] surveyed various methods and provide details that differentiate algorithms.

Superpixels survived in the deep leaning dominated era. A common application lies in active learning for semantic segmentation. For instance, SEAL (Segmentation Affinity Loss) [8] used deep convolutional networks to learn the features for superpixel generation. In an active learning framework, Cai et al. [9] and respectively Kim et al. [10] used SEEDS, followed by adapted merging to pre-segment the image, so to minimize the user effort while annotating for future supervised segmentation. Overall, various alternatives for clustering pixels in superpixels have been proposed but is not based on logit representation of colors.

Aerial and satellite image semantic segmentation. Semantic segmentation refers to the process where each pixel of the is associated to an object and the object is labeled. While in the recent years this field is dominated by the deep network approach, even that tremendous effort has put into miniaturization and efficientization of the calculus, deep networks still require a significant amount of resources, especially access to a graphical card; furthermore, they do require days for training. In contrast, non deep methods while they may be inferior as performance, they can be easily applied to larger resolution images and require significant lesser resources for training and prediction. Recently, a number of works used superpixels in remote sensing image segmentation.

For instance, Zhang et al. [11] proposed a multi-scale, spectral based superpixels followed by optimization to segment coastal city images. Cheng et al. [12] notes the limitation of deep CNN in dealing with variable sized images and proposes a superpixel-based graph convolutional network for PolSAR (Polarimetric Synthetic Aperture Radar) image classification. Ma et al. [13] incorporate the superpixel generation and merging steps into, by means of differentiable loss function into an end-to-end trainable deep network. Geng et al. [14] successfully extended the superpixel hypergraph neural network to segmentation of Polarimetric SAR image.

To conclude, the faster computation and the ability to work with variable sized images made the superpixel concept to be useful. It had been used alone or in conjunction with deep networks, in the area of semantic segmentation of remote sensing images, even in the very recent times.

2. Method

In summary, our method consists of the following steps: (I) over-segment the RGB image with enhanced, logit based, SEEDS superpixel algorithm. (II) describe each superpixel with a set of attributes (III) given the database that is split into training and testing, train a classifier to correctly label each superpixel; (IV) post process the labelled image.

2.1. Superpixel Segmentation

For the superpixel segmentation, we take as baseline algorithm the SEEDS [5] solution and enhance it. We recall that superpixels aim to group similar pixels, based on homogeneity or other criteria. Inspired by the SLIC [4] superpixel, the SEEDS also start with a regular grid of prototypes (set empirically to 500 for our experiments), that it assigns neighboring pixels based on color to each prototype. In subsequent steps, the superpixels are updated at pixel level and at block level.

The update uses an energy minimization principle, where it is considered the square of the probability for two colors to be in the same group, as a homogeneity criteria. The probabilities are taken from the 3D RGB color histogram.

The original SEEDS algorithm defines the following measure to be the minimizable color energy:

$$H(s) = \sum_k \Psi(A_k) \tag{1}$$

where $\Psi(A_k)$ is a function enforcing the colors to be concentrated in one or few values. A_k is the building superpixel. The original function proposed [5] was based on the square of probabilities in the superpixel histogram (i.e GINI index). We have found out that a logit version works slightly better:

$$\Psi(A_k) = \sum_k \frac{e^{-c(A_k)}}{\sum_j e^{-c(A_j)}} \tag{2}$$

The eq.(2) shows the main technical contribution of the proposed paper. We removed the boundary term from the original SEEDS method, but we adhere to the hill climbing principle (where maximization direction is found my going in the direction marked by the derivative gradient) in the pixel update and block update. At the end of the superpixel algorithm, it results the superpixel partition: the pixels in each image $i \in I$ are separated into a set $S_0(i)$ of superpixels, and to produce a base segmentation $S_0 := S_i \in I$. The number of superpixels in an image is variable, due to block fusion.

2.2 Features

Each superpixel is described by a feature vector of fixed size. In our implementation, the descriptor is made out of the mean for each color plane (R,G,B), the variances on the same color plane and the histogram of gradient orientation [15] (HOG) with 6 bins. Overall, the feature vector has length 12 and multiple superpixels form multiple instances.

