International Journal for Modern Trends in Science and Technology, 8(02): 219-226, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0802036
Available online at: http://www.ijmtst.com/vol8issue02.html

Identifying Trees and Their Species in Street Images Using Deep Learning

M. V. S. Haranadh[1] | A. Sureshbabu[2]

[1]Department of Computer Science and Engineering, JNTUA, Anantapuramu, Andhra Pradesh, India.
mharanadh@gmail.com
[2]Department of Computer Science and Engineering, JNTUA, Anantapuramu, Andhra Pradesh, India.
sureshalladi.cse@jntua.ac.in

To Cite this Article
M. V. S. Haranadh and A. Sureshbabu. Identifying Trees and Their Species in Street Images Using Deep Learning. *International Journal for Modern Trends in Science and Technology* 2022, *8* pp. 219-226. https://doi.org/10.46501/IJMTST0802036

Article Info
Received: 16 January 2022; Accepted: 18 February 2022; Published: 24 February 2022.

ABSTRACT

The order-based tree recognition algorithm in a timberland climate to create autonomous routes for woodland vehicles. To divide the image into tree objects, a combination of shading and surface signals was used. Because of the wide range of illumination, the impact of varied shading patterns, non-homogeneous bark surface, shadows, and foreshortening, dividing photos into tree objects is a moving errand. The methods used to do this was to find the optimal mixes of shading and surface descriptors, as well as order algorithms sing monocular vision, measure the distance between the wood vehicle and the foundation of portioned trees as an extra project. In addition, driven by the previously proposed impediment mindful single short identifier (SSD) retrained model work, the original tragedy and a tree part-consideration model to decreasing the misleading recognitions caused by hefty impediment. And by training and evaluating a few different variations of the proposed model, then approve the importance of each part. So, loss function used as performance metric. The goal is to build the accurate model to detect the trees and their species names in streets images.

KEYWORDS: Deep Learning, Object Detection Model, Single Shot Detection (SSD), Image Processing, Faster Region-based Convolutional Neural Network (Faster R-CNN)

1. INTRODUCTION

Currently, a vast quantity of geo-labeled photos from almost any part of the populated world are being captured and shared on the internet. Client-contributed photos and symbolism from web-based planning administrations are the two main sources of generally accessible images. While user-provided images primarily cover well-known sites, effective business operations provide a uniform and extensive coverage of the world's inhabited regions, particularly metropolitan areas. This includes skyward symbolism captured by satellite and aeroplane, as well as high-goal ground sceneries spread throughout the roadway network [3]. Program-based link points, such as Google Maps, provide free and well-organized access to this vast, modern, and geocoded treasure trove.

A dream-based framework for distinguishing and characterizing openly visible items. Upward and road view symbolism are combined to replenish and renew a

public tree stock with GPS location and fine-grained species at no cost to the public. Our methods were inspired by Opentreemap1, a large-scale tree planning effort that aims to provide a concentrated, freely accessible, and frequently renewed tree resource for every city on the earth. The project is suffocated by the massive amount of human labour required to inventory trees. Then hypothesized that Computer Vision could make it feasible, and carried out the combination of calculation and recognition would be most appropriate in general.

The tree has evolved into an essential component of densely populated metropolitan areas. The common of the trees are placed along the streets and place an important role in the city's structure. They serve as multifunctional frameworks for reducing pollution and sound, as well as providing cover for pedestrians. As a result, watching their health and development is critical [1]. The most important task is to determine their total. The management might be interested in knowing how many road trees there are in a certain area, as well as what kind of tree they are. Previously, this issue had to be physically resolved. Specialists were contacted to go into the street and count and arrange trees in a sequential manner, which plainly needed a large amount of time and effort. Recently, civil organizations have been able to detention a series of road view photographs in a short period of time using street view cars. These images are subsequently [10] sent to specialists who employ marking apparatuses in the computer to physically identify trees. Regardless, the manual naming errand is still difficult and time-consuming. This made it difficult to focus on this errand for long phases of time, resulting in missing or wrong marks.

