IMPROVED LINE DETECTION IN IMAGES USING NEURAL NETWORKS AND DTE SUBCLASSIFIERS

Jérôme Treboux, Dominique Genoud

HES-SO Valais-Wallis
Institute of Information Systems
Sierre, Switzerland

Rolf Ingold

University of Fribourg
Department of Informatics
Fribourg, Switzerland

2021 9th European Workshop on Visual Information Processing (EUVIP) | 978-1-6654-3230-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/EUVIP50544.2021.9484058

ABSTRACT

It is widely accepted that deep neural networks are very efficient for object detection in images. They reach their limit when multiple long line instances have to be detected in very high resolution images. In this paper, we present an original methodology for the recognition of vine lines in high resolution aerial images. The process consists in combining a neural network with a subclassifier. We first compare a traditional U-Net architecture with a U-Net architecture designed for precision agriculture. We then highlight the significant improvement in vine line detection when a DTE is added after the customized U-Net. This methodology addresses the complex task of dissociating vine lines from other agricultural objects. The trained model is not sensitive to the orientation of the lines. Therefore, our experiments have improved the precision by around $15\,\%$ compared to our improved neural network.

Index Terms— Machine Learning, Neural Network, Decision Tree Ensemble, Image Recognition, Line Detection, Line Recognition, Vineyard Lines, Object Recognition

1. INTRODUCTION

Machine Learning (ML) is widely used for image recognition. Indeed, deep Learning (DL) algorithms such as AlexNet Neural Network [1], Convolutional Neural Networks (CNN) [2], VGG-16 Neural Network [3] and U-Net [4] have a high performance in the classification and segmentation of objects in images.

The algorithms are trained on large datasets to improve the accuracy of image recognition, such as the detection of objects in images. Many research groups focus on the detection of small objects in high-resolution images, such as [5] and [6]. However, when several instances of identical long objects cross an image, the performance of the segmentation algorithms is limited, as for example in the case of the detection of vine lines in a high-resolution image taken by a UAV.

Furthermore, in the case of precision farming, the distinction between agricultural objects in images is a complicated task: colors may be similar (e.g. trees, grape leaves, grass) and shapes may be comparable (e.g. lines of bushes and vines).

This paper, based on precision agriculture, compares the efficiency of the U-Net [4], often used for the segmentation of objects in images, and a customized U-Net combined with a DTE (Decision Tree Ensemble) that extracts additional information and refine the classification. Our previous experiments have shown that traditional algorithms and simple neural networks do not achieve sufficient precision [7][8]. For this reason, we decided to use a U-Net. The training is performed on perfectly identical data according to a precise methodology described in the section 4.1. The experiments allow to segment the image and detect the vine lines.

This research document is structured as follows: the state-of-the-art of the algorithms exploited is addressed in the next chapter. Then, the data processing is explained in section 3. The experiments are detailed in section 4 followed by the results. Finally, the conclusion and the outlook are presented in section 6.

2. STATE-OF-THE-ART

Most recent Deep Learning models are based on artificial neural networks. Deep neural networks (NN) such as the ResNet [9] or Inception-v4 [10] are very efficient for image classification and segmentation. The ResNet-151 gets an error of $5.7\,\%$ and the Inception-v4 gets an error of $4.2\,\%$ for the classification of pictures in the ImageNet dataset [9] [10].

The U-Net neural network, used for objects segmentation in images, is often used in the biomedical domain, for cell detection or segmentation of medical images [11] [12]. The FCN-8 model, which is partially composed with an architecture similar to the U-Net, achieves a classification efficiency greater than $80\,\%$ for segmentations on aerial images [13] [14].

Research is being done on the detection of lines and edges in images. For problematics such as the detection of edges of documents or specific objects, Hough Transformation is very efficient [15]. In more complex cases, such as the detection of power lines in images [16], the use of various neural networks are promising to determine the areas containing lines in the image. However, line segmentation at pixel level is missing. Other researches on the detection of vine lines have interesting results but are dependent and limited by the colors [17] [18].

