ORIGINAL RESEARCH Open access

UAV-Based Automated Detection and Mapping of Olive Trees Using an OBIA–CNN Hybrid Approach

Mateo Alvarez[1*], Sofia Herrera[2], Diego Cruz[1], Andres Castro[3]

Abstract

The task of distinguishing individual trees stands as a crucial subject of investigation across forestry, remote sensing, and computer vision. It provides a mechanism for efficiently and effectively administering and preserving forests and orchards. However, this undertaking is not as simple as it may appear; tree identification and enumeration can demand considerable time, entail prohibitive costs, and suffer from limited accuracy, particularly when executed manually over broad expanses. The availability of very high-resolution UAV imagery coupled with remote sensing can make the quantification process less arduous, more rapid, and more accurate. As technology evolves, this workflow can be further automated by deploying intelligent algorithms, such as CNNs. This paper proposes an OBIA-CNN (Object-Based Image Analysis-Convolutional Neural Network) framework that integrates CNNs with OBIA for the autonomous identification and counting of olive trees from Phantom 4 advanced drone imagery. To begin with, the CNN-based classifier was constructed, trained, validated, and employed to produce olive tree probability maps on the ortho-photo. A post-classification enhancement rooted in OBIA was then carried out. Superpixel segmentation and the Excess Green index were applied, and an exhaustive accuracy evaluation was conducted to assess the proposed methodology's suitability. The application to an RGB ortho-mosaic of an olive plantation in the eastern part of Morocco succeeded, using a handcrafted training set comprising 4500 24×24-pixel image patches. In the end, the CNN pinpointed and tallied 2934 olive trees on the ortho-photo, securing an overall accuracy of 97%, which rose to 99% after the OBIA refinement step. The performance of the proposed OBIA-CNN method was also measured against the classification outcomes of Template matching, the standalone CNN method, and the standalone OBIA analysis. Our findings imply that leveraging very high-resolution imagery in combination with object-based deep learning holds considerable promise for automated detection and enumeration of olive trees, thereby bolstering precise and sustainable agricultural surveillance.

Keywords Precision agriculture, Deep learning, CNN, UAV, OBIA, Olive trees

*Correspondence:
Mateo Alvarez
mateo.alvarez@unal.edu.co
[1] Department of AI for Quinoa and Native Potato Cultivar Identification, Faculty of Agriculture, National University of Colombia, Bogota, Colombia
[2] Department of Remote Sensing for Water Deficit Monitoring, Faculty of Agriculture, University of Chile, Santiago, Chile
[3] Department of Smart Soil Erosion Modeling, Faculty of Agriculture, University of Antioquia, Medellin, Colombia

Introduction

Unmanned Aerial Vehicles, or drones, are seeing growing adoption as a remote sensing platform across multiple domains, with agriculture standing out [1]. Their capacity to deliver spatial detail at the centimeter scale in vegetated settings stems from their ability to resolve the core ecological entity: the individual plant [2]. Within precision agriculture,

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

drone applications predominantly target weed mapping [3], yield forecasting [4], pest and disease identification [5], land surveillance and cartography [6, 7], alongside tree detection and enumeration [8, 9].

Against this backdrop, a spectrum of techniques has been deployed to identify and delineate individual trees, ranging from conventional image processing in earlier decades to deep learning-driven strategies in recent years. Nonetheless, clarifying the interplay between specific algorithms and particular remote sensing datasets remains ambiguous, as each investigation operates under its own set of premises.

DEEP learning and CNN

Deep learning designates a branch of machine learning and AI. It includes a variety of algorithms that can learn features autonomously from a massive corpus of training examples and excel with voluminous datasets [10]. Neural networks rank among the most widely used DL algorithms; their adoption in geosciences and remote sensing is proving revolutionary, given their proficiency in tackling analytical challenges such as classification and segmentation [11].

The convolutional neural network (CNN) is among the most widely used deep learning architectures, engineered primarily for image classification [12-14]. Fukushima [15] first conceived it, but it gained wider traction following the rollout of Google's TensorFlow platform and the unveiling of AlexNet in 2012 [16]. Operating as a supervised algorithm, it relies on labeled data. Architecturally, a CNN comprises an input layer, a series of hidden layers, and an output layer. The hidden units mirror biological neurons, each linking to every neuron in the layer that precedes it. The input layer receives the image as a multi-dimensional value, which traverses a succession of hidden layers before yielding the final output.

The CNN pipeline can be separated into two principal stages: feature extraction and classification [16]. The feature extraction (or feature learning) stage incorporates both convolutional and pooling layers. In contrast, the classification stage is built around a fully connected layer **(Figure 1)**. The performance of the CNN scales with the size of the training corpus [12].

© 2024 The Author(s). | Published by Agro Intelligence Research Press

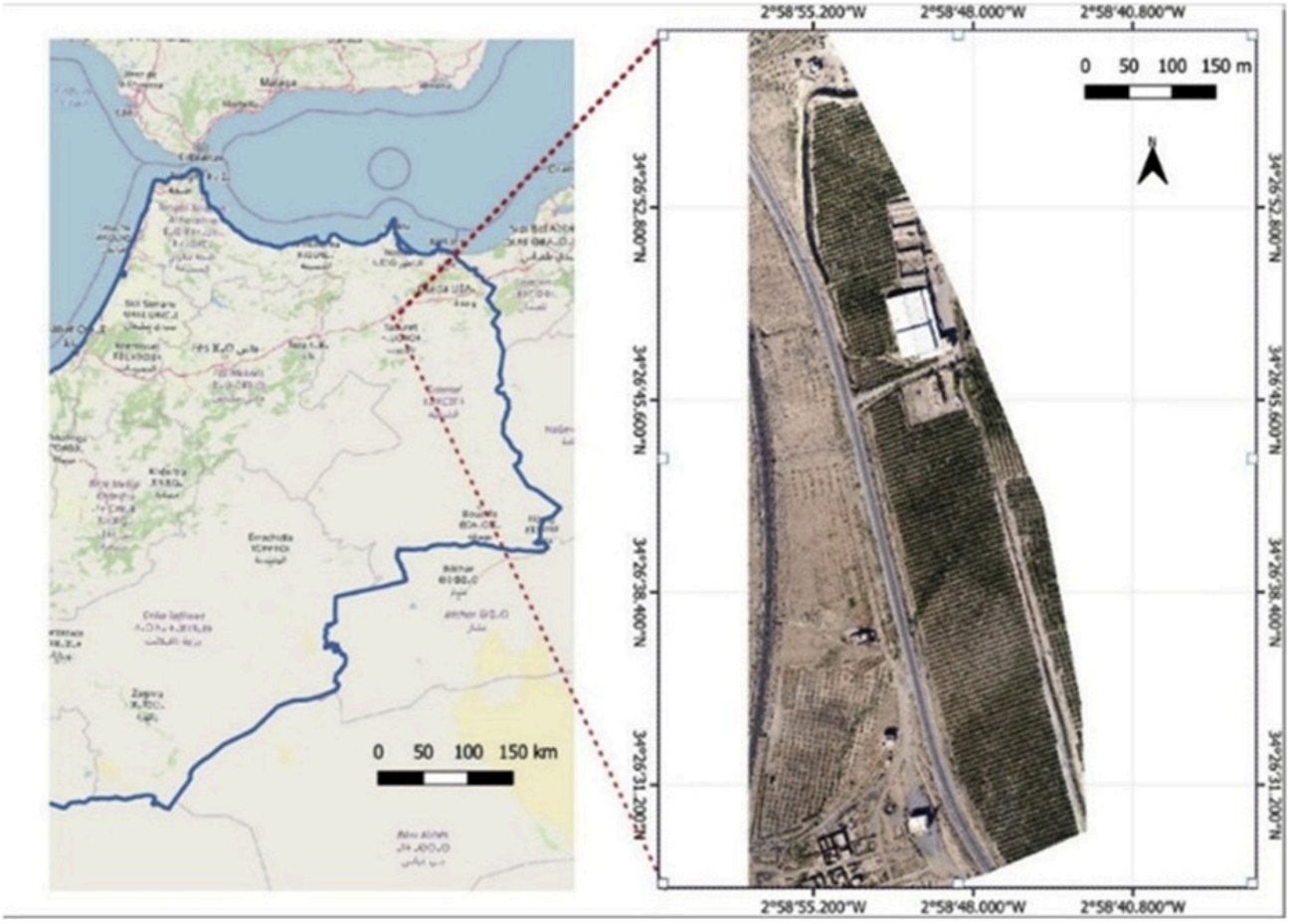

Figure 1. The location of the study area in eastern Morocco.

