



**5th International Conference of the International
Commission of Agricultural and Biosystems Engineering
(CIGR)**

Hosted by the Canadian Society for Bioengineering (CSBE/SCGAB)
Virtually from Québec City, Canada – May 11-14, 2021



**A DEEP LEARNING PLANTS IDENTIFICATION MODEL FOR AUGMENTED REALITY
TOURING IN URBAN PARKS**

KONSTANTINOS P. FERENTINOS¹, MYRTO S. BARDA²

¹ Dept. of Agricultural Engineering, Soil & Water Resources Institute, Hellenic Agricultural Organization “Dimitra”, k.ferentinos@swri.gr

² Dept. of Agricultural Engineering, Soil & Water Resources Institute, Hellenic Agricultural Organization “Dimitra”, myrto.barda@yahoo.gr

**CSBE21477 – Presented at 5th International Conference of the International Commission
of Agricultural and Biosystems Engineering (CIGR) (General Session)**

ABSTRACT Urban parks usually contain a wide variety of plant and tree species, which form suitable conditions for the attraction of city residents and visitors. Enhancing the touring experience of these areas, can increase the environmental awareness among citizens and assist in the development of alternative forms of “green tourism”. New technologies are a key tool in enhancing the experience of touring urban parks, as they can make park visits much more entertaining and at the same time educative, highlighting interesting information about the flora and fauna of the park, as well as various other points of interest. In the context of an ongoing research project called “VR-Park”, an augmented reality system is developed for the enhancement of the urban park touring experience. In this work, the development of a computational model for the automatic identification of specific plant and tree species through real-time video of a smartphone or tablet device is presented. The model is based on deep learning methodologies (convolutional neural networks) for fast object identification in video stream, and 24 different plants and trees are identified. The model is tested and assessed in real conditions in an urban park. It will be eventually incorporated into the augmented reality application developed in the VR-Park project, to enhance the user experience by presenting information about the identified plants and trees, and thus transforming the system into an agriculture-specific educational tool which would further improve the park visitor’s experience.

Keywords: Artificial intelligence, deep learning, plants detection, augmented reality, urban parks

INTRODUCTION Urban parks constitute not only recreational spaces for citizens and city visitors, but also important ecological entities with crucial environmental benefits for the atmosphere of urban areas. In the vast majority of cases, the degree of interest that citizens show in an urban park, depends on but also determines the conditions that exist inside the park, concerning the quality of plants and trees, the overall area conditions and

the maintenance level of its infrastructure. Thus, it is of the benefit of the citizens' life conditions that urban parks are attractive to them and to city visitors.

The use of modern technology can make the touring of urban parks not only more attractive and useful, but also safer and more rewarding to the park visitor (Pombo et al., 2018; Dieck & Jung, 2018). In addition, useful information can be provided to park administrators, regarding several aspects of park management and touring services, such as visitors demographics, popular areas and paths around the park, problem reporting and relevant feedback from visitors, etc. In the context of a research and development project called "VR-Park", the design and development of an augmented reality (AR) mobile application for urban park touring and management is proposed, which adopts a user-centered approach, focusing on using artificial intelligence methodologies to make the park visiting experience more attractive to the visitor and the park administration more data-rich so that data analytics techniques can be used by park managers to further improve park operation.

One of the main goals of the AR mobile application is to provide real-time on-screen information about the several objects of interest of the park, which are in the field of view of the mobile device and in close proximity to the user / park visitor. GPS and digital compass of the mobile devices provide position and orientation data, however, their accuracy is not sufficient for a seamless and well-aligned overlay onto the user's screen. This operation of the mobile application concerning the localization and tracking of some specific object of interest on the screen, becomes problematic when the distance between the user and the object of interest becomes smaller than the GPS accuracy. In these cases, the GPS-provided position makes the object to appear on the user's screen at different positions than its actual one, or even to disappear. Another problem, specifically concerning plants and trees in the park, is the fact that they constitute a dynamic setting, because throughout the year, plants are replanted at different spots, seasonal plants or damaged plants are removed, etc. In addition, several plants are too close to each other for a GPS-based device to be able to distinguish them. Thus, the registered to a geospatial database information concerning their exact coordinates, is not actually useful.

