Evaluating Multispectral Data Fusion for Dense Instance Segmentation in Vegetation and Artificial Objects Point Clouds

Clodoaldo Souza Faria Junior[1], Isabella Subtil Norberto[2], Marcos Ricardo Omena de Albuquerque Maximo[1], Antonio Maria Garcia Tommaselli[2], Mauricio Galo[2]

[1]Aeronautics Technological Institute, São José dos Campos, São Paulo 12228-900, Brazil – clodoaldo.junior.101406@ga.ita.br, mmaximo@ita.br
[2]Faculty of Science and Technology, São Paulo State University (UNESP) at Presidente Prudente, São Paulo 19060-900, Brazil – (isabella.subtil, a.tommaselli, mauricio.galo)@unesp.br

Keywords: Terrestrial data, deep learning classification, multispectral point cloud, instance segmentation, digital agriculture.

Abstract

Multispectral data improves instance segmentation in digital agriculture by combining geometric and spectral information to distinguish complex natural features. While geometric information captures structural details, it often falls short when dealing with complex natural features that exhibit high spectral similarity, rather than due to limitations inherent to geometric representation itself. This work presents a feasibility analysis of instance segmentation using a spectral point cloud. A combination of spectral bands is selected based on class separability and proximity to a normal distribution as estimated by the Shapiro–Wilk test. The aim is to identify the minimum number of bands required to produce optimum results. For the normality analysis, Euclidean magnitude normalisation was applied, and it was also used alongside standard scaling to support the Multilayer Perceptron (MLP) for classification and segmentation. To refine the MLP predictions and consolidate instance labels, a graph-based post-processing step was applied, linking each point to its nearest neighbours and using a majority-voting scheme, resulting in spatially coherent clusters and refining the MLP predictions. The results demonstrate that multispectral data can reliably segment individual objects, with ten spectral bands being sufficient to achieve highly satisfactory segmentation and accurately delineate natural features such as leaves and tree trunks. Further increasing the number of bands improved spectral definition even more, with 14 bands achieving the highest performance across all metrics (mIoU: 96.59%; AP_{50}: 96.14%). These findings highlight the strong potential of multispectral point clouds for precise and scalable object-level segmentation in agricultural environments.

1. Introduction

Three-dimensional (3D) point clouds have become an important part of advanced automation, with possible uses in areas including agricultural robotics, especially for difficult tasks like self-driving harvesting (Abbasi et al., 2022). In this context, deep learning models are notable for their ability to learn discriminative representations from data, capturing complex structural and spectral patterns (Charisis and Argyropoulos, 2024), thereby overcoming the limitations of traditional methods. Semantic segmentation (Guo et al., 2018), object detection (Zhao et al., 2019) and, most notably, instance segmentation (Gu et al., 2022) are techniques that significantly broaden the scope of computer vision. These techniques enable precise and individualised identification of elements in dense scenes.

These advances have driven the development of intelligent systems in several domains, including robotics and precision agriculture. Single-mode data strategies still face major problems, particularly when dealing with small, hidden, or contextually complex objects (Cui et al., 2023). In this scenario, 3D segmentation models must efficiently integrate multiple sources of information. Architectures such as PointNet (Qi et al., 2017*a*) and PointNet++ (Qi et al., 2017*b*) introduced important milestones. These included point-wise feature extraction and hierarchical grouping strategies. Subsequent models have enhanced robustness against structural disorder and both local and global variations. Examples of these models include DGCNN (Wang et al., 2019), PAConv (Xu et al., 2021) and Transformer-based networks.

In addition to these architectural advancements, there is an increasing interest in portable devices that are equipped with multimodal sensors, which allow for the flexible collection of high-resolution spectral and geometric data (Xie et al., 2024). The appeal of these systems lies in their suitability for use in plant phenotyping and digital agricultural monitoring. However, acquiring complete point clouds in complex environments is challenging, as it requires unobstructed and well-positioned views (Schor et al., 2017). This challenge emphasises the importance of intermodal fusion strategies to complement incomplete or low-quality data.