GEOBIA

Image analysis methodologies fall into two primary camps: pixel-oriented and object-oriented [17-19]. The pixel-oriented paradigm assigns each pixel to its most probable thematic category [19, 20]. OBIA, conversely, initiates the workflow by partitioning the image into segments and subsequently assigns these segments (or objects) to the most fitting thematic class based on a suite of attributes [21].

The adoption of GEOBIA (Geographic Object-Based Image Analysis) for classification and feature extraction has been expanding, facilitated by the emergence of software suites such as Orfeo Toolbox, Imagic, eCognition, Spring, and others **(Figure 2)**.

© 2024 The Author(s). | Published by Agro Intelligence Research Press



© 2024 The Author(s). | Published by Agro Intelligence Research Press



© 2024 The Author(s). | Published by Agro Intelligence Research Press

(2) Furthermore, the methodology relies solely on RGB data; no multispectral, Lidar, or thermal data streams were incorporated, rendering the approach substantially more economical.

(3) It ranks among the scant body of research that explicitly addresses the synthesis of CNNs and OBIA for single-tree detection and mapping.

(4) Additionally, the specific challenge of detecting olive trees arises from the distinctive architecture and coloration of their crowns, and, to the best of our knowledge, this constitutes the inaugural study to deploy a hybrid CNN-OBIA framework for detecting this emblematic Mediterranean tree from UAV imagery.

Materials and Methods

Research area

The experimental site lies at coordinates (42.904549°S, 147.328536°E) in Taourirt province, in the northeastern sector of Morocco, within the Oriental administrative region **(Figure 1)**, approximately 100 km west of the city of Oujda. The surveyed plot constitutes an olive plantation spanning 13.4 ha and hosting roughly 3800 olive trees.

UAV imagery acquisition

Aerial data were collected using a Phantom 4 Advanced UAV equipped with an RGB sensor. This platform is among the most sophisticated and widely used drones on the commercial market. Its 4K camera delivers imagery of exceptional clarity. The aircraft is equipped with convenient features, including “ActiveTrack” and “Drawing” modes.

The Phantom 4 sustains flight for 30 minutes, delivering a 25% extension in usable flight time over the Phantom 3 Professional. It can operate across a 7 km radius while maintaining full control and transmitting a 720p HD live stream of the camera perspective **(Table 1)**. Beyond that, leveraging “Active Track” simplifies subject following to merely a handful of screen interactions. The drone also offers POI functionality that, once triggered, circles the aircraft around a designated subject during motion or allows the photograph to be reframed by sliding the subject across the screen **(Table 1)**.

Table 1. Characteristics of the drone used.

Vidéo	**UHD 4 K (4096 × 2160) à 60 ips**
Photo	Capteur cmos 1″ - 20 mpx
Stabilization	Cadran à 3 axes
Temps de vol maximum:	30 min
Distance de transmission	7 km
Stabilization image	Une nacelle 3-axes
Taille	30 cm x 25 cm x 9 cm
Poids	1368 gmes avec batterie
Temps de charge	180 min

The imaging payload integrates a 1-inch, 20-megapixel sensor capable of capturing 4K/60fps footage and rapid-sequence stills at 14fps. The platform’s “Flight Autonomy” suite incorporates five optical detectors to perceive obstacles along two axes and to bypass forward obstructions. Using a titanium and magnesium alloy boosts structural stiffness while reducing weight, leaving the Phantom 4 Advanced lighter than the standard Phantom 4.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

The camera employs a mechanical shutter, affording control over exposure duration and, consequently, the interval during which the sensor remains illuminated. Its enlarged individual pixels harvest additional light, and a peak sensitivity reaching ISO 12,800 facilitates enhanced contrast rendition.

Pix4D Mapper handled mission planning and flight oversight. This photogrammetric suite reconstructs 3D geometry from an image sequence. Pix4D applies SfM to yield a point cloud and, subsequently, an ortho-image mosaic. The ortho-photo derived from the olive parcel contains red, green, and blue channels **(Figure 4)** and delivers a fine ground sampling distance of 0.24 m per pixel. These three channels were used to compute an Excess Green Vegetation Index (EXG) layer using the expression (2Green-Blue−Red).

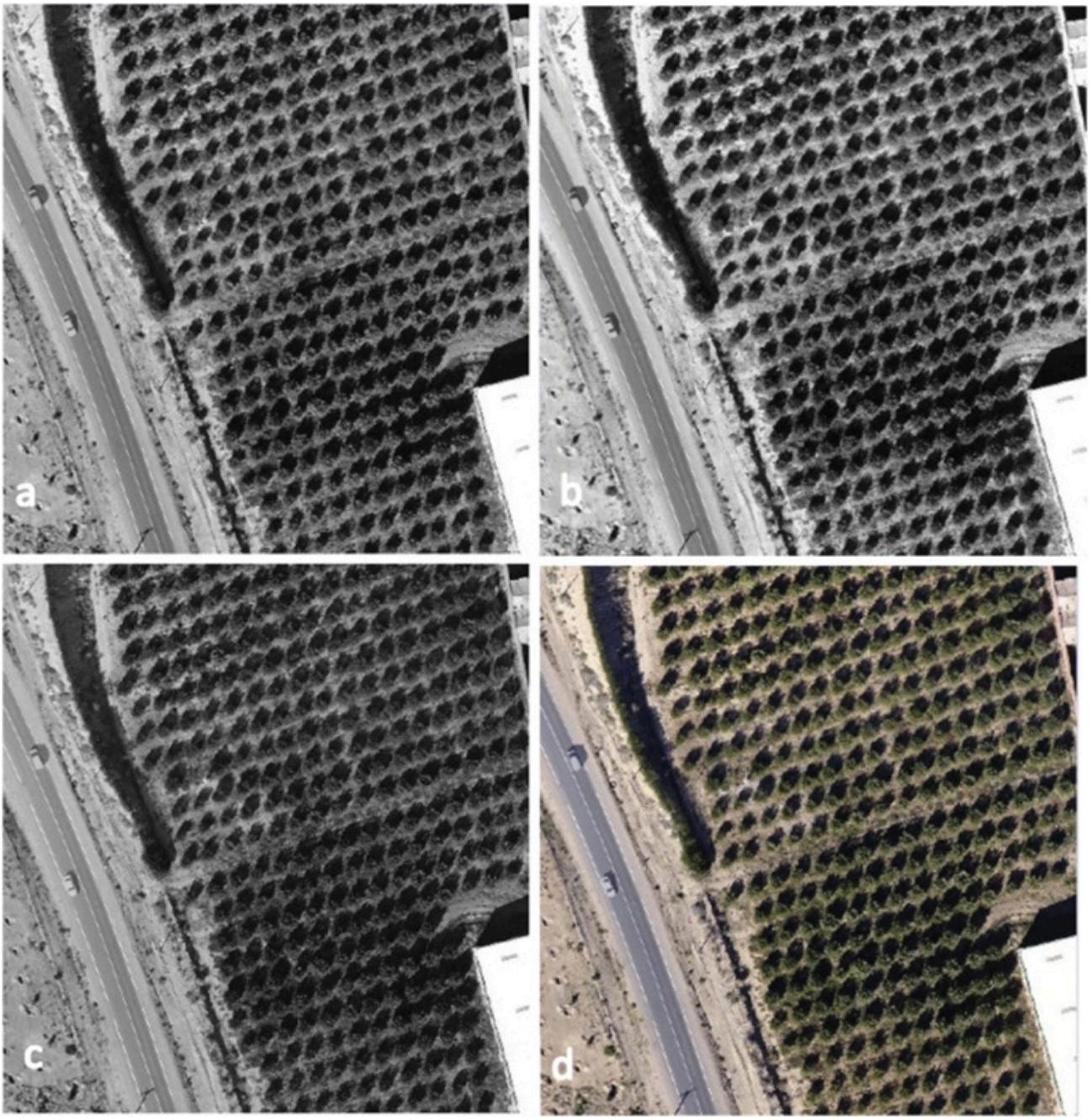

Figure 4. Example of the green (a), red (b), blue (c) bands and the true color image of the ortho-photo.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

CNN workflow

All computations were executed on a machine with a 64-bit architecture, 16GB RAM, and an Intel(R) Core(TM) i7–7700 CPU clocked at 3.60 GHz. Tree recognition was accomplished through the CNN module embedded in Trimble's eCognition Developer 9.4 **(Figure 5)**, which builds upon Google TensorFlow. As one of the foremost OBIA platforms, selecting this environment enabled tight coupling between the CNN pipeline and OBIA-driven post-processing of its output, completing the entire analytical chain within a single piece of software.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