This topic can be considered an exemplary article placement issue that has been discussed in the PC vision industry for quite some time. Both precision and efficacy for general object detection in supplied photographs have been dramatically enhanced, thanks to the application of profound learning procedures, such as profound convolutional[14] neural organisations (CNNs). In any case, there are still a few challenges to overcome for a few specified assignments, such as person on foot position [3], ailing tissue recognition in clinical image [5], or shallow break identification in metal materials [5]-[8]. Impediment is likely to be the most difficult aspect of the road tree finding task, especially in a swarming scenario. In Fig. 1, the taken image may also hurt from the negative impacts of poor illumination. Interclass obstacle and intraclass impediment are the two most common types of hindrance. Interclass impediment occurs when trees areobstructed by things or goods of various classifications, and intraclass impediment occurs when trees obstruct each other. By using a few cutting-edge general article identification structures to the introduced road tree image informational index, such as Faster R-CNN and

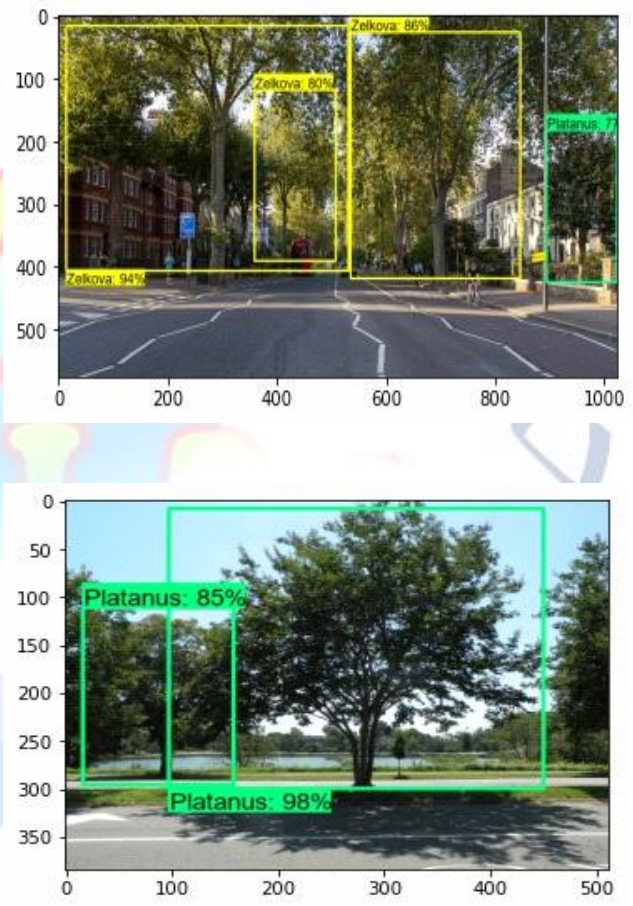

Fig. 1. Sample Street image with annotation and accuracy

YOLOv3, the location result is frequently noticed to be improper. Then the common of these erroneous discoveries are produced to a large extent by obstruction. Because of growth and advancement, two close trees beside the road may even be coordinated with one another, making it impossible to differentiate

them in any case for natural eyes. A few studies have been conducted to address the obstacle problem in the item finding task. Regardless, none of them could be directly applied to the road tree identification application and effectively fix the tree-to-tree obstacle [20] issue. On the other hand, no existing approach for recognizing road trees has been offered. As a result, then to present a tree identification organization as a solution to this problem. To concentrate on addressing the obstacle issue in the road tree detection project in order to advanceprecision. In general, the goal of this paper is to look into the use of deep learning-based photo inspection to device a tree detection approach in road view images, mostly in congested settings. In addition, to deal with low illuminance situations, a simple yet effective programmed picture splendour modification strategy is proposed.

The contribution of the paper as follows:

To predict the tree images in street images. Initially to collect some street images along with trees and find the specious also. After data collect perform some data preprocessing and data cleaning steps. Have to do annotation manually by using some annotation technique. By LabelImg to annotate the images and draw the bounding boxes by giving class names manually. In model training first download pre-trained SSD model and retrain that model on custom data. Perform retrain SSD model on test data predict the accuracy and loss function.

The section 2 provides related work, Section 3 gives Proposed methodology, Section 4 presents the results and analysis, section 5 gives Conclusion.

2.RELATED WORK

Taking advantage of the ubiquitous road view pictures available from Google Streets Views (GSVs), Maxillary, and other sources, a slew of projects have been launched to better understand metropolitan plant life (Li et all. 2018; Li et all. 2015), improve present guides with fine-grained division classifications (Mattyuss et al. 2016), [11] investigate metropolitan morphologies by planning the conveyance of picture areas (Crandall et al. 2009), and dissect (Li et al. 2017). Furthermore, raised symbolism has been used with road view photos to realize tree[13] location/characterisation (Wegners et al. 2017), land use characterization (Workmans et al. 2018), and fine-grained street division[16] (Wegner et all. 2017). (Mattyus et al. 2016). To estimate object areas from road level photographs, various techniques, like (Timofte and Van Gool, 2011), rely on an improved on locally level territory model.