Then, to refine the classification, we decided to implement decision trees based on our experience. They are used for classification but also for the detection of objects in images, such as for face recognition [19]. They are frequently used for the extraction of specific parameters as well as for the selection of the most important ones. They allow to reduce dimensions by selecting the most important and dependent parameters [20] [21].

3. DATASET AND DATA PROCESSING

The dataset is created using an UAV (Unmanned Aerial Vehicle) that flies over the Alpine vineyards of Switzerland. The UAV is programmed to take a high-resolution RGB photo every 5 meters (see Figure 1). Each image contains meta-information such as the UAV's altitude, speed and geolocation. Using a photogrammetry software, we combine the images to create a complete landscape image of the vineyard. The detailed description of the dataset is presented in the Table 1.

Fig. 1: Example of an aerial image of a Swiss Alpine vineyard taken at an altitude of 50m using an UAV. The red square represents a **patch** of the original image (145px).

Table 1: Detailed description of the complete dataset available for the experiments reported in this research.

	Settings
Images	790 Aerial images
Geolocalisation	TRUE
Resolution per image	4.000x3.000 pixels
Colour	RGB
Frequency	Every 5m
Altitude	50 meters

The Figure 2 describes the methodology used to create the dataset for the experiments. (1) Eight images are used to train the models, two are used to test the models and the other three are used for the validation. (2) Each image is manually labelled through a mask with white lines representing vine lines. (3) The original images are divided into **patches** of identical size. The corresponding labels are also divided into **patches** [22]. (4) Finally, due to the small amount of data, the vine lines do not represent all orientations and therefore do not correspond to the reality of the field. To overcome this limitation, an augmentation of the training data has been performed. The specifications of the data augmentation are based on the research of [23]. The augmentations performed are described in the Table 2.

The size of the patches for the experiments on this paper is 145x145 pixels. This size is optimal for the problem of vine line detection.

Table 2: This table details the augmentation applied to the data increasing the variations in the image. This augmentation forces different orientations for the lines of the vine.

Augmentation name	**Augmentation settings**
Flip	- Horizontal - Vertical
Rotation	- [-10°;10°] every 1° - 90°
Scale	- [50 %;90 %] every 10 %

4. METHODOLOGY AND EXPERIMENTS

4.1. Methodology

To determine the optimal segmentation and classification algorithms for our problem, the following methodology is applied during the experiments.

1. Preparation of the dataset for training and validation, following the process described in Figure 2. The original images and annotations are divided into **patches** [22].
2. The U-Net is trained and validated using first the original network structure proposed by [4] and then the altered U-Net proposed in this research (see Figure 3). The input of the network corresponds to the **patch** of the original image. The output of the network, classifies pixels as vine or non-vine.
3. To improve vine line detection, a Decision Tree Ensemble is implemented after the U-Net. The DTE filters the rectangular areas representing the vine lines.
4. The results are established at an optimized threshold. They are compared statistically on the validation images. The measures used for the results are precision, recall and confidence interval.

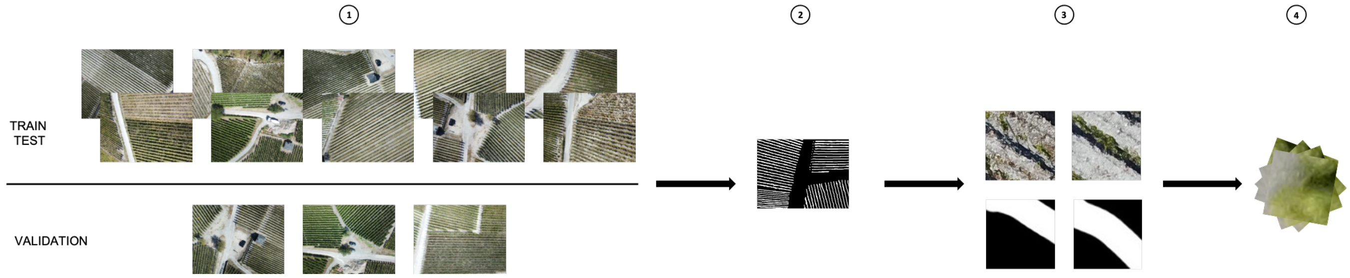


Fig. 2: Methodology for the preparation of the dataset. (1) Creation of the training, testing and validation dataset. (2) Image annotation. (3) Division of the original and annotated images into **patches**. (4) Training dataset augmentation.