These problems could be overcome by making the user's device "aware" of the actual position of each object of interest, in relation to the position and orientation of the user. This can be realized with the use of object detection and identification models, which can be integrated into the AR mobile application. In this paper, a deep learning model for the identification of plants and trees in urban parks is presented, with an ultimate goal to integrate it into the "VR-Park" AR mobile application for the touring of urban parks, so that useful information about specific plants and trees of interest can be provided to the user during park touring.

MATERIALS AND METHODS The automated identification of plant species in the visitor's field of view through the "VR-Park" mobile application provides an important advantage in the availability of specific information about the plants of interest in an urban park. Specific plant species are objects of interest of the park which have such density and such

dynamics in terms of their locations (due to possible replanting at different locations, seasonal plants and problems leading to plant destruction and removal) that their spatial digitization becomes practically impossible. Therefore, their automated visual identification can become a great asset towards the targeted provision of information to the visitor regarding specific plant species. An overview of the pilot park of the project (“Pedion Areos” park in central Athens, Greece) and the general structure of the basic components of the proposed system are depicted in Fig. 1. The interaction of the developed subsystems and the basic features of the overall system are described in Ferentinos et al. (2020).



Figure 1. The pilot park of the “VR-Park” project and the general structure of the overall proposed touring system.

The automated identification of objects through visual recognition from the information provided by the camera of the mobile device during the operation of the “VR-Park” application, i.e. through the video stream, is possible by using sophisticated artificial intelligence algorithms for the detection of objects in image and video. These algorithms are based on deep learning methodologies (LeCun et al., 2015) and in particular on Convolutional Neural Networks (CNNs) models (LeCun et al., 1998), which are an evolution of "traditional" Artificial Neural Networks (ANNs). ANNs are computational models that, in some way, simulate the structure of the brain's neurons, so that through parallel computations they are able to model, after appropriate “training”, the operation of complex systems. Their development usually requires a large amount of data derived from the system to be modelled, which are used to train the neural network, i.e. to adjust, through specific training algorithms, the weights with which the artificial neurons (nodes) of the network are interconnected, so that the output of the model simulates the output of the system to be modelled, for specific input data.

For applications related to image recognition, several basic CNN architectures have been developed and successfully applied to complex problems of this type, e.g. AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), Overfeat (Sermanet et al., 2013), VGG (Oxford Visual Geometry Group), etc. In particular, for object recognition in a

video stream, where besides the performance of object recognition, the time of recognition plays an important role so that the whole process can work in real time, the algorithm YOLOv3 (Redmon & Farhadi, 2018), which recently (June 2020) evolved to its 5th version, has been very successful, while other very successful and fast models include R-CNN (Ren et al., 2015), Mask R-CNN (He et al., 2017), U-Net (Ronneberger et al., 2015) and UNet++ (Zhou et al., 2018).

For the specific application of detecting plants and trees of interest via video streaming from mobile devices, the YOLOv3 model was chosen, both because of its speed and the fact that it performs object detection and not only segmentation at the pixel level of the image (object segmentation). The algorithms that perform object segmentation essentially create the exact outline of the object to be searched for, which is not necessary for this application. On the other hand, algorithms that perform object detection, usually return the coordinates and size of rectangles enclosing the identified objects (bounding boxes), which is what is required in the current application (localization of objects of interest). Therefore, the models developed were based on the methodology of the YOLOv3 algorithm.

Dataset and model development The first step in the development of the plants and trees identification model was the selection of candidate plant species to be included in the model. For that purpose, the project's pilot park was used and three approaches were followed: Candidate plants a) were situated on commonly visited paths and areas of the park, b) had impressive flowers or fruits, and c) were the most commonly found ones in the entire park. Each approach initially identified 10 species, resulting in 30 plant species, plus some additional species suggested by the park managers, 24 were ultimately selected and used to develop the identification model (Table 1).

The required data (photographs of the plants and trees in real conditions) for model training and testing was collected in various stages between November 2018 and March 2020, mostly using mobile phones with high-resolution cameras, but also using a DSLR camera. Images were captured under various weather conditions and at different times of the day so that a high variation in lighting conditions was included. An initial deployment of the model was carried out using a total of 1477 photos. Subsequently, the plants with the greatest difficulty in being currently identified by that initial model, were spotted and the photographic material was enhanced with a focus on these plants, reaching a total of 3461 photos. Finally, the dataset was further improved with some additional images captured during summer season, leading to a final dataset with a total of 4816 images.