In this context, fusion methods have been used to combine spectral and depth information by integrating imagery with LiDAR data. The effectiveness of 3D data exploitation has been enhanced through the application of instance segmentation techniques to point clouds, which enable a more structured and interpretable representation of objects in the scene (Jia et al., 2025). Some deep learning models use multimodal inputs, combining RGB images and 3D point clouds to improve segmentation accuracy. For example, Frustum PointNet (Qi et al., 2018) integrates 2D detections with 3D segmentation.

This study contributes to the field by proposing an instance segmentation method for dense multispectral point clouds, based on a lightweight architecture for analysing spectral information, a feature still largely underexplored in computer vision, especially within digital agriculture. Additionally, a graph-based post-processing step is incorporated, aiming to obtain accurate instance-level segmentations.

The remainder of this work is organised as follows. Section 2 presents the background and theoretical foundations related to multispectral point clouds, instance segmentation, and multimodal data fusion in agricultural applications. Section 3 describes the materials and methods, including data acquisition, colourisation procedures, preprocessing steps, network

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-279-2026 | © Author(s) 2026. CC BY 4.0 License.

architecture, training configuration, post-processing techniques, and evaluation metrics. Section 4 presents the results and discussion, encompassing training performance, segmentation quality, comparative analyses across datasets, and noise quantification. Finally, Section 5 summarises the conclusions and outlines potential directions for future research.

2. Background

2.1 Multispectral Point Cloud

The combination of geometric and spectral information into a single data structure is made possible by 3D multispectral point clouds, enabling physio-biochemical analyses of plant structures. This integration has proven to be promising in digital agriculture applications, particularly in the development of Digital Replica, which are accurate virtual representations of real-world objects and elements (Verdouw et al., 2021).

Some studies have focused on the fusion of LiDAR data with satellite imagery, given its wide availability (Reji and Nidamanuri, 2023). Although optical sensors such as multispectral and hyperspectral cameras have shown effectiveness in crop classification at the field scale (Handique et al., 2017), their application at the plant level presents additional challenges for terrestrial acquisitions. Variations in lighting conditions, vegetation structure, and viewing angles significantly affect spectral quality (Xie et al., 2024), making it essential to adopt suitable, unobstructed viewpoints (Schor et al., 2017).

According to Teixeira et al. (2023), classification and segmentation studies confirm that multimodal fusion strategies are effective in overcoming the limitations of individual sensors and enhancing segmentation performance in complex agricultural environments. Nevertheless, factors such as spectral and spatial resolution, the presence of non-vegetation elements, and sample quality directly influence the outcomes. Furthermore, some studies have demonstrated that classification relying solely on radiometric data tends to underperform and it is more susceptible to errors (Arrizza et al., 2024), highlighting the need to integrate geometric information.

Combining 2D and 3D information is fundamental to precision agriculture, especially for robotic harvesting, which requires the accurate identification and localisation of objects in complex environments. This process relies heavily on instance segmentation, as this is the only way to isolate individual objects and analyse their shape and position. Fusing point cloud data with spectral imagery increases the robustness of models, especially in challenging conditions, including occlusions and varying lighting.

2.2 Instance Segmentation Using Point Clouds

Among the explored architectures in the literature, DaSNet-v2, proposed by Kang and Chen (2020) , stands out for its YOLO-based approach to apple detection and segmentation, leading accuracy values above 87% (mIoU = 87.3%). Subsequent work was carried out by the same group, with variants such as Mobile-DasNet, which was combined with PointNet for 3D pose estimation, resulting in AP_{50} of 86.3% and IoU of 82% (Kang et al., 2020). The potential of multimodal approaches for processing agricultural scenes is highlighted by this fusion between convolutional networks for 2D feature extraction and networks operating directly on point clouds.

Other relevant examples include the use of Mask R-CNN by Coll-Ribes et al. (2023) for grape harvesting, which showed a 6.6% improvement in AP_{50} (from 85.9% to 92.5%) when depth information was added to the input model. Additionally, the Apple 3D Network (A3N), developed by Wang et al. (2022), combined YOLACT for instance segmentation with a modified PointNet for grasping pose estimation, resulting in an IoU of 87.3% for the masks. In this study, objects are modelled as independent semantic classes during training, with instance-level coherence enforced through spatial clustering and graph-based post-processing, consolidating point-wise classifications into coherent object-level segments.