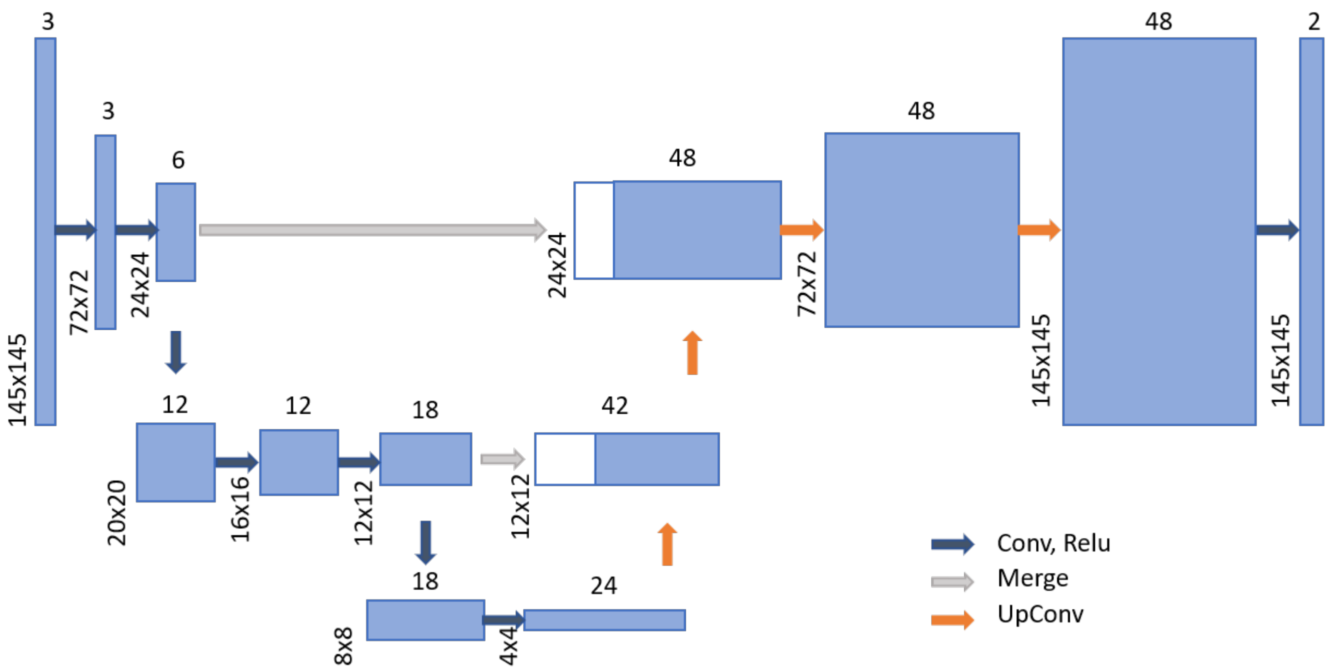


Fig. 3: Altered structure of the U-Net, with an input of 145x145 pixels of 3 channels (RGB) and an output of 145x145 pixels of 2 channels (classes).