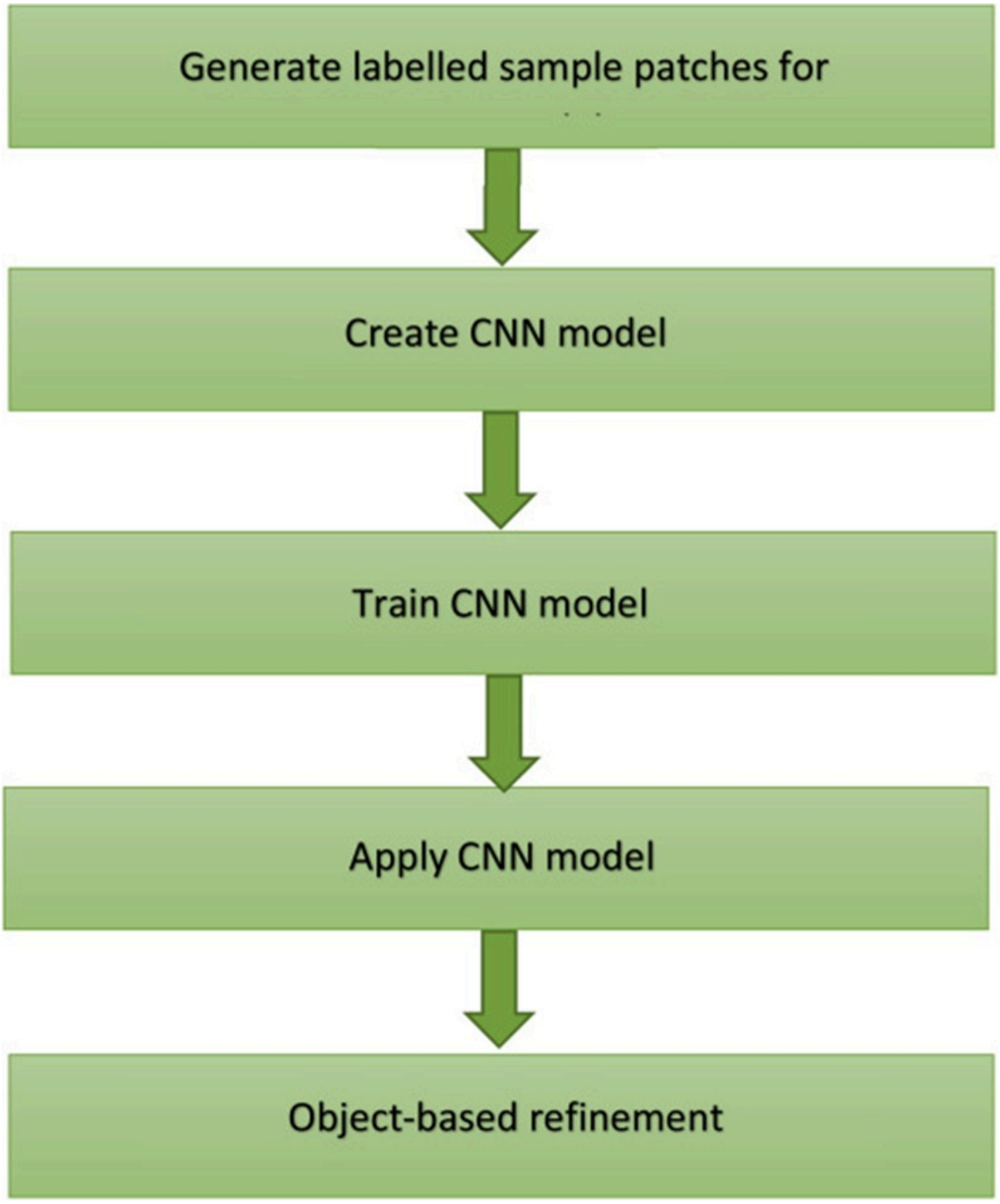

Figure 5. Flowchart showing convolutional neural network (CNN) workflow in eCognition software.

Generating patches from labelled samples

Initially, the olive grove was partitioned into a northern zone destined for training and a southern zone reserved for validation **(Figure 6)**.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

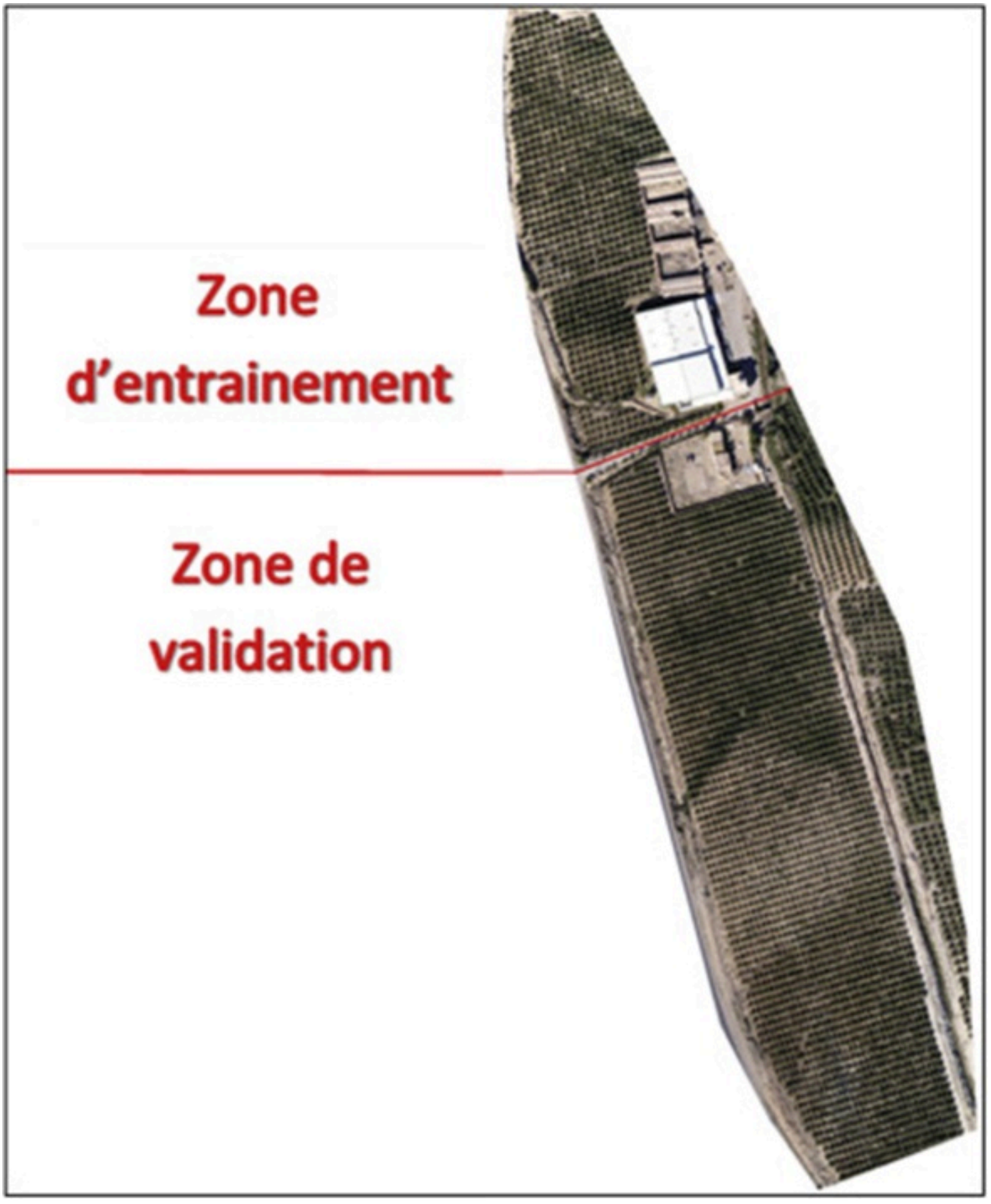

Figure 6. Division of the study area into two parts: northern for training and southern for validation.

To train the CNN, three semantic categories were defined—trees, non-trees, and shadows—each comprising roughly 1500 examples. Sample patches were extracted from a manually delineated set of individual tree references (northern trees for training, southern trees for validation). A buffer of 2 × 2 pixels around every training point was introduced to

© 2024 The Author(s). | Published by Agro Intelligence Research Press

expand the sample pool and strengthen the CNN's resilience. Through this mechanism, each tree is characterized by four pixels surrounding its centroid, from which the algorithm randomly picks during training **(Figure 7)**.