3. Materials and Methods

3.1 Materials

3.1.1 Field Data

The dataset was acquired using high-performance equipment widely employed in 3D mapping applications: the Agrowing 7Rxxx Sextuple multispectral camera (Agrowing Development Team, 2025) and the Faro Focus Premium terrestrial laser scanner (TLS). The Agrowing camera is equipped with six lenses, capturing a total of fourteen spectral bands. Point clouds were coloured using selected bands from each of the six sub-images (Figure 1). The camera’s focal length is 21.6 mm, and the pixel size is 0.0037 mm. Figure 1 shows the wavelengths (in nm), available in each sub-image (S).

Figure 1 – The six sub-images (S1, S2, …, S6) of the multispectral Agrowing camera and the respective wavelengths (in nm) for each sub-image.

A black flat wooden board (Figure 2) was positioned facing North with several targets. Data collection was carried out from three distinct angles relative to this reference: 15° (Fig. 2b), 0° (Fig. 2c) and 345° (Fig. 2d). An example of the resulting point cloud is shown in Fig. 2a with station at 3 m with axis normal to the board - 0°, resulting in a PSOSU (Pixel Size in Object Space Units) of 0.51 mm. This experiment was designed to simulate objects in an agricultural environment. The board's surface was affixed with various natural and artificial elements using double-sided tape. These included 49 objects belonging to eight distinct categories, such as healthy and senescent leaves. They included sandpaper, pieces of tree bark, white, grey and black EVA (ethylene-vinyl acetate) foam, pebbles, lemon and orange peel and coded targets, following the Agisoft Metashape standard.

For the colouring process, the calibration targets were used as a reference in bundle adjustment. The board's black background was excluded as an object of interest during the training, validation, and testing stages to prevent bias and ensure an impartial segmentation of the relevant elements.

The experiments were conducted in a controlled setup, with all the training, validation and testing data acquired from the same scene and object configuration. This design allows for a focused

evaluation of the proposed methodology and ensures consistency across datasets.

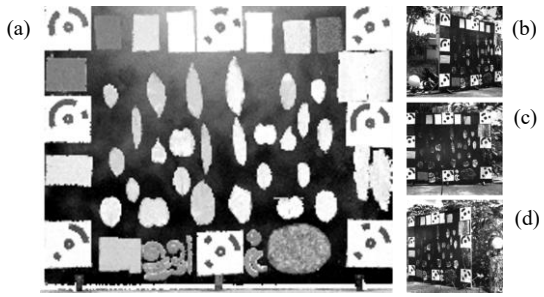

Figure 2 – (a) Point cloud acquired from the 0° position (facing North), and scanner placements at (b) 15°, (c) 0°, and (d) 345° relative to North.

3.2 Methods

The overall methodological workflow is summarised in Figure 3, which provides a high-level overview of the main processing stages adopted in this study. The figure is intended to illustrate the pipeline’s logical sequence rather than to convey all implementation details. Each stage is therefore described explicitly in the following subsections, guiding the reader through the acquisition, colourisation, preprocessing, model training, and post-processing steps that compose the proposed methodology.

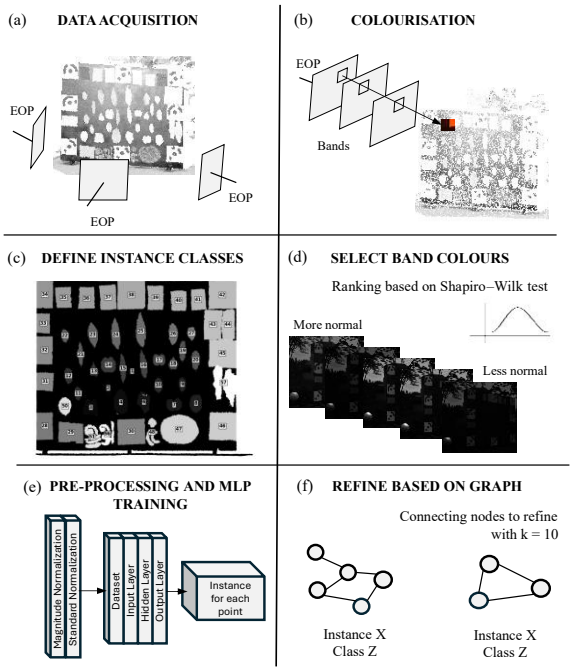

Figure 3 – Workflow for Multispectral Point Cloud Instance Segmentation. The procedure comprises six main stages: (a) data acquisition using terrestrial laser scanning and multispectral imaging; (b) bundle adjustment and point cloud colourisation based on the estimated exterior orientation parameters; (c) manual definition of object classes and instances; (d) spectral band selection and dataset construction guided by normality analysis; (e) preprocessing and training of a lightweight MLP model for point-wise classification; and (f) graph-based post-processing to enforce spatial coherence and consolidate instance-level segmentation results.