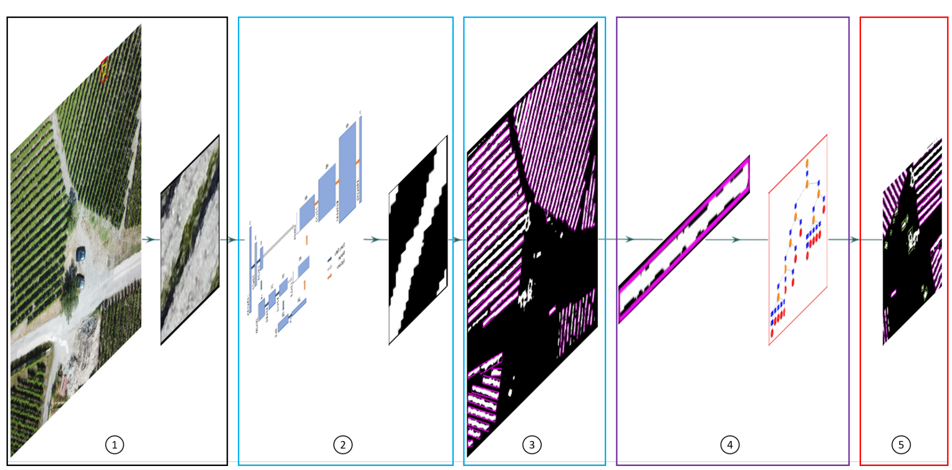


Fig. 4: Methodology for the recognition of the vine lines. (1) Original image is divided into patches of 145x145 pixels. (2) Segmentation and classification of vine lines with the U-Net. (3) Contours recognition with the edge detection algorithm. (4) Extraction of the vine lines and their parameters to improve the classification with a DTE. (5) Result produced by the DTE, with a reduction of False Positive.

4.2. Segmentation and classification using a U-Net

As a baseline, the U-Net neural network is implemented with its original structure allowing an optimal segmentation and classification of images, especially areas of interest [4].

For a more precise recognition of the vine lines, the original U-Net architecture is adapted to create a reduced structure, allowing the analysis of **patches** of optimal size required for the experiments [22]. The adapted U-Net architecture is presented in the Figure 3.

The network is configured and trained to detect two classes: **vines and non-vines.**

4.3. Classification improvement using a DTE

To reduce pixel detection errors around vine lines, the edge detection algorithm proposed by [24] is applied. The algorithm determines the surroundness relations among the borders of a binary image. The output of the algorithm is one rectangle per vine line, containing the detected pixels as a vine. Each rectangle is then filtered to eliminate false detection. When a rectangle is too small (ratio between long edge and short edge), it is merged with its very closest correctly detected neighbor or eliminated.

To improve the detection of the vine lines and to capture the longitudinal shapes, we propose to use a Decision Tree Ensemble (See the Figure 5). The output image of the U-Net is used for training. Each of the rectangles of the image is labeled as vine / non-vine (Figure 4, pt.3). Then, each rectangle is passed to the DTE to train the detection model. The settings used are the dimensions of the rectangle and the characteristics Haralick [25], Tamura [26] and First Order Statistics [27] (Figure 4, pt.4). The output of the DTE provides a final detection and classification of the vine lines (Figure 4, pt.5). The Figure 4, pt.5, shows the major impact of using DTE. The improvements of this classification are presented in the chapter 5.

Finally, the classification validation is done on images containing lines of vines with varied orientations. It's executed according to the process described in Figure 4, from the original image, through U-Net, edge detection, filtering and finally DTE.

5. RESULTS

This chapter presents the results obtained with the different algorithms. The edge detection algorithm is applied to the results of neural networks, generating rectangles that are used to determine the quality of detection using IoU (Intersection over Union) [28]. The comparison of the results allows to evaluate the quality of the classification of the U-Net, the modified U-Net and the impact of the DTE on the classification.

Precision, recall and IoU (Intersection over Union) are used for the model evaluation. Thanks to the IoU, it is possible to determine the precision of recognition at the level of the location [28]. Indeed, this measurement makes it possible to determine whether the detections are completely overlapping or offset. The closer the IoU is to 1.0, the more accurate is the detection. If the IoU is too low, the detection is too offset. In our experiment, when the IoU drops below 0.5, the detections can no longer be considered as TRUE. Considering the very close lines, it is no longer possible to determine to which vine line the detection corresponds.

The threshold of 0.75 for the IoU allows to keep the lines distinct. It is used to calculate precision and recall. The experiment results are presented in the Table 3.

Our results include the Standard Error (SE) calculated with the Equation 1.