The rapid growth of CNNs and CNN-based picture content analysis in recent years has been remarkable. It had been shown to be capable of extracting highlight depictions from a huge dataset (LeCuns et all., 2016)[18]. In addition, since it was planned, metropolitan investigations, including road level images, have been generally updated. Various studies employ deep learning for object discovery and grouping, as well as picture semantic division to screen area change (Naika et all. 2018)[21], to evaluate metropolitan discernment on a worldwide scale (Dubeys et all. 2015), to appraise segment cosmetics (Gebrus et al. 2017) [26], to predict apparent security reactions to pictures (Naik et al. 2014), to anticipate financial pointers (Arietta et al. 2014), and to investigate[19] a variety (Mirowski et al. 2018). Separating traffic components inside street crossing places from road view iconography, on the other hand, has received less attention. Furthermore, these tactics make use of GSV as a source of data, however GSV demands a charge after downloading a certain amount for free, which is unquestionably not a good option for organizations or individuals with limited exploratory resources. Mapillary, a free, publicly sponsored, almost continuous renewed, and universal road level iconography, is thus introduced into our work.

To date, only a few methods have been made accessible to plan certain types of articles from road view symbolism: transportation[15] signals (Jensens and al., 2017; Tread et all., 2015), street signs (Soheilians et ali., 2014), and sewage vents (Soheilian et all., 2013). (Timofte et al., 2011). These techniques use position triangulation to determine the locations of street resources based on individual camera views. When multiple articles appear in a similar setting, they all rely heavily on different visual and mathematical features to match. As a result, when multiple indistinguishable articles exist at the same time, the display of these

tactics is bad. As a result, a more effective technique is offered. Hebbalaguppes[14] et al. (2018) present the problem as an article recognition task, and then use the sound system[18] vision (Seitz et al. 2016) method to measure item organises from sensor plane directions using GSV. In any event, unlike GSV, Mapillary road view photographs don't have any camera intrinsics or projective change in their EXIF data, therefore they can't be used to adjust the camera later. Thus, unable to propose a similar technique for traffic signals/signs restriction based on Maxillary [22] images. Krylovs et all. (2019) have recently combined the use of monocular depth assessment and triangulation to enable planned planning of confusing situations with the synchronous presence of various, externally equivalent objects of interest, and achieve a location accuracy of roughly 2m.

To focus on the evaluation of crossing sites in order to improve articles related to OSM convergences, such as traffic signs and lights, and to locate them for independent driving or route reference. In light of picture semantic division from street convergences photographs, So, to offer a whole pipeline to separate scene components such as structures, sky, streets, walkways, traffic signals, and signage. Because there are understandable relationships between geographies, traits, and semantics of street objects, the order of semantic components should be used for restriction purposes. In addition, an ascribed topological twofold tree (ATBT) in view of metropolitan language can be set out to depict the geographies among street items, along with the division outcomes. These are then synchronized with OSM map highlights. Finally, as promising outcomes, street objects can be limited.

3. PROPOSED METHODOLOGY

In the proposed work, look at a whole pipeline for limiting tree on road images. Three modules make up the pipeline: (1) preparation and cleansing of data; (2) object division and acknowledgment; and (3) limitation module, the full construction is depicted in Figure 2. The primary module is in charge of preparing pre-processed and cleaned data for the following two modules (see Section 3.1). The following module, which involves picture semantic division as well as object identification and characterisation, mostly removes street-related data (see Section 3.2). An attributed topological parallel tree (ATBT) is constructed in the final module to handle the general position connection between removed items at convergences and to locate articles with metropolitan punctuation (see Section 3.3). In low-illuminance and crowded settings, to provide a unique structure for discovering road trees. Our system pipeline is depicted in Fig. 2. Two crucial stages are included in the technique: 1) input picture splendour change and 2) road tree location.

Based on this proposed architecture initially collect the data then perform some data cleaning and data pre-processing technique. Then annotations must be carried out manually. While it is time consuming process and need to do in correct manner or else our model accuracy will be affect. Then split the data for training and testing. In this case testing means making some data as a unseen data. Download the pre trained SSD model change the parameter based on out data set and classes then retrain that model on custom data. After that save that trained model perform it on test data. Using those results can conclude our model performance. Which is defined our proposed work in figure 2.