Data annotation of the 24 plant/tree species in all images was performed manually using an open-source software (LabelImg, available at: <https://github.com/tzutalin/labelimg>) as shown in Fig. 2. The models were trained using 90% of the annotated images, while they were tested with the remaining 10% of the images, but mainly with videos taken in different environmental conditions (different weather conditions) and times (different stages of the day) using a mobile phone, in the pilot park. Three different types of YOLOv3 models were used: i) the regular type, ii) the "tiny" type which has lower complexity thus

lower memory and computational power requirements, and can achieve higher object recognition speeds (the most suitable for implementation on mobile devices), and iii) the “spp” type which is an intermediate implementation between the regular and tiny types, with some improvements. PyTorch was used for model implementation, and training was performed using two NVIDIA® RTX 2080 GPUs on Linux operating system (Pop!_OS 19.04).

Table 1. Final list of plants/trees included in the identification model.

	Common name	Scientific name
1	Bay laurel	<i>Laurus nobilis</i>
2	Black locust	<i>Robinia pseudacacia</i>
3	Blue Jacaranda	<i>Jacaranda mimosifolia</i>
4	Brachychiton	<i>Brachychiton sp.</i>
5	Canary Isl date palm	<i>Phoenix canariensis</i>
6	Carob	<i>Ceratonia siliqua</i>
7	Chinaberry tree	<i>Melia azedarach</i>
8	Cottonwood	<i>Populus spp.</i>
9	Evergreen oak	<i>Quercus ilex</i>
10	Goldenrain tree	<i>Koelreuteria paniculata</i>
11	Japanese cheesewood	<i>Pittosporum tobira</i>
12	Japanese photinia	<i>Photinia glabra</i>
13	Judas-tree	<i>Cercis siliquastrum</i>
14	Mulberry	<i>Morus spp.</i>
15	Myrtle	<i>Myrtus communis</i>
16	Oleander	<i>Nerium oleander</i>
17	Persian silk tree	<i>Albizia julibrissin</i>
18	Platanus	<i>Platanus spp.</i>
19	Pyracantha	<i>Pyracantha spp.</i>
20	Rose	<i>Rosa sp.</i>
21	Ulmus minor	<i>Ulmus campestris</i>
22	Viburnum	<i>Viburnum spp.</i>
23	Washingtonia	<i>Washingtonia spp.</i>
24	Whistling pine tree	<i>Casuarina equisetifolia</i>

Model performance metrics Model performance was first evaluated on the test set, which, as mentioned before, consisted of 10% of the total available photographic material, using the widely used metrics for evaluating the performance of object detection in images:

- Precision
- Recall
- F1 score
- mAP (mean average precision)

RESULTS The trained model was tested on the testing set, which consisted of randomly selected 10% of all available images (i.e., on 480 images). Overall performance results were not very high, with Precision, Recall, F_1 score and mAP values all around 0.6 - 0.7. However, these values are not representative of the actual performance of the model in real video conditions, because, in order to train the model in the best possible way, all annotated images (included the ones randomly selected for testing) included annotations of too many different occurrences of the target plants, even ones in quite long distances from the mobile device (an example can be seen in Fig. 2). The actual goal of the developed model is to be able to correctly identify these plants in rather small distances from the user, so that meaningful and useful information about the identified plants can be provided on the screen of the mobile device. Thus, the identification of plants far away from the user is not important to the system. In real-time testing on video stream, the model performed very satisfactorily, as it can be seen from some screenshots with correct identifications shown in Figures 3-5 (labels are in “Greek-Latin” format, numbers are the confidence degree for each identification in the range 0-1). In addition, in some test images, several plants do not seem to be identified (thus, counted as False Negatives), however, in the exact same settings and lighting conditions in video mode, these plants are identified under slightly different angles, which is something totally acceptable, because once an object is identified with a specific confidence, its identification is stored in the application and the appropriate information is presented to the user, even if its identification is then lost by the model.