3.2.1 Colourisation

The methodology relies on associating to each 3D point in the point cloud a Digital Number (DN) provided by the acquired images. Therefore, it is essential to determine the EOPs and Interior Orientation Parameters (IOPs) through bundle adjustment (BA). The next step of the process defines the sensor’s point of view relative to the scene. Since the EOPs and IOPs are determined, the DNs from the sub-images are to each point from the LiDAR point cloud, limited to the areas within the camera’s field of view (FoV). To avoid duplicated projections on overlapping regions, we adopted the occlusion-handling strategy described by Katz et al. (2007), which identifies and labels non-visible points. These occluded points were excluded from the projection step. Consequently, each sub-image is considered an independent capture of the scene, offering a unique geometric perspective of the point cloud.

3.2.2 Data preprocessing, model architecture, training, and post-processing

The input vectors were defined according to the number of spectral bands under analysis, which were selected using the Shapiro–Wilk test (Mahlayeye et al., 2024) to prioritise those exhibiting behaviour closer to a normal distribution. Based on this criterion, each dataset was constructed by incrementally adding one additional spectral band, resulting in configurations ranging from a single band to a total of fourteen spectral bands, resulting in multiple datasets with different combinations of bands, thus forming the MV (Multispectral Values) dataset. This approach aimed to assess the contribution of each spectral band and identify the minimum number of bands required to achieve robust results, optimising data efficiency and model performance.

Initially, the input vectors were normalised by their Euclidean magnitude (L2 norm normalisation) aimimg to eliminate variations related to the absolute value of the DN in the subimages, preserving only the relative shape of the spectral signatures. This type of normalisation is particularly important for multispectral and hyperspectral data, as factors such as illumination, sensor distance, and viewing angle can affect the amplitude of values without changing their intrinsic spectral relationships (Chanchí Golondrino et al., 2023).

Subsequently, the L2-normalised data were standardised using Z-score standardisation, ensuring zero mean and unit variance for each spectral feature (Aksu et al., 2019). While L2 normalisation operates at the level of individual spectral vectors, Z-score standardisation acts across features, reducing disparities in scale and variance among spectral bands. This two-stage normalisation strategy was adopted to prevent features with higher variance from disproportionately influencing the learning process and to promote stable and balanced convergence of the MLP during training. Before encoding, each object was manually identified and assigned to one of 50 distinct classes in a labelling process, providing a clear mapping between objects and their respective classes, as shown in Figure 4.

To address class imbalance, the Random Oversampling technique was applied to equalise the number of samples across all class labels, which correspond to those illustrated in Figure 4. Before oversampling, one of the objects contained 58,944 points (tree bark), while the smallest had only 1,119 (leaf), a difference primarily attributable to the physical size of each object within the point cloud and the resulting variation in point density. All

classes were thus adjusted to contain the same number of samples as the majority class, ensuring balanced datasets for both training and validation (Kashongwe et al., 2024). In this configuration, two independent point clouds were used as input: the first was randomly split into 70% for training and 30% for validation, while the second point cloud was used exclusively for testing. This separation prevents data leakage, enabling a more rigorous evaluation of the model’s generalisation capability. Adopting oversampling retains valuable patterns from minority classes that would otherwise be underrepresented through undersampling, while effectively mitigating model bias towards majority classes (He and Garcia, 2009).