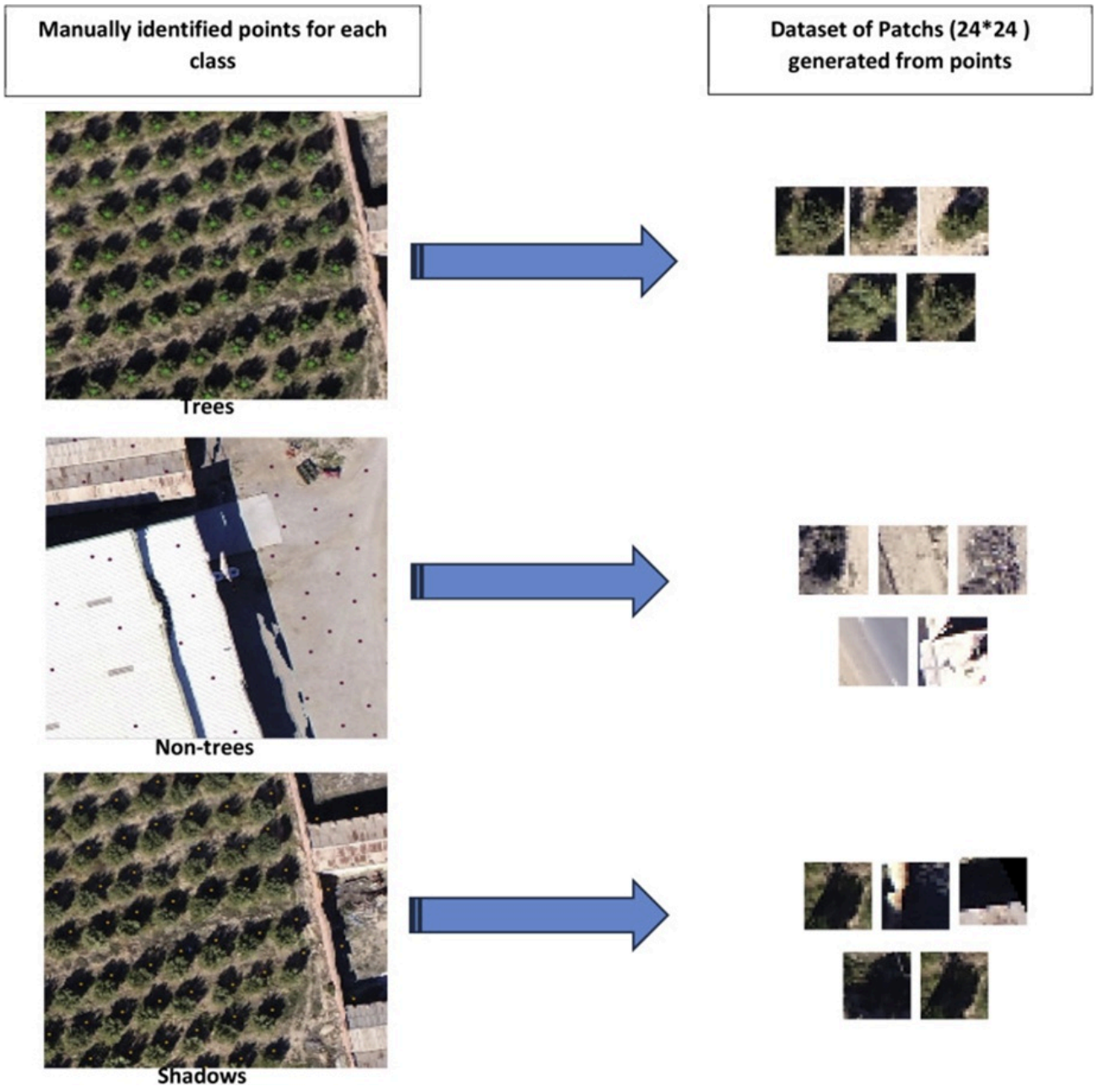

Figure 7. Generation of patches from points for each class.

The most suitable patch dimension was determined to be 24 × 24 pixels, selected through iterative experimentation. Excessively small patches triggered multiple crowns being detected per tree, whereas overly large patches overlooked juvenile trees **(Table 2)**.

Table 2. Samples generated for the tree classes.

© 2024 The Author(s). | Published by Agro Intelligence Research Press



Creating the CNN model

The CNN topology adopted here features a solitary hidden layer. The input image dimensions match those used for patch generation. Kernel dimensions, the number of feature maps, and the max pooling operation parameterize the hidden layer. To align hidden units with pixel borders, an even-sized (11 × 11) kernel paired with 40 feature maps was chosen. Max pooling down-samples the feature maps using a 2 × 2 window and a stride of two along both spatial axes. Consequently, the hidden-layer kernel has a weight tensor of shape 3 × 11 × 11 × 40. The initial factor (3) captures the number of input image channels, the (11 × 11) factors denote the count of units inside the local receptive field projecting to the hidden layer, and the terminal factor (40) corresponds to the number of produced feature maps. Thus, 40 distinct 3 × 11 × 11 kernels undergo training within this network, implying that 14,520 individual weights are optimized in its single layer.

Training the CNN model

The model is trained on labeled sample patches, with weights updated via backpropagation. A learning rate of 0.0015 was selected by trial and error. This parameter dictates the magnitude of weight adjustments per iteration. An elevated learning rate shortens the training duration but risks missing the true optimum. Meanwhile, an overly modest learning rate prolongs convergence and may trap the optimization in an unfavorable local minimum, producing suboptimal weights. In our setup, a total of 5000 training steps was prescribed, with each step drawing upon 50 training samples.

Application of the CNN model

Once the trained CNN model is deployed in the validation area (southern part), it generates a heat map that encodes the likelihood of tree presence at each pixel. This map is smoothed with a 7 × 7 Gaussian kernel, followed by a 9 × 9 morphological dilation filter to enhance local peaks in the smoothed probability surface. Tree instances are finally extracted by imposing a threshold of 0.3 on these local maxima.

Classifying and refining

© 2024 The Author(s). | Published by Agro Intelligence Research Press

The initial CNN pipeline erroneously labelled certain roof edges and patches of exposed soil as trees. To suppress these false positives and further boost the CNN's reliability, the generated heatmap and the Excess Green Index band were jointly segmented through superpixel segmentation. Here, the process encouraged segmentation boundaries to conform to circular objects in the heatmap while also compelling the partitioning to distinguish bare ground from vegetated surfaces in the EXG layer. The resulting segments (superpixels) were then classified under two requirements: first, the number of pixels overlapping the thematic layer that marks tree centers (Local maxima of trees) must be greater than zero; second, the average Excess Green value within each segment must likewise be above zero. The segments satisfying both conditions were ultimately converted to point features representing individual tree locations.

Validation

A reference set of manually digitized tree crowns (n = 2952) served as the ground truth for validation. The assessment targeted trees exclusively, excluding non-tree and shadow classes, because the reference database contained only tree instances. Every tree detected by the CNN was matched to its closest manually delineated counterpart. Standard performance metrics for binary classification were adopted [24, 25], namely tallying true positives (TP) (a tree correctly retrieved), false positives (FP) (a detection that does not correspond to a real tree; a commission error), and false negatives (FN) (an actual tree left undetected; an omission error). FP, FN, and TP, respectively, indicate over-detection, under-detection, and perfect detection. The F-score (F), precision (P), and recall (R) were computed. The F-score, being the harmonic mean of recall and precision, conveys overall accuracy while jointly penalizing both commission and omission errors **(Eq. 1)**; Precision (also termed positive predictive value) reflects the proportion of detected trees that are correct and thus captures the algorithm's resilience to false positives **(Eq. 2)**; and recall (or sensitivity) reflects the fraction of actual trees that were successfully detected, indicating the algorithm's handling of false negatives **(Eq. 3)**.

$$F - score = 2 * ((R * P) / (R + P)) \quad (1)$$

$$P = TP / (FP + TP) \quad (2)$$

$$R = TP / (FN + TP) \quad (3)$$

Results and Discussion

Output of CNN

Executing the CNN workflow produces a probabilistic heat map that conveys the estimated likelihood of tree occupancy at each location within the training portion of the site. Values on this map are bounded between 0 and 1, with the lower extreme reflecting a negligible chance and the upper extreme reflecting near-certain presence. The accompanying graphic renders this heat map using a colour gradient: red zones pinpoint areas of maximal tree probability, whereas blue zones mark areas of minimal tree probability **(Figure 8)**.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