Figure 3. Screenshot of correct identifications during a test video. Green: Bay laurel, Blue: Oleander, Purple: Brachychiton, Orange: Canary Isl date palm.

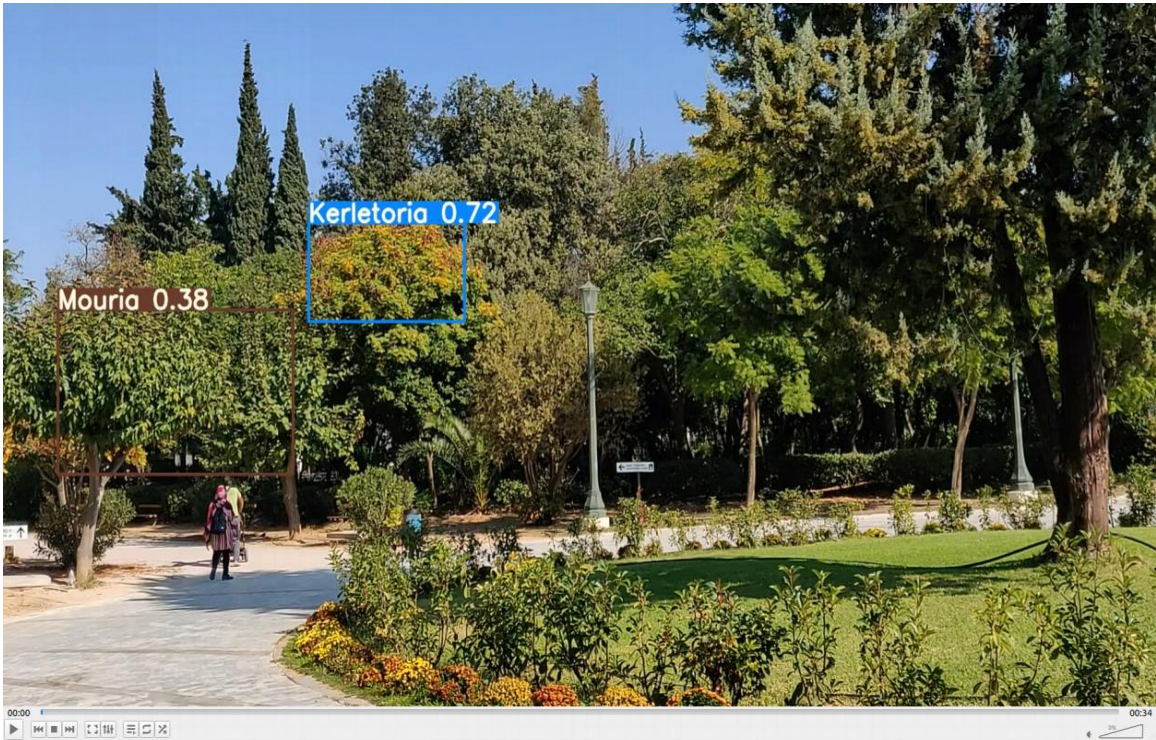


Figure 4. Screenshot of correct identifications during a test video. Brown: Mulberry, Blue: Goldenrain tree.