2.3. Classifier

While many non-deep classifiers have been proposed in the literature, our version for this application is an ensemble of boosted trees. The boosting paradigm, has been introduced for binary classification, but it has been extended to multiple classes by the SAMME (Stagewise Additive Modelling with a Multi-class Exponential loss function) [16] algorithm. The boosting, intuitively, refers to building classifiers from an ensemble in a sequential manner, where the current one tries to compensate for the errors of the previous classifiers. A multiclass boosting and arcing procedure was shown to work with high capacity classifiers such as SVM [17], but is as efficient with smaller ones and it is faster.

Intuitive motivation of the proposed method lies in the fact that, here, we use strong attributes, compared to weak ones, there [17]. The procedure is as following:

1. Let be given a training database with *n* examples
2. Initialize the observation weights : $\mathrm{W}_{\mathrm{i}}^{(1)} = 1,\ \mathrm{i} \in \{1, \dots, \mathrm{n}\}$
3. For $m=1:M$
   a. Randomly select a classifier $\mathcal{T}_{\wp}^{(m)} with\ \mathrm{p} \in \{1 \dots \mathrm{Q}\}$ Also select $\mathrm{X}_{(\mathrm{p})}$
   b. Select a random bootstrap sample of the data
   c. Fit the chosen classifier $\mathcal{T}_{\wp}^{(m)}$ to the training data using weights $\mathrm{W}^{(\mathrm{m})}$
   d. Compute the recognition error:

$$\varepsilon_m = \left(\sum_{i=1}^{n} W_i^{(m)} \left[c_i \neq \mathcal{T}_{\wp}^{(m)}(x_i)\right]\right) / \sum_{i=1}^{n} W_i^{(m)} \tag{3}$$

   e. Compute the update:

$$\alpha^{(m)} = \min\left(\log\frac{1-\varepsilon_m}{\varepsilon_m} + \log(K-1)\,, \alpha_{max}\right) \tag{4}$$

f. Set the new weights, to emphasize examples poorly labeled:

$$W_i^{(m+1)} \leftarrow W_i^{(m)} \cdot \beta^{\alpha^{(m)}\left[c_i \neq \mathcal{T}_{\wp}^{(m)}(x_i)\right]} \tag{5}$$

4. The final classifier is given by:

$$C(X) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)}\left[\mathcal{T}_{\wp}^{(m)}\left(X_{(p)}\right) = k\right] \tag{6}$$

In this procedure, *[$\boldsymbol{a}_i$ = $\boldsymbol{b}_i$]* is the Iverson bracket notation for the number of occurrences. For the proposed solution, $\alpha_{max} = 10$ and β is linearly decaying, with respect to the iteration from 6 to 3, while Q = 2. The trees used as individual predictors have been cut early and prediction uses probability of being in a class. Next, the boosted ensemble classifier is able to provide probabilities for each class due to the number of individual classifiers working and aggregation of individual probabilities. The original SAMME work [16] offers strong theoretical justification for the convergence of the algorithm, justification which stands to this version too.

3.4. Post-processing

This step takes place after prediction, and it has been introduced to limit the noisy labelling and to adapt the solution to the database.

As one can see in Figure 2, annotations are not precisely at pixel level, but with polygons, thus being rather vague on the edges; yet regions are compact. As we further discuss database acquisition, there is physical motivation for these observations.

In terms of prior (i.e. at training stage), for each class it has been determined the minimum and maximum area of the annotated shapes. Furthermore, on the validation subset, again, for each class, we have determined a pair of thresholds with respect to the probability provided by the classifier, as a lower confidence and an upper confidence values.

After prediction, the mask for each class has been taken separately and processed. Each compact area that was found to be too small (given the minimum size) was removed. For areas large enough, starting from superpixels passing the high confidence threshold, they have been merged with neighboring ones that pass the lower confidence, in a process similar to the one used in the Canny edge detector (i.e. hysteresis-based merging).