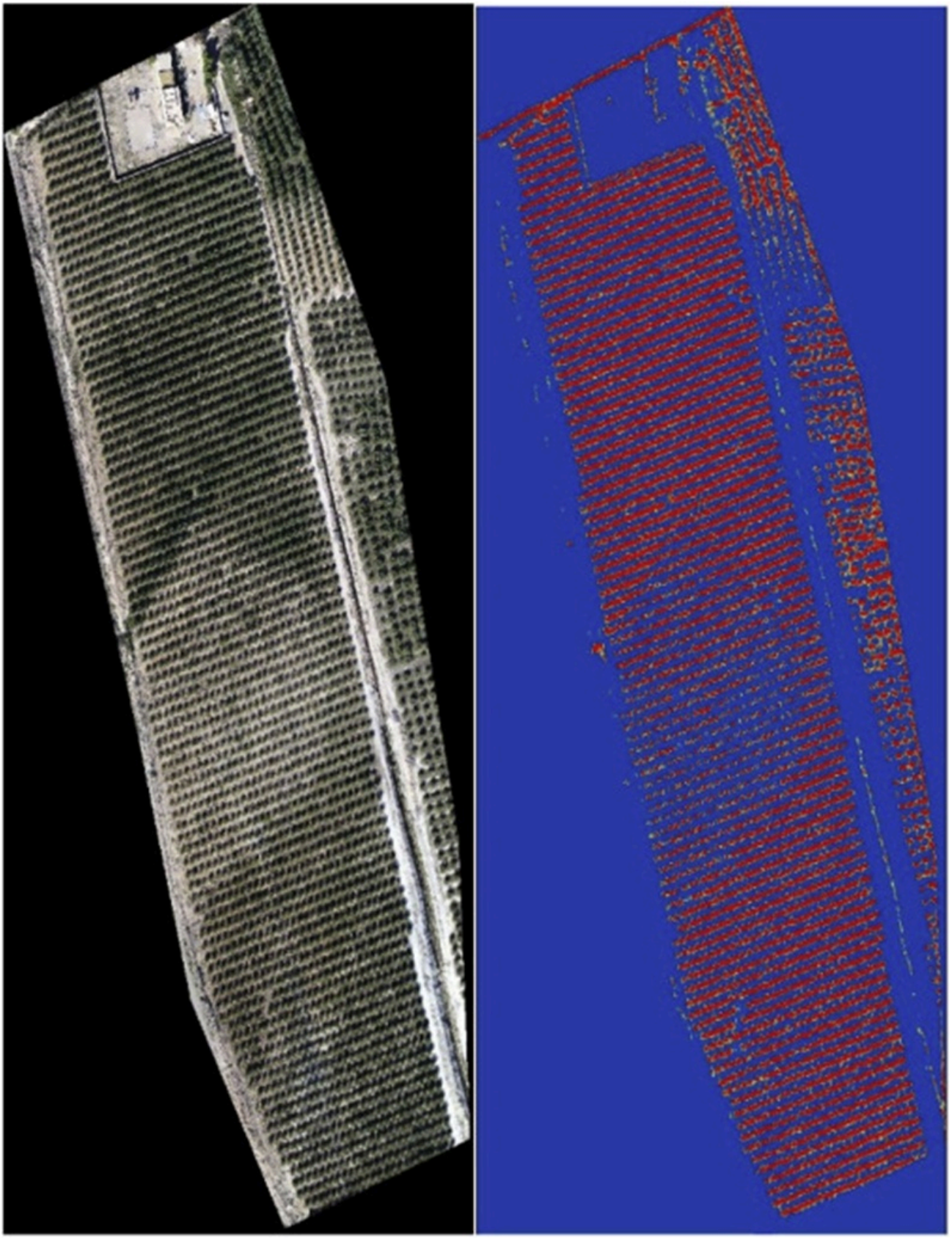

Figure 8. The heatmap result of CNN.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

In the definitive classification layer, 2204 individual tree crowns were discriminated, juxtaposed against a ground-truth tally of 2518 reference trees. Breaking this down, the TP component (trees matched correctly) amounted to 2002 instances, the FN component (trees entirely missed) equalled merely 2, and the FP component (non-tree pixels wrongly labelled as trees) totalled 316 **(Figure 9)**.

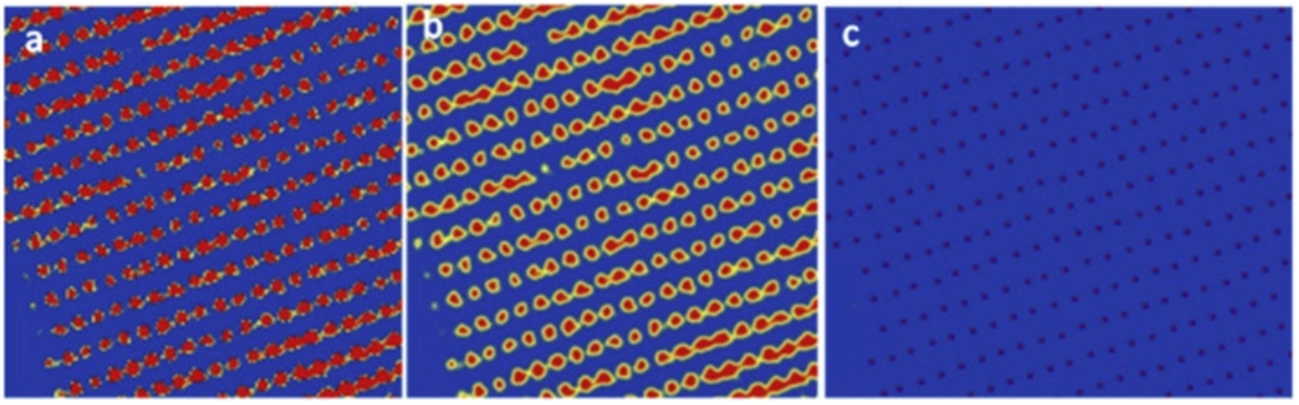

Figure 9. Results of classification by the CNN: (a) The heatmap of tree presence, (b) The heatmap smoothed by a Gaussian filter, and (c) Centers of the detected trees.

Refining the classification

The visual product of simultaneously partitioning the CNN-generated heat map and the Excess Green index band through a superpixel segmentation routine is presented in **(Figure 10)**. Here, red cross marks correspond to tree centroids placed by the CNN. At the same time, the blue-toned polygons represent the homogeneous image regions segmented by the algorithm.

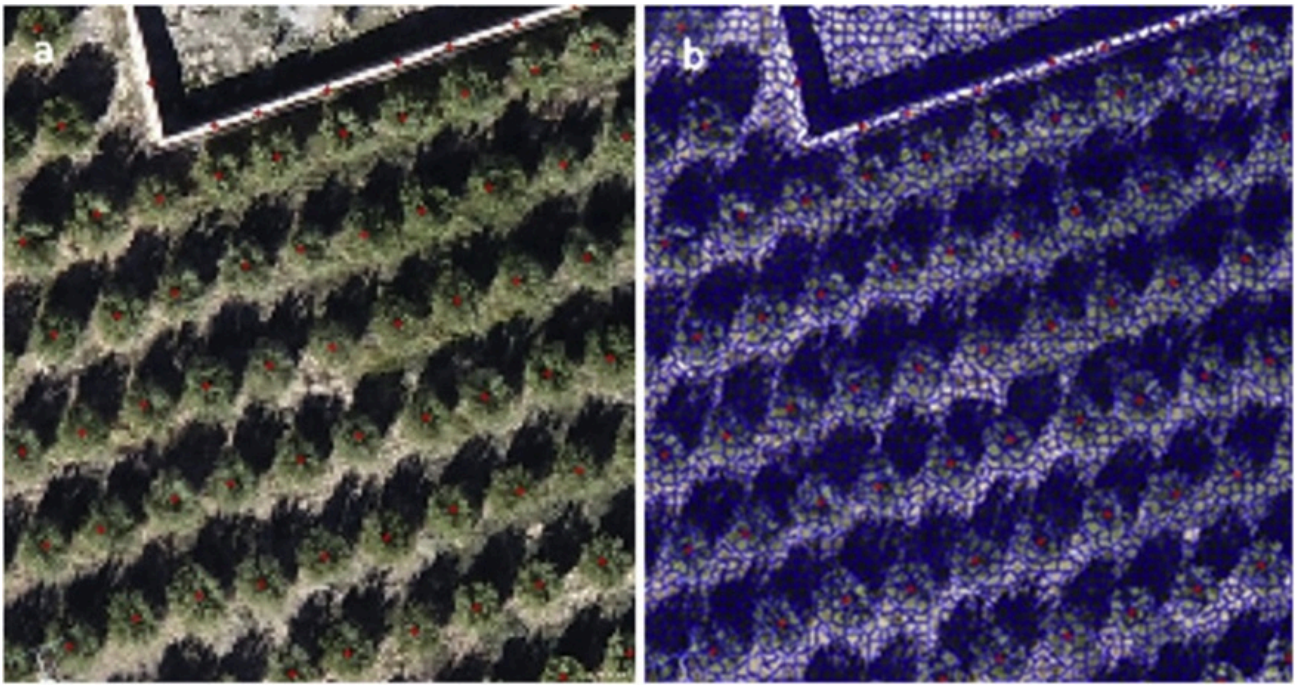

Figure 10. Result of the segmentation by Superpixel: (a) The original picture, and (b) The segmented picture.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

Once segmentation was complete, the resultant objects were assigned labels by enforcing the dual criteria previously formalized in Section 2.4. The subsequent **Figure 11** renders these successfully labelled objects in a yellow palette, superimposed on red cross marks that retain the original CNN-derived tree centre coordinates.