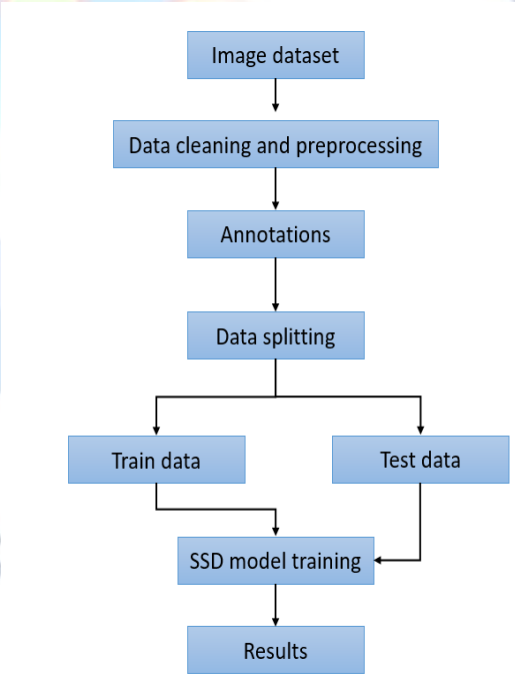

Fig. 2. Proposed Architecture

Data Pre-processing and Cleaning

The primary goal of the information element is to prepare information for the next two modules. Every one of the available photographs can be retrieved by interrogating significant Mapillary APIs, and a support has been made up for each street crossing point to separate picture arrangements contained in the cushion. A picture arrangement relates to a consumer walking down the street in a certain direction. So, the hypothetically creation offour picture tracks for a convergence with four street branches by integrating various picture successions based on their geo-locations due to four types of hard driving bearings, such as west-east, east-west, south-north, and north-south. Furthermore, the camera area, including scope and longitude, as well as the camera point, has been eliminated.

Furthermore, by observation the Mapillary's picture arrangements frequently show the GPS position float of the pictures, which could be due to the topographical climate during the trip (for example, being near tall structures or under dense tree cover that blocks the GPS signal), or it could be due to a problem with the shooting device's inherent GPS collector. Fortunately, one of Mapillary's major advantages is that road view photographs of a same street fragment can be transferred multiple times by different volunteers. Furthermore, there is a certain amount of cross-over between the two contiguous images, allowing us to address their shooting positions.

Fig. 3. Image with bounding boxes for image localization

Annotation

Though, with ability to handle information naming precision in order to build a pipeline of great preparation data that affects your CV calculations. Picture grouping, semantic division, object placement and acknowledgment, and example division are all supported by our foundation. Bouncing boxes, polygons, and key point comment are examples of naming tools. By using this LabelImg technique to annotate our images. It will ask the person which class and bounding box area. Then select the area and class label. Based on this we build the classification model. Classification can be divided into two parts the first one is drawing bounding boxes and the second one is classifying the box with appropriate label. LabelImg is a graphical image annotation tool. It is written in Python and relies on Qt for its graphical interface. In the PASCAL VOC design, which ImageNet uses, comments are kept as XML entries. Additionally, it supports SSD, YOLO and CreateML designs.

Object Detection and Classification using SSD model

The single shot multibox indicator [13], which includes two key steps, highlight map extraction and convolutional channel applications, to identify objects, is probably the best identifier in terms of speed and precision.

$$L\left(x,c,l,g\right)=\frac{1}{N}\left(L_{conf}\left(x,c\right)+\alpha L_{loc}\left(x,l,g\right)\right)$$

The SSD engineering builds on the VGG-1616 organisation [16], and this decision was made in light of the organization's strong performance in top-notch picture grouping assignments and its ubiquity in challenges involving move learning. Rather than the first VGG entirely associated layers, a slew of assistant convolutional layers alter the model, allowing it to separate items at several scales while decreasing the magnitude of the contribution to each subsequent layer. The jumping box age considers the use of matching pre-figured, fixed-size bouncing boxes called priors with the first dispersion of ground truth boxes. These priors are chosen to keep the convergence over association (IoU) proportion equivalent to or more noteworthy than 0.50.5.

SSD Model formula

The general misfortune work described in Eq. (1) is a direct mix of the certainty misfortune, which uses unmitigated cross-entropy to estimate how sure the organization is of the figured jumping box, and the area misfortune, which uses the L2 standard to estimate how far away the organizations expected bouncing boxes are starting from the earliest stage ones.