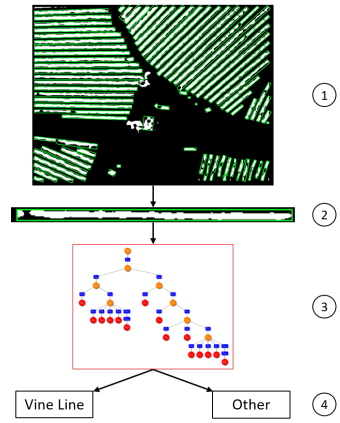


Fig. 5: Process to improve vine line detection with a DTE. (1) Output image from the U-Net. (2) Extraction of parameters for each rectangle. (3) Training of a DTE for line detection. (4) Classification of the object as vine or non-vine line.

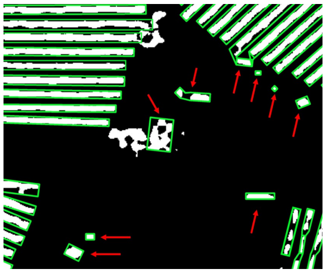
Table 3: This table details the precision, standard error and recall obtained with a U-Net and an altered U-Net and an altered U-Net combined with a DTE.

	Precision	Recall
U-Net from [4]	0.705 ± 0.08	0.740
Altered U-Net	0.815 ± 0.08	0.746
Altered U-Net with DTE subclassifier	**0.963** ± 0.03	0.746

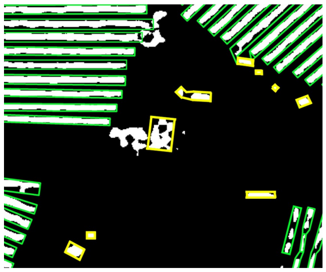
$$SE = Z_\alpha \sqrt{\frac{p(1-p)}{n}} \tag{1}$$

Where:

- Z_α: confidence level at 95 %, $Z_\alpha = 1.96$
- p: is the precision
- n: is the number of data, including both classes



(a) **Before** the DTE subclassifiers



(b) **After** the DTE subclassifiers

Fig. 6: Visualization of the classification results for lines of vines with the altered U-Net. Image (a) represents the classification at the output of the U-Net, by applying an edge algorithm. Image (b) represents the output of the U-Net after filtering using a DTE on the detected areas. The green are the rectangles detected as vine lines. The red arrows show false positives. The yellow are the rectangles with a corrected classification after the DTE.

The results in the Table 3 shows that our altered U-Net have a higher precision compared to the original U-Net [4]. In addition, our experiments show a significant improvement of the precision of $14.8\,\%$ with the use of a DTE to refine vine line detection. On the other hand, recall does not vary with the inclusion of a decision tree as a decision aid. Indeed, DTE mainly allows to remove **False Positive** and increases the precision of our image recognition models. The DTE extracts additional information to capture the lines. Noise due to false detection is lowered, eliminating image-like agricultural objects such as bushes or trees.

In the Figure 6, the detection of vine lines is presented graphically. The white areas are the pixels detected as being a vine line with our U-Net. The green rectangles are the vine line detected as a vine. The red arrows highlight the areas wrongly classified as vineyard. Finally, the yellow rectangles point out the corrections made by the DTE, which removes the False Positive.

These visualizations highlight the improvements made by the Decision Tree Ensemble. Indeed, the False Positive are removed (represented with yellow rectangles in the Figure 6), such as the white area in the middle of the image which is actually a tree. But DTE also allows to remove objects such as bush lines detected as vines.

6. CONCLUSION

The detection of lines of vines crossing a complete image is a complicated task. Traditional Deep Learning algorithms such as U-Net are not precise enough so that objects that are similar (e.g. bushes) or with identical colors (e.g. trees) are often incorrectly detected.

The use of a Decision Tree Ensemble to aid decision-making, subsequent to a first analysis performed with a neural network, allows a significant improvement of the classification. The combination of an U-Net and a DTE improves the precision by around $15\,\%$. This allows the removal of False Positive and thus reduces the noise associated with the misclassification of similar objects in terms of imaging. By appending a DTE to the U-Net it is possible to extract additional information to quickly converge to a more precise classification and line detection. Furthermore, using a DTE allows the classification to be adapted to a new similar problem without having to train the neural network from the beginning, considerably reducing training time [7].