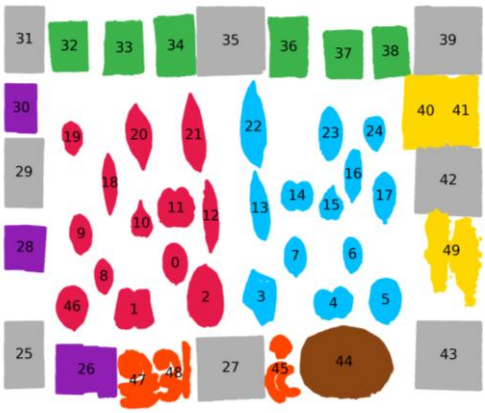

Figure 4 – Labelled dataset representing the classes and instances considered for the instance segmentation. Labels highlighted in red correspond to the healthy leaves class, while blue represents senescent (dead) leaves. The green label denotes EVA material, orange corresponds to orange and lemon peels, and brown identifies aggregations of small rocks. The yellow label represents tree bark, purple indicates sandpaper surfaces, and grey marks reference targets employed to support calibration and processing procedures.

A Multilayer Perceptron (MLP) was employed owing to its architectural simplicity and its suitability for the analytical framework adopted in this study, in which each point in the cloud is treated as an independent sample, with its spectral and geometric attributes serving as input features. Although the entire point cloud is processed in batches during training, the network classifies each point individually, producing class-probability scores for the target classes. The architecture was selected empirically through a trial-and-error process to achieve high performance while maintaining low computational demands. Unlike more complex and computationally intensive architectures such as PointNet (Qi et al., 2017), PointNet++ (Qi et al., 2017*b*), and DGCNN, the proposed MLP offers a lightweight alternative that efficiently captures relevant spectral information without the overhead of deep geometric feature extraction.

The architecture used in this paper comprises an input layer whose dimensionality is determined by the number of spectral bands present in the dataset, indicating that the size of the input vector is directly conditioned by the specific attributes included in each data configuration. This layer is followed by two fully connected layers with 128, and 64 neurons (hidden layer as mentioned in Figure 3.d), respectively. All hidden layers use the ReLU activation function and are followed by Batch Normalization (Ioffe and Szegedy, 2015). The output layer contains several neurons equal to the target classes and employs the SoftMax activation function, which produces class probabilities for each point. Discrimination between instances of the same class is subsequently performed in the post-processing stage, which groups the classified points.

The model was compiled using categorical cross-entropy as the loss function and the Adam optimiser (Kingma and Ba, 2017). Training employed early stopping, halting after ten (10) consecutive epochs without improvement in validation loss, and restoring the best weights automatically. Other relevant parameters, including learning rate, batch size, and number of epochs, were tuned empirically to optimise performance while ensuring reproducibility using random state 42.

To refine initial MLP predictions, a graph-based post-processing step was applied as shown in Figure 3.d. Conceptually inspired by graph-based neighbourhoods in Dynamic Graph CNN (DGCNN) for point cloud feature learning (Wang et al., 2019), this method operates as a post-processing step rather than as part of the network architecture. Each point was represented as a node in an undirected graph, with edges connecting each node to its k = 10 nearest neighbours, chosen empirically to balance label smoothing and preservation of instance boundaries. Nearest neighbours were computed using the Euclidean distance in the 3D space. Then, connected components were identified using the NetworkX library, which efficiently detects spatially coherent clusters of points (Hagberg et al., 2008; Feldbauer et al., 2020).

Within each connected component, a majority voting scheme was applied: the point labels were updated to the most frequent class only if they exceeded a vote threshold of 70% within the component; otherwise, the original labels were retained. Here, the component is formed by connecting each point to its k = 10 nearest neighbours in 3D space, so the 70% vote considers all points within this spatially connected cluster, not just the close neighbours of a single point. This threshold was empirically determined to reduce isolated misclassifications while preserving small instances and fine-grained structures. The procedure effectively smooths and consolidates the segmentation, ensuring that each instance is represented by a spatially coherent cluster of points and improving both visual consistency and quantitative evaluation metrics.