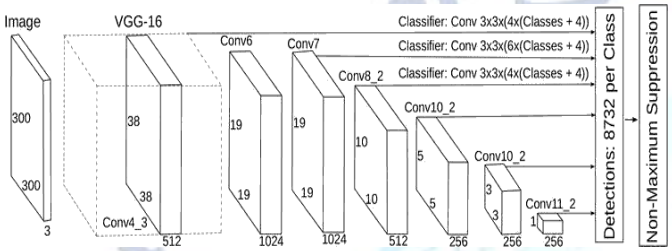

Fig. 4. SSD Model Architecture

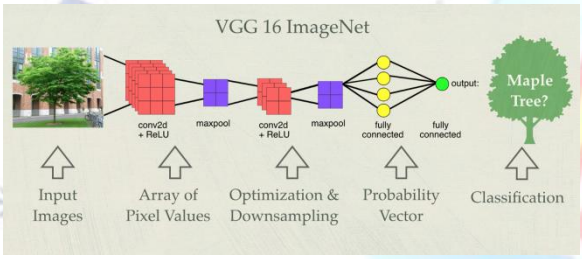

Fig. 5. VGG 16 ImageNet

4. RESULTS AND DISCUSSION

By using loss function, accuracy as performance metrics. Use loss function for predicting the bounding boxes and accuracy to classify the tree names. The loss values lies between 0 to infinity loss 0 is best case and accuracy lies 0 to 100 in best cases accuracy is 100. The annotations of the data using labelImg then download the pre-trained SSD model and retrained on our custom dataset. Tested with up to 2000 steps in first stage which has 1.23 loss then the step size increase the loss is decreased to 0.23 and accuracy increased to 82 percent. After getting this loss and accuracy, the trained model has to be tested on the new data so to can check the reliability and performance. After testing with the new data the model has done well. Its working fine predicting bounding boxes as well as classifying the trees very accurately. The results will be in the format of figure 6.