The next steps in our work are focused on extending the process, by applying research on the detection of vine lines on low-resolution images, such as satellite maps imagery. This will allow us to confirm that the re-training of the DTE is fast, efficient and essential. The exploration of other algorithms similar to DTE will also be able to extend our experiments.

7. REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[2] Waseem Rawat and Zenghui Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[3] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] Z. Meng, X. Fan, X. Chen, M. Chen, and Y. Tong, "Detecting small signs from large images," in *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, 2017, pp. 217–224.

[6] Kang Tong, Yiquan Wu, and Fei Zhou, "Recent advances in small object detection based on deep learning: A review," *Image and Vision Computing*, vol. 97, pp. 103910, 2020.

[7] Jérôme Treboux, Rolf Ingold, and Dominique Genoud, "Towards retraining of machine learning algorithms: An efficiency analysis applied to smart agriculture," in *2020 Global Internet of Things Summit (GIoTS)*. IEEE, 2020, pp. 1–6.

[8] Jérôme Treboux and Dominique Genoud, "High precision agriculture: An application of improved machine-learning algorithms," in *2019 6th Swiss Conference on Data Science (SDS)*, 2019, pp. 103–108.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, vol. 31.

[11] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al., "U-net: deep learning for cell counting, detection, and morphometry," *Nature methods*, vol. 16, no. 1, pp. 67–70, 2019.

[12] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *annual conference on medical image understanding and analysis*. Springer, 2017, pp. 506–517.

[13] Dimitrios Marmanis, Jan D Wegner, Silvano Galliani, Konrad Schindler, Mihai Datcu, and Uwe Stilla, "Semantic segmentation of aerial images with an ensemble of cnss," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016*, vol. 3, pp. 473–480, 2016.

[14] Jamie Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," *arXiv preprint arXiv:1606.02585*, 2016.

[15] Bilal Iqbal, Waheed Iqbal, Nazar Khan, Arif Mahmood, and Abdelkarim Erradi, "Canny edge detection and hough transform for high resolution video streams using hadoop and spark," *Cluster Computing*, vol. 23, no. 1, pp. 397–408, 2020.

[16] Rabab Abdelfattah, Xiaofeng Wang, and Song Wang, "Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[17] Mohamed Kerkech, Adel Hafiane, and Raphael Canals, "Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images," *Computers and electronics in agriculture*, vol. 155, pp. 237–243, 2018.

[18] Carlos Poblete-Echeverría, Guillermo Federico Olmedo, Ben Ingram, and Matthew Bardeen, "Detection and segmentation of vine canopy in ultra-high spatial resolution rgb imagery obtained from unmanned aerial vehicle (uav): A case study in a commercial vineyard," *Remote Sensing*, vol. 9, no. 3, pp. 268, 2017.

[19] Emir Kremic and Abdulhamit Subasi, "Performance of random forest and svm in face recognition.," *Int. Arab J. Inf. Technol.*, vol. 13, no. 2, pp. 287–293, 2016.

[20] Jung Hwan Cho and Pradeep U Kurup, "Decision tree approach for classification and dimensionality reduction of electronic nose data," *Sensors and Actuators B: Chemical*, vol. 160, no. 1, pp. 542–548, 2011.

[21] Kajal Rai, M Syamala Devi, and Ajay Guleria, "Decision tree based algorithm for intrusion detection," *International Journal of Advanced Networking and Applications*, vol. 7, no. 4, pp. 2828, 2016.

[22] Libin Jiao, Lianzhi Huo, Changmiao Hu, and Ping Tang, "Refined unet: Unet-based refinement network for cloud and shadow precise segmentation," *Remote Sensing*, vol. 12, no. 12, pp. 2001, 2020.

[23] Vincent Schülé and Dominique Genoud, "Détection automatique d'objets agricoles avec du machine learning," M.S. thesis, University of Applied Sciences and Arts Western Switzerland (HES-SO), 2018.

[24] Satoshi Suzuki et al., "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.

[25] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, , no. 6, pp. 610–621, 1973.

[26] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, man, and cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[27] Jiwen Lu, Gang Wang, and Pierre Moulin, "Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 329–336.

[28] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.