4. Experimental evaluation

Implementation. The method has been implemented on Python using standard libraries. The superpixel segmentation code is derived from SEEDS. Due to the nature of the segmentation problem (potentially overlapping masks), for each image, masks for each class have been taken separately.

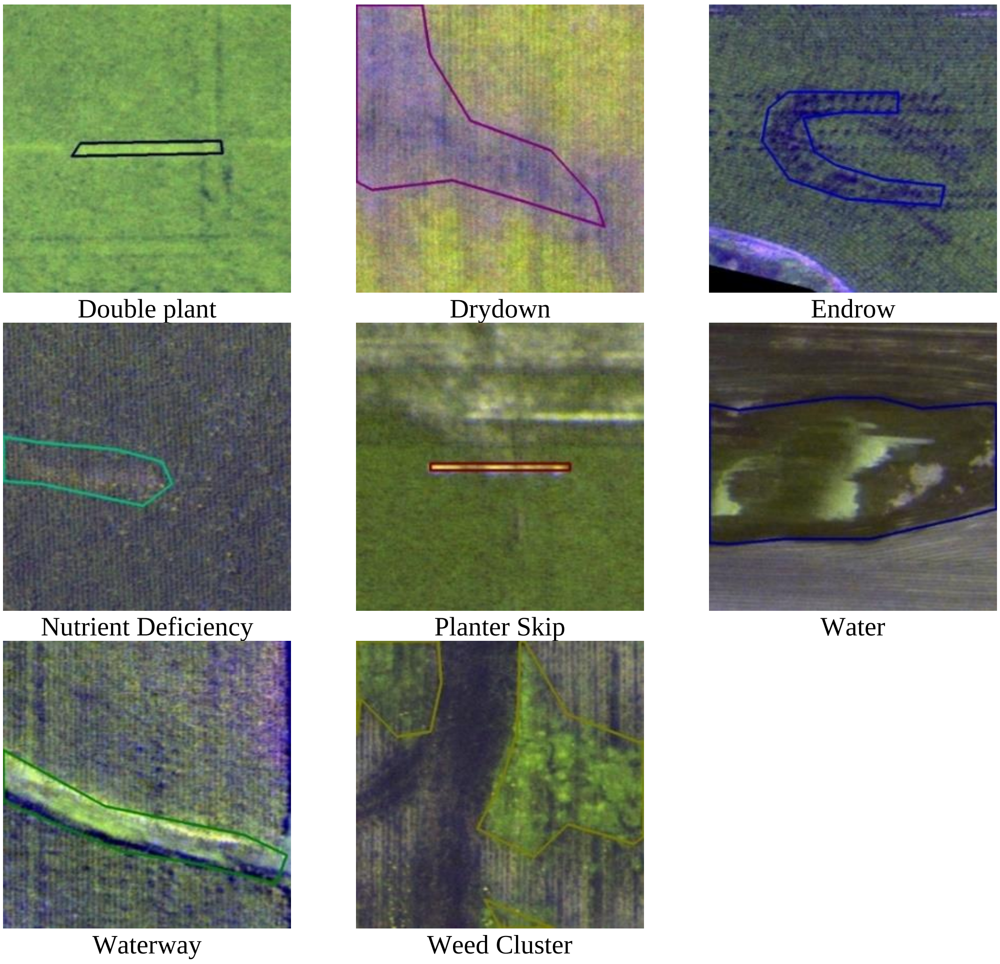


Fig. 2. The classes that are available in Agriculture Vision dataset. In addition, pixels that are in none of the above classes are taken as background.

Database. The method has been evaluated on a subset from Agriculture Vision [18], the 2021 version. For this version, raw images have been collected within a sequence of 2-7 flights from 54 fields from 2017-2020 for a total of 261 full field images. The majority of the images have been acquired with standard DSLR camera (Canon, Nikon D850, Nikon D800E) recording standard RGB images. Additionally, some of these fields contain Sentinel-2 imagery at 10m resolution. Each image was labelled by annotators trained by expert agriculturalists using polygons (for segmentation). In total, there are 94,986 images split as follows: 56,944/18,334/19,708 as train/val/test images. They contain the following classes: “double plant”, “drydown”, “endrow”, “nutrient deficiency”, “planter skip”, “water”, “waterway”, “weed cluster” and “background” and for each image there is a mask for each class. Also, we restrict ourself only to RGB channels from the images.