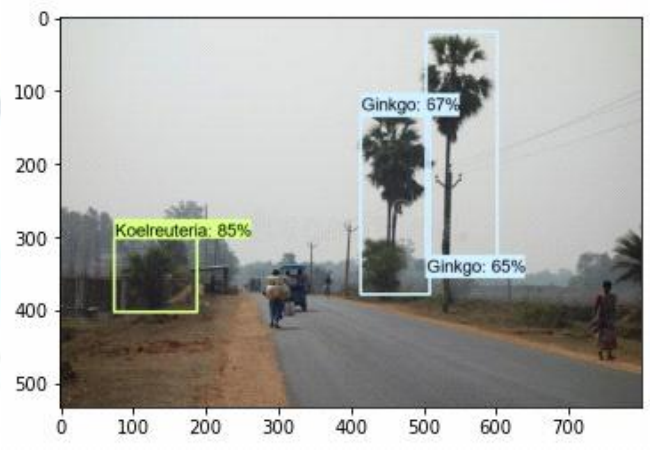

Fig. 6. Results output

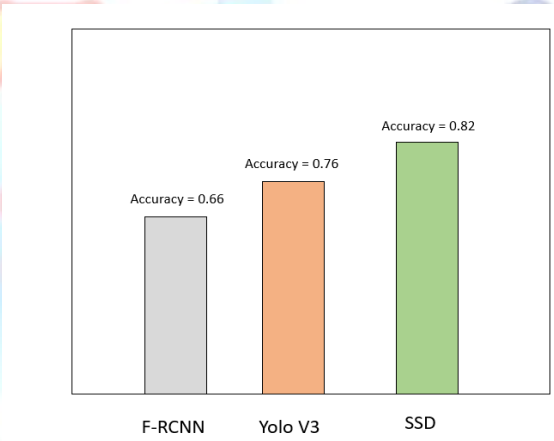

Fig. 7. Models accuracy function performance

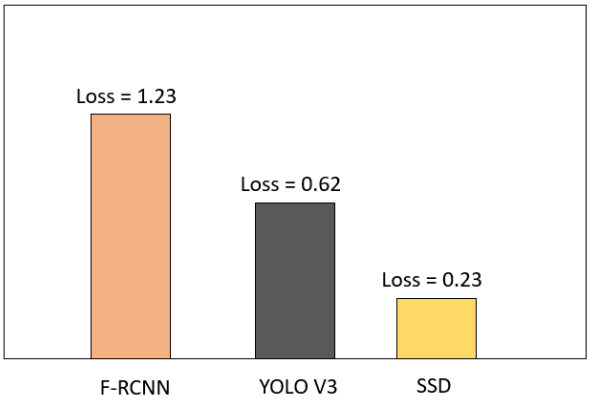

Fig. 8. Models loss function comparisons plot

Table 1:- Results comparison table

S. No	Models	Accuracy	Loss
1	FasterR-CNN	0.66	1.23
2	Yolo v3	0.76	0.62
3	SSD	0.82	0.23

Before building the SSD model we tried faster RCNN and Yolo also. Compared to SSD Faster R-CNN and Yolo models are old model. SSD is updated model. In faster RCNN and Yolo some limitations are there that is the reason to build SSD in this project. By using faster RCNN got 66% accuracy 1.23 loss. In this case the miss classification error is very high then tried Yolo v3 model. By using Yolo model also got 76% accuracy and 0.66 loss. It is some better compared to faster RCNN. In this Yolo also miss classification errors there. At last the SSD model got 82% accuracy and 0.23 loss by using SSD model is trained. The graphs are plotted in figure 7, figure 8 and table 1. Using this results comparison with the Yolo and faster RCNN, SSD performing well. Then created interface and tried this model on some unseen data and got best results on unseen data also. Finally SSD is best model in this project. The SSD model accuracy by increasing the number training images and adding parameter tuning technique.

5. CONCLUSION

To address the issue, this study proposes a novel tree detection framework that combines cutting-edge deep learning-based detection algorithms with two ground-breaking advances in training loss definition and network module designation. To integrate past structural information from trees into the SSD detector with occlusion prediction, then used the part attention module. The occluded components of the suggested unit would be given lower weights in order to lessen their detrimental impact on the subsequent classification job. Our tree part-attention network is used to solve the problem of detecting street trees. The suggested technique displays good fidelity detection across the provided street tree data set captured by street-view collecting vehicles. Extensive experimental comparisons reveal that our proposed framework out performs the baseline in terms of detection accuracy, even in congested contexts with occlusion concerns. However, due to variances between similar tree species and resemblance across different tree species, street tree categorization remains a tough task. One possible future route is to design an effective classification network to differentiate these trees based on their species once they have been detected. Furthermore, we proposed preprocessing approach is directly connected to the photographs used in this paper. It may not be generalizable to other picture data sets. As a result, developing a more general method for the low-illumination problem is another possible research direction. In future the location can be added and GPS per each tree and monitor it by using this model.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008.

[2] Uijlings, Jasper RR, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. "Selective search for object recognition." International journal of computer vision 104.2 (2013): 154-171.

[3] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[5] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. Springer, Cham, 2014.

[6] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, and Dragomir Anguelov, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

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

[8] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[9] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[10] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object

detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[11] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[12] Li, Yi, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. "Fully convolutional instance-aware semantic segmentation." arXiv preprint arXiv:1611.07709 (2016).

[13] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." arXiv preprint arXiv:1703.06870 (2017).

[14] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." European Conference on Computer Vision. Springer, Cham, 2014.

[15] Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. "Pyramid scene parsing network." arXiv preprint arXiv:1612.01105 (2016).

[16] O. Nevalainen et al., "Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging," Remote Sens., vol. 9, no. 3, p. 185, 2017.

[17] Victor F. Strîmbu and B. M. Strîmbu, "A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data," ISPRS J. Photogramm. Remote Sens., vol. 104, pp. 30–43, Jun. 2015.

[18] J. Aval et al., "Individual street tree detection from airborne data and contextual information," in Proc. GEOBIA From Pixels Ecosyst. Global Sustainability, 2018. [Online]. Available: https://prodinra.inra.fr/?locale=en#!ConsultNotice:458921

[19] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," Remote Sens., vol. 9, no. 1, p. 22, 2016.

[20] W. Li, R. Dong, H. Fu, and L. Yu, "Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks," Remote Sens., vol. 11, no. 1, p. 11, 2019.

[21] Z. Zhang, S. Fidler, and R. Urtasun, "Instance-level segmentation for autonomous driving with deep densely connected MRFs," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 669–677.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2005, vol. 1, no. 1, pp. 886–893.

[24] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in Proc. Brit. Mach. Vis. Conf., London, U.K., 2009.

[25] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2241–2248.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[27] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, Apr. 2013.