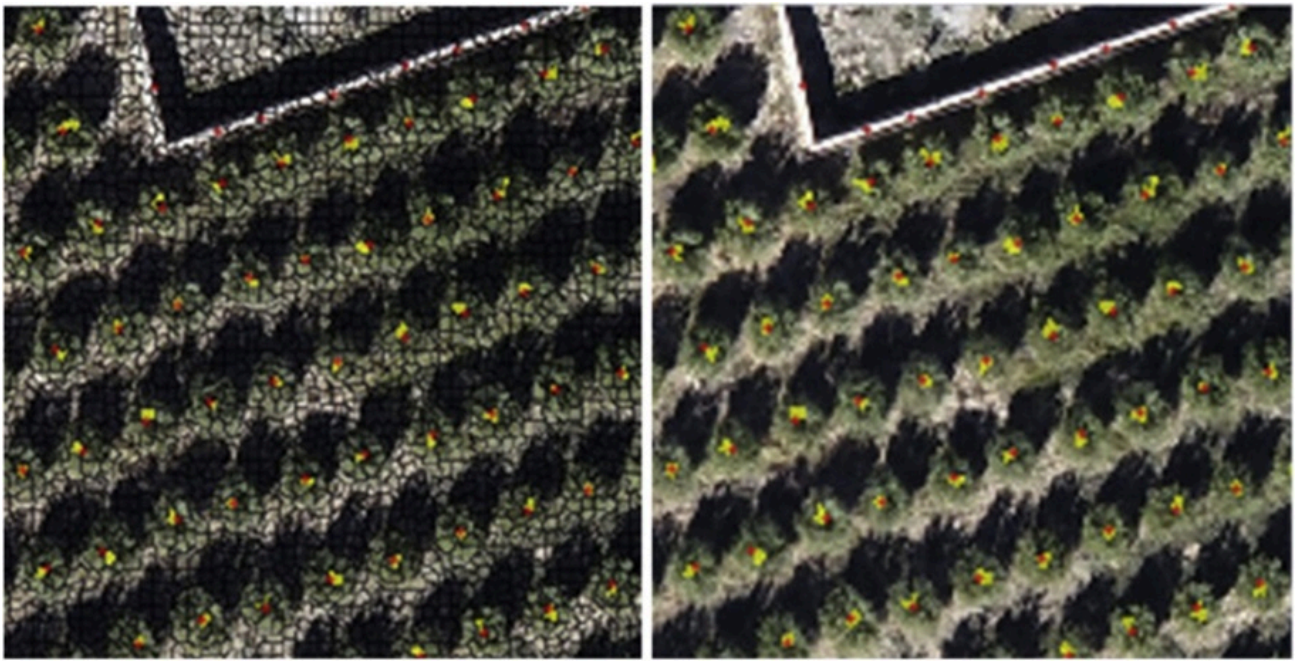

Figure 11. Classification of the objects representing the olive trees (in yellow).

A final conversion step re-expressed the classified objects as yellow cross symbols anchored at the centroid of every detected olive crown **(Figure 12)**. A visual side-by-side inspection—red crosses for the raw CNN output versus yellow crosses for the CNN+OBIA refined output—makes it unmistakably clear that the OBIA post-processing stage effectively purged spurious identifications that had appeared along portions of building rooftops and expanses of exposed soil. These corrected zones are visually emphasized in **Figure 12** with yellow circular outlines.



© 2024 The Author(s). | Published by Agro Intelligence Research Press

Figure 12. Transforming the classified objects to tree centres (in yellow).

Assessing the accuracy

The quantitative evaluation demonstrates that the refined dual-technique classification (CNN+OBIA) attains superior performance metrics: a comprehensive accuracy of 99%, a recall rate of 99%, and a harmonized F-score of 99% **(Figure 13)**. Ranking second, the unrefined CNN on its own secures a comprehensive accuracy of 95.16%, a recall rate of 99.39%, and an F-score of 97.25%.

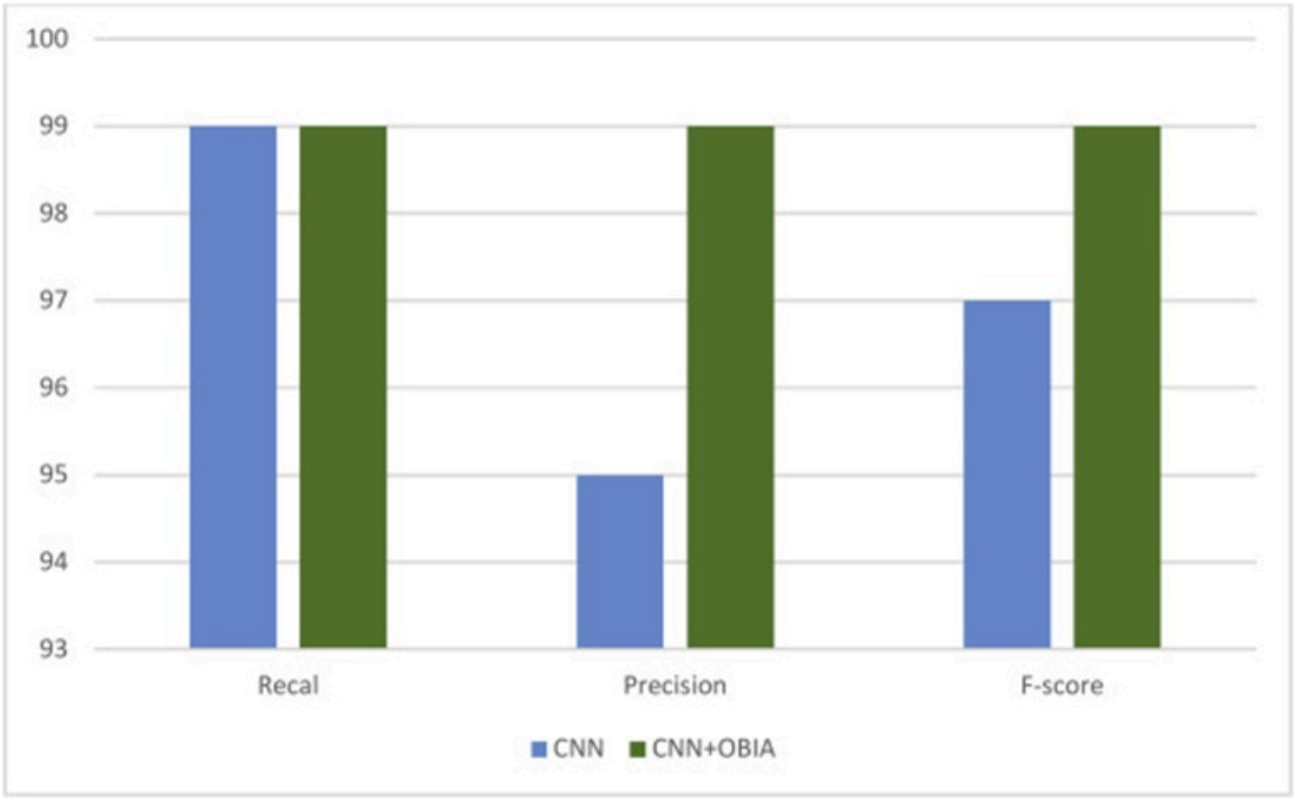

Figure 13. Precision, Recall, and F-score of the CNN and refined CNN by OBIA.