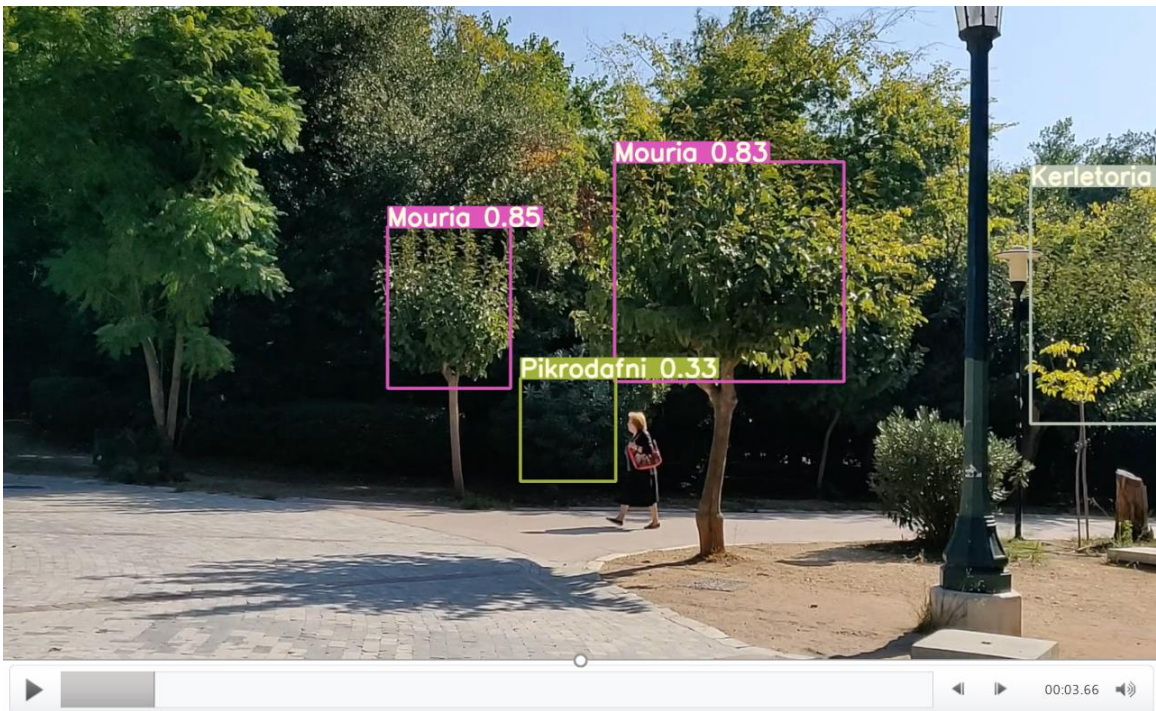


Figure 5. Screenshot of correct identifications during a test video. Purple: Mulberry, Green: Oleander, Grey: Goldenrain tree.

CONCLUSION A model for the identification of 24 plants and trees commonly found in urban parks was presented, based on deep learning methodologies, and specifically using the YOLOv3 model architecture. Its goal was to correctly identify plants and trees of interest in video so that it can be integrated into an augmented reality mobile application in a way that relevant information about the identified plant species can be provided to the user / visitor of the park. The performance of the developed model was satisfactory in several testing videos in a pilot park under several different weather and lighting conditions, making the model suitable for integration into the final mobile application.

Acknowledgements. This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-02182).

REFERENCES

- Pombo, L., M.M. Marques, V. Carlos, C. Guerra, M. Lucas, and M.J. Loureiro. 2018. Augmented reality and mobile learning in a smart urban park: pupils' perceptions of the EduPARK game. In: Mealha, Ó., M. Divitini, and M. Rehm, M. (Eds), *Citizen, Territory and Technologies: Smart Learning Contexts and Practices*. Springer, Cham.
- Dieck, M. C., and T. Jung. 2018. A theoretical model of mobile augmented reality acceptance in urban heritage tourism. *Current Issues in Tourism* 21: 154-174.
- Ferentinos, K.P., Y. Stavrakas, H. Nakos, K. Pristouris, M.S. Barda. 2020. Initial design and features of an augmented reality system for urban park touring and management. *International Journal of Computer Theory and Engineering* 12(5): 106-112.
- He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision*: 2961-2969.
- Krizhevsky, A., I. Sutskever, G.E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, Eds. F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger (Curran Associates, Inc.), pp. 1097–1105.
- LeCun, Y., Y. Bengio, G. Hinton, G. 2015. Deep learning. *Nature* 521: 436-444.
- LeCun, Y., Y. Bottou, Y. Bengio, P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11): 2278-2324.
- Redmon, J., and A. Farhadi. 2018. YOLOv3: An incremental improvement. *arXiv preprint, arXiv:1804.02767*.
- Ren, S., K. He, R. Girshick, and J. Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NIPS)*: 91-99.
- Ronneberger, O., P. Fischer, and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*: 234-241.
- Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv:1312.6229*.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. 2015. Going deeper with convolutions. *Proc. of the IEEE Conference on Computer Vision and Pattern*

Recognition.

Zhou, Z., M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang. 2018. UNet++: A nested U-Net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning*. Springer, Cham.