Examples of images may be seen in Figure 2.

As quality measure, following the introductory work [18], modified Intersection over Union” (mIoU) was used. The modification is due to masks being potentially overlapped:

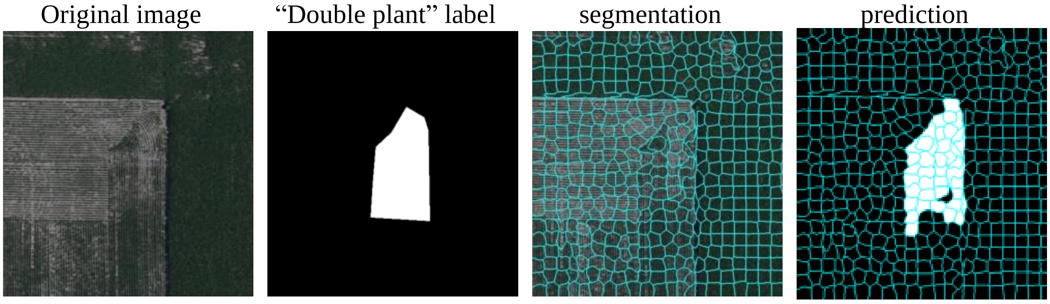
$$mIoU = \frac{1}{c}\sum_{c}\frac{TP_c}{Prediction_c + Target_c - TP_c}$$

where TP – stands for true positive.

Results. Visual examples may be seen in figure 3. From Figure 2 and Figure 3 one may notice that the problem is difficult, with issues being generated both by confusing classes as poor delineation in images. Examining closely examples shown in figure 3, one will notice that the polygonal markings are not very precise and superpixel boundaries follow more curved shapes. Furthermore, variability in the training set of each class limits the prediction accuracy and, as showed in the last row from figure 3, the “weed” greenish characteristic tint is identified in a totally different region, leading to a poor result.

A summary of the results, accumulated over the entire database may be seen in Table 1. We compared the proposed method with various baselines. For the segmentation, part we have considered the standard SLIC [4] method, as well as standard SEEDS [5] as being the most popular choice and respectively the baseline of the proposed method. For the descriptor side we have investigated, in addition, the quantized histogram (a simplification of Color Structure Descriptor [19]).

Original image | “Double plant” label | Superpixel segmentation | “Double plant” prediction