These performance indicators were placed in context by benchmarking against two alternative strategies: a “Template Matching” pipeline and a pure “OBIA” methodology built upon multi-resolution segmentation. TM achieves approximately 99% overall accuracy, yet its precision drops to 71% and its F-score to 83%. Pure OBIA, for its part, reports an overall accuracy of 84%, a precision of 96%, and an F-score of 89%, as shown in the figure that follows **(Figure 14)**.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

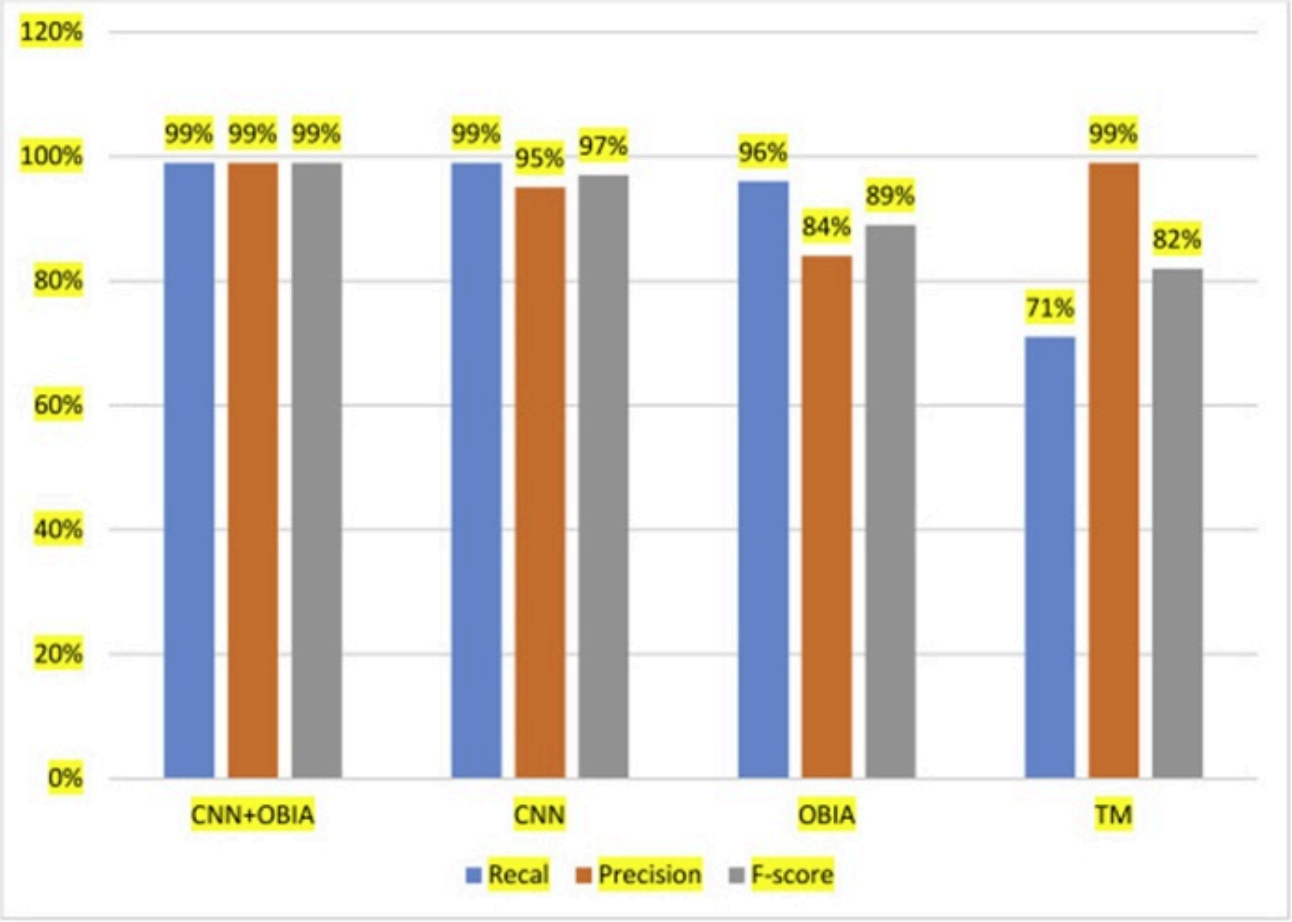

Figure 14. Precision, recall, and F-score for CNN, CNN + OBIA, Template Matching, and OBIA.

Precision was notably high for both the CNN and TM pipelines (95% and 99%, respectively), suggesting that false-positive detections were rare under either paradigm; in practice, the vast majority of flagged detections correspond to olive trees. TM edged ahead of CNN in the precision metric alone, yet this came at the expense of detection completeness, with its recall languishing at 71%. The CNN, meanwhile, maintained both high completeness and high precision simultaneously (99% recall). In essence, TM adopts a conservative detection strategy—flagging fewer trees and incurring fewer false alarms—while CNN adopts a more sensitive approach that captures nearly all trees while still filtering out most non-tree targets.

Pure OBIA, by contrast, exhibited an elevated recall statistic (0.96), indicating that it captured a large fraction of the true tree population. This thoroughness, however, was offset by relatively weaker precision (0.84), resulting in a larger pool of false-positive identifications. When using the F-score—a summary metric that combines precision and recall into a single balanced indicator of detection quality—CNN clearly outperformed the competing TM and OBIA methods (97% versus 89% and 82%, respectively).

Reflecting on the relative merits of these analytical paradigms, CNNs typically deliver high precision in tree extraction tasks. This stems from their intrinsic design: they learn hierarchically structured, discriminative features straight from labelled exemplars, granting them a refined ability to disentangle tree signals from background interference. On the other side of the ledger, CNNs constitute sophisticated model architectures that incur considerable computational overhead during both the parameter-learning and prediction phases. Their typical construction stacks many convolutional filters, pooling operations, and non-linear activation transformations, each layer adding to the aggregate processing burden. CNNs further exhibit pronounced sensitivity to a constellation of external and internal factors—paramount among them

© 2024 The Author(s). | Published by Agro Intelligence Research Press

the representativeness and volume of the training corpus, the calibration of hyperparameters, and the network's architecture. Guarding against overfitting and steering the model toward its optimal configuration demands deliberate tuning and the application of regularization measures.

By shifting the focus to OBIA-centric approaches, these techniques can yield credible precision in tree mapping by jointly leveraging spectral reflectance patterns, spatial arrangement cues, and contextual neighbourhood information. Their foundational step—segmenting the continuous image plane into discrete, semantically coherent parcels—enables classification logic to operate on object-level descriptors rather than isolated pixel values, often improving fidelity. The OBIA workflow is inherently multi-staged, typically requiring sequential execution of image segmentation, attribute computation, and rule-based or machine-learning classification. Though it may not match the sheer computational weight of deep CNNs, OBIA is far from trivial, demanding judicious parameterization of segmentation routines and deliberate selection of discriminating features. The particular segmentation algorithm adopted, along with the suite of object attributes fed into the classifier, can dramatically sway end accuracy. Detection performance may also oscillate depending on the spatial granularity and land-cover diversity of the scene under study.

Template matching sits at a different point in the methodology spectrum. It can achieve high precision when the handcrafted template faithfully mirrors the typical visual appearance of the target class (olive crowns, here). Its Achilles' heel, however, lies in its brittleness: natural variability in crown morphology, illumination geometry, and partial occlusion by shadows or adjacent canopies tend to degrade its detection rate, generally yielding lower overall accuracy figures than either CNNs or OBIA. Judged on complexity, TM is comparatively parsimonious. The algorithmic core is a straightforward convolution-like scan in which the template kernel is rasterized across the image lattice. At the same time, a local similarity index is computed at each position—an operation that can be executed with modest computational cost. Template matching remains susceptible to nuisance variation in radiometric conditions, geometric scale, and viewing perspective. The selection of template dimensionality and the specific pixel-derived features encoded in it exert first-order control on the quality of the resulting detections. It exhibits limited capacity to generalize beyond the range of tree appearances and occlusion scenarios represented in its template library.

In summary, CNNs—and their hybridized CNN+OBIA counterparts—generally set the upper bound for tree detection precision, but this performance premium comes with substantial computational demands and multifaceted sensitivities **(Table 3)**. OBIA methodologies present a compromise, trading a degree of peak accuracy for more manageable complexity while exploiting complementary spatial and spectral evidence streams. Template matching, by design, is the most lightweight and transparent option but frequently sacrifices both precision and resilience to scene variability **(Table 3)**. Selecting the most suitable paradigm hinges on project-specific constraints: the computational budget available, the structural complexity of the landscape mosaic, and the practitioner's priority ranking along the precision-versus-efficiency continuum.