3.2.3 Metrics

For the evaluation of the results obtained in this paper, two approaches were employed: a visual inspection to assess the consistency of the segmentations, and a quantitative analysis using performance metrics. The metrics considered were Mean Intersection over Union (mIoU) and Average Precision at IoU 0.5 (AP_{50}) to assess the segmentation quality, calculated as follows:

$$\text{IoU} = \frac{|P_i \cap G_i|}{|P_i \cup G_i|}, \qquad mIoU = \frac{1}{N}\sum_{c=1}^{N}\frac{|P_c \cap G_c|}{|P_c \cup G_c|}, \tag{1}$$

$$AP = \int_0^1 P(r)\, dr, \tag{2}$$

where N is the total number of classes according to the Confusion Matrix; P represents the precision-recall curve; P_i and G_i correspond to the prediction and ground truth of the pixel i. Further details regarding the confusion matrix and these metrics can be found in Heydarian et al., (2022).

4. Results and Discussion

4.1 Multispectral Data

After the bundle adjustment, the Root Mean Square Error (RMSE) for the control points ranged from 0.4 to 0.6 mm for the planimetric and altimetric coordinates. Figure 5 presents the resulting multispectral point clouds acquired from the 0° position (facing North).

These values can be considered of high quality because the camera was previously calibrated, and the interior orientation of each sub-image was used in the BA of each data set. The results proved to be highly satisfactory for the research purpose, enabling the generation of a multispectral point cloud with geometric accuracy aligned to the study’s goals.

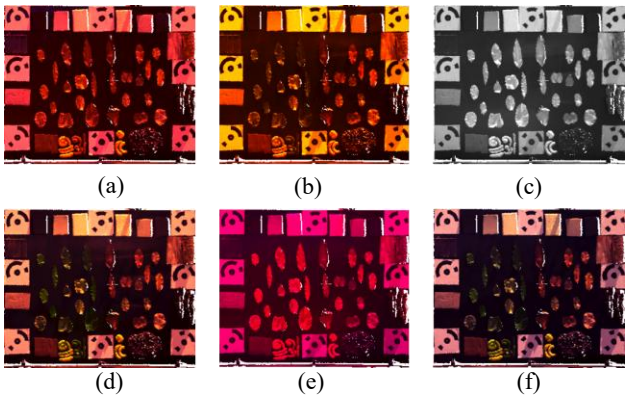

Figure 5 – Multispectral point cloud coloured by sub-images sets: (a) S1 (405, 570, and 710 nm), (b) S2 (525 and 630 nm), (c) S3 (850 nm), (d) S4 (430, 550, and 650 nm), (e) S5 (490 and 735 nm), and (f) S6 (450, 560, and 685 nm).

4.2 Select multispectral bands

Bands were selected via the Shapiro–Wilk test after DN magnitude normalisation, choosing those closest to normality for stable grouping. The resulting sequence of bands was 685, 710, 735, 430, 550, 650, 850, 630, 560, 450, 570, 405, 490, and 525 nm. Therefore, combinations of 7–14 Spectral Band Values (SB) were formed in this order, enabling the cumulative impact of including less normalised bands on the model's performance to be evaluated. For example, the 7MV dataset comprises the bands 685, 710, 735, 430, 550, 650, and 850 nm, whereas the 8MV configuration includes the next band in the sequence, 630 nm, and this procedure continues until all fourteen spectral bands are included.

4.3 Training Analysis

Training epochs varied across the multispectral point cloud datasets, with 7MV (75 epochs), 8MV (61), 9MV (51), 10MV (72), 11MV (59), 12MV (74), 13MV (57), and 14MV (44) each requiring different epochs for the training model. The analysis of the training and validation epochs are shown in Figure 6. At this stage, the process involves classifying each object in the point cloud as an individual class. Subsequent refinement groups point with the same label, completing instance segmentation by producing coherent clusters for each object.

Model training was conducted with early stopping, which halted learning after ten consecutive epochs without significant improvement in loss to prevent overfitting. For each combination of bands, the final loss gap recorded at the point at which early stopping was implemented was used as a reference metric for the model's generalisation capacity. The loss gap represents the difference between the average loss on the training and test sets, providing a direct indication of how well the model fits the data without overfitting.