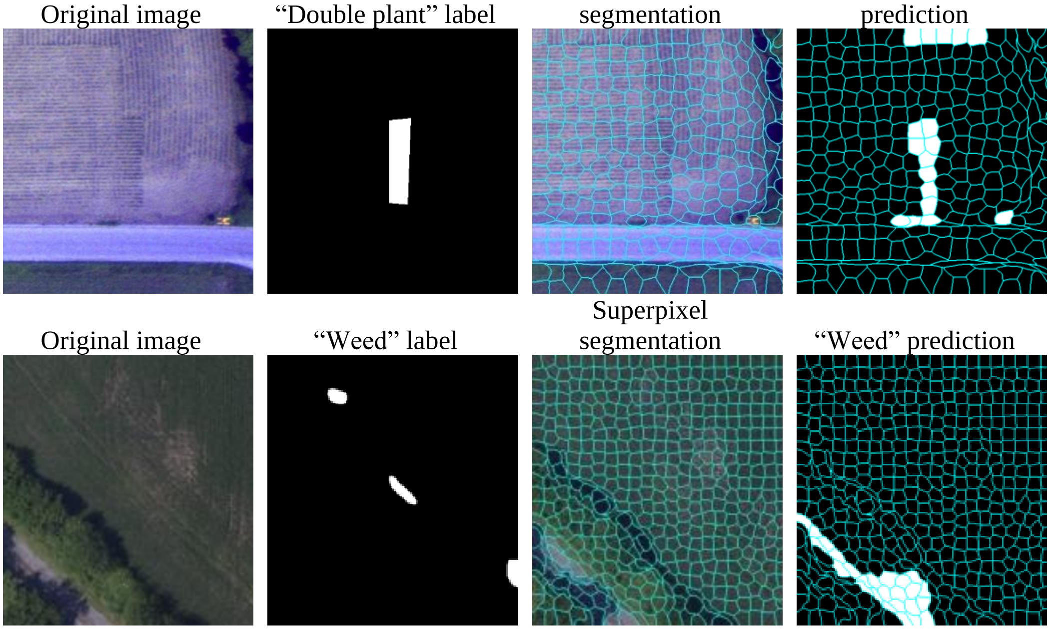


Fig. 3. Example of results. First row: a better example, the middle row an average example, while in the last row, a wrong example with noticeable errors

For the classifier part, following the large evaluation carried by Fernandez-Delgado et al. [20], which identify non-deep machine learning system able to obtain strong performances on various problems, we compared against Support Vector Machine (SVM), with Gaussian kernel and random forest. It is noticeable that the proposed combination reaches the best performance.

In the era of deep learning, a natural question is how much deep methods improve with respect to non-deep methods. In the paper introducing the database, Chui et al. [18] report several solutions. A comparison of the best option from the proposed set with the best deep solution, and, respectively, with an of-the-self solution on class level performance is in Table 2. As one can see, the proposed method is inferior to deep learning solutions, but it is much closer to an off-the-shelf semantic segmentation method, that is such an off-the-shelf reported to a specifically engineered one.

Table 1.

Performance (modified IoU) of various versions of proposed solution. With gray background, we have marked the best solution

Method			**Performance**
Superpixel	**Descriptor**	**Classifier**	**mIoU**
SLIC	Mean, Var, HOG	Random Forest	29.15

SEEDS	Mean, Var, HOG	Random Forest	28.85
Modified SEEDS	Mean, Var, HOG	Random Forest	28.45
Modified SEEDS	Mean, Var, HOG	SVM	25.41
Modified SEEDS	Mean, Var, HOG	Boost Trees	29.96
Modified SEEDS	Mean, Var	Boost Trees	24.25
Modified SEEDS	Mean, Var, HOG, color hist	Boost Trees	26.45

5. Conclusions

In this paper we introduce a method for semantic segmentation of remote sensing images into various classes as a prerequisite of automated cartography and monitoring of agriculture crops in the broad context of precision agriculture or Agriculture 5.0.

The image was first segmented using a modified SEEDS superpixel algorithm. Next, each superpixel has been described in terms of color, homogeneity, and texture. For classification, ensembles of boosted trees have been involved. The proposed method has taken into account the specific forms and shapes existing in the studied database and a post processing step has been implemented. Our choices for each step of the method aimed at avoiding the high computation needed by deep learning-based methods.

Table 2

Comparison, of class level performance, between the proposed method and deep learning-based solutions. The reported metric is mIoU and higher is better

Method	Overall	Background	Double	Dry-down	Endrow	Nutrient	Planter	Water	Waterway	Weed Cluster
Ours - Non deep	29.96	56.42	17.16	43.18	8.05	31.26	18.9	42.67	32.84	19.16
Chui et al. [18] Best Deep	43.40	73.31	28.25	57.43	21.74	38.86	33.55	73.59	34.37	28.33
Chui et al. [18] DeepLab	35.28	73.01	21.32	56.19	12	35.22	20.1	42.19	35.04	22.51

The method has been evaluated positively on images from AgricultureVision dataset. Each of the proposed changes was shown to improve the overall performance with respect to other popular choices. The limitations of the method are derived from the limited amount of resources available. From an end-user point of view, the method, due to the training step, is limited in prediction of images similar to those used to build the classifier. The post-processing exploits characteristics of the database, that are derived from acquisition, thus also limiting the generality.