Table 3. Comparing the performance, complexity, and sensitivity for CNN, CNN + OBIA, template matching, and OBIA.

Empty cell	CNN+OBIA	CNN	OBIA	TM
Performance	++++	+++	++	+
Complexity	+++	+++	++	+
Sensitivity	+++	+++	++	+

The accuracies ultimately obtained in this study are highly encouraging (both overall accuracy and F-score reaching 99%) and stand shoulder to shoulder with previously published investigations that drew upon UAV-borne imagery for tree inventory—a domain widely acknowledged as posing exceptional difficulty within remote sensing and precision farming circles. Earlier contributions enriched their analyses by fusing multispectral reflectance data with DSMs (Digital Surface Models) [26], CHMs (Canopy Height Models) [27], or LIDAR-derived point cloud datasets [28]. To the best of our

© 2024 The Author(s). | Published by Agro Intelligence Research Press

knowledge, the work reported here breaks new ground as the first to successfully deploy a combined CNN-OBIA framework operating exclusively on three-band RGB imagery and thereby achieving an elevated level of accuracy in the location and enumeration of solitary olive trees.

Conclusion

The adoption of UAV technology for crop surveillance and management continues to accelerate; these platforms possess the transformative capacity to reduce the time required for tree censusing compared with traditional field-based tallying, especially when integrated with deep learning algorithms. In this concise communication, a hybrid analytical scheme that marries CNN-based detection with OBIA-based refinement was fielded to locate olive trees in RGB-format drone photographs autonomously.

The terminal accuracy figures are eminently satisfactory (overall accuracy of 97.25% attributable to the CNN module alone, climbing to 99% for the full CNN+OBIA pipeline) and align closely with the performance benchmarks reported by prior studies that tackled tree detection from UAV-derived data—a problem consistently ranked among the most intractable in remote sensing and precision agricultural research. Earlier efforts augmented their data stacks by incorporating multispectral channels alongside DSMs (Digital Surface Models) [26], CHMs (Canopy Height Models) [27], or LIDAR point clouds [28]. Uniquely, the present endeavour demonstrates that a CNN-driven workflow, fuelled solely by standard RGB imagery, can achieve first-rate accuracy in singling out and counting discrete olive trees. Looking ahead, forthcoming research could productively explore the adoption of deeper, region-oriented CNN variants—such as the R-CNN family of architectures—to move beyond point-based detection toward full crown delineation and per-tree enumeration.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

Hereby, I, Oumaima AMESLEK, consciously assure that for the manuscript “Identification and mapping of Individual Trees from Unmanned Aerial Vehicle Imagery Using an object-oriented Convolutional Neural Network” the following is fulfilled:

1) This material is the authors’ own original work, which has not been previously published elsewhere.

2) The paper is not currently being considered for publication elsewhere.

3) The paper reflects the authors’ own research and analysis truthfully and completely.

4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

5) The results are appropriately placed in the context of prior and existing research.

6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

Violating the Ethical Statement rules may result in severe consequences.

Received: 21 Dec 2022 Revised: 25 Feb 2023 Accepted: 14 Mar 2023
Published online: 25 May 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Velusamy P, Rajendran S, Mahendran RK, Naseer S, Shafiq M, Choi JG. Unmanned aerial vehicles (UAV) in precision agriculture: Applications and challenges. Energies. 2022;15(1). https://doi.org/10.3390/en15010217.

2. Marconi S, Graves SJ, Weinstein BG, Bohlman S, White EP. Estimating individual level plant traits at scale. bioRxiv. 2020.

3. Vijayakumar V, Ampatzidis Y, Schueller JK, Burks T. Smart spraying technologies for precision weed management: a review. Smart Agric Technol. 2023;6:100337. https://doi.org/10.1016/j.atech.2023.100337.

4. Ariza-Sentís M, Valente J, Kooistra L, Kramer H, Mücher S. Estimation of spinach (Spinacia oleracea) seed yield with 2D UAV data and deep learning. Smart Agric Technol. 2023;3.

5. Kool J, Evenhuis A. Early detection of Phytophthora infestans in potato plants using hyperspectral imaging, local comparison and a convolutional neural network. Smart Agric Technol. 2023;6:100333. https://doi.org/10.1016/j.atech.2023.100333.

6. Bose A, Roy Chowdhury I. Monitoring and modeling of spatio-temporal urban expansion and land-use/land-cover change using Markov chain model: a case study in Siliguri metropolitan area, West Bengal, India. Model Earth Syst Environ. 2020;6(4):2235–49.

7. Tsouros DC, Bibi S, Sarigiannidis PG. A review on UAV-based applications for precision agriculture. Information (Switzerland). 2019;10(11). https://doi.org/10.3390/info10110349.

8. Ocer NE, Kaplan G, Erdem F, Kucuk Matci D, Avdan U. Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. Remote Sens Lett. 2020;11(9):847–56.

9. Santos AA dos, Marcato Junior J, Araújo MS, Di Martini DR, Tetila EC, Siqueira HL, et al. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. Sensors (Switzerland). 2019;19(16):1–11.

10. Liu T, Abd-Elrahman A, Morton J, Wilhelm VL. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. GISci Remote Sens. 2018;55(2):243–64. https://doi.org/10.1080/15481603.2018.1426091.

11. Csillik O, et al. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. Drones. 2018;2(4):1–16.

12. Fu T, Ma L, Li M, Johnson BA. Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. J Appl Remote Sens. 2018;12(2):1.

13. Zhang L, Zhang L, Du B. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci Remote Sens Mag. 2016;4(2):22–40.

14. Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F, et al. Deep learning in remote sensing: A review. IEEE Geosci Remote Sens Mag. 2017. https://doi.org/10.1109/MGRS.2017.2762307.

15. Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural Netw. 1988;1(2):119–30.

16. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. The history began from AlexNet: A comprehensive survey on deep learning approaches. 2018.

17. Lei G, Li A, Bian J, Zhang Z, Jin H, Nan X, et al. Land cover mapping in southwestern China using the HC-MMK approach. Remote Sens (Basel). 2016;8(4).

18. Warth G, Braun A, Assmann O, Fleckenstein K. Prediction of socio-economic indicators for urban planning using VHR satellite imagery and spatial analysis. Remote Sens (Basel). 2020;12(11). https://doi.org/10.3390/rs12111730.

19. Ye S, Pontius RG Jr, Rakshit R. A review of accuracy assessment for object-based image analysis: from per-pixel to per-polygon approaches. ISPRS J Photogramm Remote Sens. 2018;141:137–47.

20. Juniati E, Arrofiqoh EN. Comparison of pixel-based and object-based classification using parameters and non-parameters approach for the pattern consistency of multi scale landcover. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Archives. 2017;42(2W7):765–71.

21. Htitiou A, Boudhar A, Lebrini Y, Lionboui H. Classification and status monitoring of agricultural crops in central Morocco: a synergistic combination of OBIA approach and fused Landsat-Sentinel-2 data. J Appl Remote Sens. 2021;15(1):014504. https://doi.org/10.1117/1.JRS.15.014504.

22. Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, et al. Geographic object-based image analysis – towards a new paradigm. ISPRS J Photogramm Remote Sens. 2014;87:180–91. https://doi.org/10.1016/j.isprsjprs.2013.09.014.

23. Pearse D, Tan AYS, Watt MS, Franz MO, Dash JP. Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. ISPRS J Photogramm Remote Sens. 2020;168:156–69.

24. Li W, Guo Q, Jakubowski MK, Kelly M. Photogrammetric engineering and remote sensing. Photogramm Eng Remote Sens. 2012;78(1):75–84.

© 2024 The Author(s). | Published by Agro Intelligence Research Press

25. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3):1–21.

26. Caruso G, Zarco-Tejada PJ, González-Dugo V, Moriondo M, Tozzini L, Palai G, et al. High-resolution imagery acquired from an unmanned platform to estimate biophysical and geometrical parameters of olive trees under different irrigation regimes. PLoS ONE. 2019;14(1):1–19.

27. Gu J, Grybas H, Congalton RG. A comparison of forest tree crown delineation from unmanned aerial imagery using canopy height models vs. spectral lightness. Forests. 2020;11(6).

28. Timilsina S, Sharma SK, Aryal J. Mapping urban trees within cadastral parcels using an object-based convolutional neural network. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. 2019;4(5/W2):111–7.

© 2024 The Author(s). | Published by Agro Intelligence Research Press