Acknowledgement

This work was funded from the AI4AGRI project entitled "Romanian Excellence Center on Artificial Intelligence on Earth Observation Data for Agriculture". The AI4AGRI project received funding from the European Union's Horizon Europe research and innovation program under the grant agreement no. 101079136.

R E F E R E N C E S

[1] *X. Ren, J. Malik, "*Learning a classification model for segmentation, in: International Conference on Computer Vision, 2003, pp. 10–1.
[2] *P. F Felzenszwalb and D. P Huttenlocher*. Efficient graph-based image segmentation. International journal of computer vision, 59(2):167–181, 2004.
[3] *M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa*. Entropy rate superpixel segmentation. In Proceedings of IEEE Computer Vsion and Pattern Recognition 2011, pages 2097–2104. IEEE, 2011
[4] *R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk*. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2274–2282, 2012.
[5] *Van den Bergh, M., Boix, X., Roig, G. and Van Gool, L.* SEEDS: Superpixels extracted via energy-driven sampling. International Journal of Computer Vision, 111, pp.298-314. 2015
[6] *Z. Li and J. Chen*. Superpixel segmentation using linear spectral clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1356–1363, 2015
[7] *D. Stutz, A. Hermans, and B. Leibe*. Superpixels: An evaluation of the state-of-the-art. Computer Vision and Image Understanding, 166:1–27, 2018.
[8] *W.-C. Tu, M.-Y. Liu, V. Jampani, D. Sun, S.-Y. Chien, M.-H. Yang, and J. Kautz*. Learning superpixels with segmentation-aware affinity loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 568–576, 2018.
[9] *L. Cai, X. Xu, J. H. Liew and C. S. Foo* Revisiting Superpixels for Active Learning in Semantic Segmentation with Realistic Annotation Costs In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021
[10] *H. Kim, M. Oh, S. Hwang, S. Kwak and J. Ok* Adaptive Superpixel for Active Learning in Semantic Segmentation In Proceedings of the IEEE International Conference on Computer Vision, 2023
[11] *Zhang, A., Zhang, S., Sun, G., Li, F., Fu, H., Zhao, Y., Huang, H., Cheng, J. and Wang, Z.*, Mapping of coastal cities using optimized spectral–spatial features based multi-scale superpixel classification. Remote Sensing, 11(9), p.998. 2019.
[12] *Cheng, J., Zhang, F., Xiang, D., Yin, Q. and Zhou, Y.*, PolSAR image classification with multiscale superpixel-based graph convolutional network. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-14, 2021.
[13] *Ma, F., Zhang, F., Xiang, D., Yin, Q. and Zhou, Y.*, Fast task-specific region merging for SAR image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-16, 2022.

[14] *Geng, J., Wang, R. and Jiang, W.*, Polarimetric sar image classification based on feature enhanced superpixel hypergraph neural network. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-12, 2022.

[15] *N. Dalal, B. Triggs*, Histograms of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[16] *Hastie, T., Rosset, S., Zhu, J. and Zou, H*., Multi-class adaboost. Statistics and its Interface, 2(3), pp.349-360. 349–360. 2009.

[17] *C. Florea, F. Gieseke* Artistic Movement Recognition by Consensus of Boosted SVM Based Experts, Journal of Visual Communication and Image Representation, 56 (2018) 220–233.

[18] *Chiu, M.T., Xu, X., Wei, Y., Huang, Z., Schwing, A.G., Brunner, R., Khachatrian, H., Karapetyan, H., Dozier, I., Rose, G. and Wilson, D.*, 2020. Agriculture-vision: A large aerial image database for agricultural pattern analysis. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2828-2838).

[19] *B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, A. Yamada*, Color and texture descriptors, IEEE Trans. Circ. Syst. Video Technol. 11 (6) (2001) 703–715.

[20] *M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim*, Do we need hundreds of classifiers to solve real world classification problems?, J Mach. Learn. Res. 15 (1) (2014) 3133–3181.