Analysis of the results reveals a clear trend of decreasing loss gap as the number of multispectral bands increases: the loss gap was 0.1651 with 7MV, decreasing to 0.1596 with 8 MV and continuing to decrease to 0.0225 with 14MV. This progressive reduction suggests that incorporating additional spectral bands improves the model's capacity to recognise relevant variations in the data, thereby enhancing its generalisation ability. The sharp decrease observed between 9MV (0.1296) and 10MV (0.0572) suggests that, once a certain threshold of bands is overcome, the model gains access to critical spectral information that enables more stable learning. Furthermore, the proximity of values such as 10MV (0.0572) and 11MV (0.0575) suggests that the inclusion of additional bands around 570 nm may yield only marginal contributions to the model fit. This behaviour is likely attributable to the spectral properties of the analysed targets, which may exhibit limited discriminative variability within this wavelength region. However, the gradual improvement observed up to 14MV demonstrates that each additional band further refines the model's representation of the multispectral spectrum.

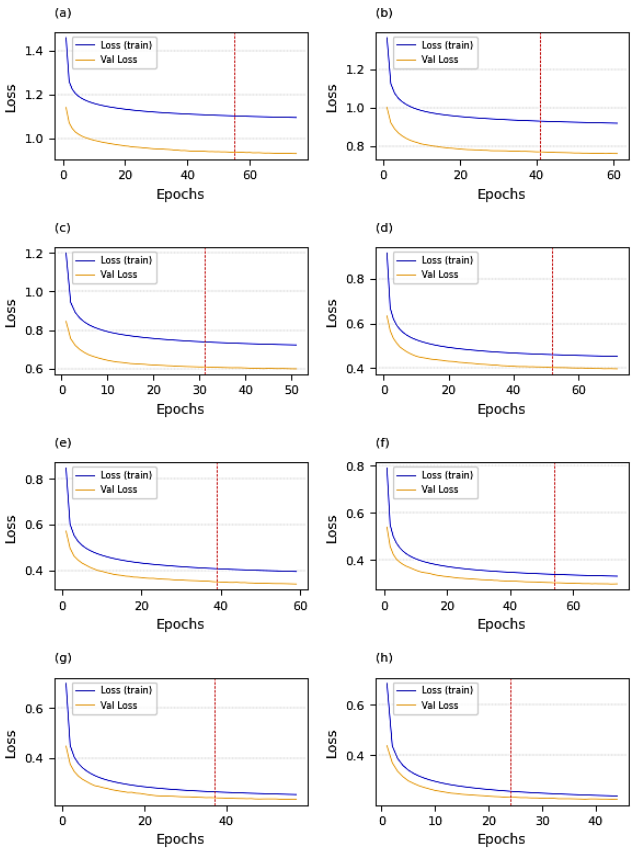

Figure 6 – Training (blue line) and validation loss (orange line) considered by early stop (red line) analysis of datasets: (a) 7MV, (b) 8MV, (c) 9MV, (d) 10MV, (e) 11MV, (f) 12MV, (g) 13MV and (h) 14MV.

4.4 Segments Analysis

Table 1 summarises the performance metrics obtained for each dataset during the evaluation process. These metrics include mIoU and AP_{50}, before and after post-processing (pp). The

results show consistent improvements in segmentation performance as the number of multispectral bands increases. Both the mIoU and the AP_{50} show progressive gains from the 7MV to the 14MV, which highlights the model’s enhanced ability to capture spectral information. Initial performance with 7MV (bands 685, 710, 735, 430, 550, 650 and 850 nm) was relatively slight, with an mIoU of 48.08% and an AP_{50} of 42.89%, suggesting limited overlap and lower detection accuracy. However, with the addition of further bands, performance improves consistently, reaching an mIoU of 63.29% and an AP_{50} of 57.59% with 9MV (630 and 560 nm), reflecting significant enhancements in the segmentation and detection of regions of interest.

Dataset	mIoU	AP_{50}	mIoU (pp)	AP_{50} (pp)
7MV	48.08	42.89	56.29	51.70
8MV	55.41	49.42	64.35	60.36
9MV	63.29	57.59	77.90	75.55
10MV	74.64	69.36	92.74	91.77
11MV	77.14	72.00	93.59	92.53
12MV	80.09	75.45	96.37	95.68
13MV	84.64	79.90	96.52	95.92
14MV	**85.02**	**80.42**	**96.59**	**96.14**

Table 1 – Results analysis of the datasets in percentage.

Post-processing provided additional gains across all band combinations. The most significant improvement occurs when moving from 9MV to 10MV (with the addition of the 450 nm band), with mIoU rising from 77.90% to 92.74 (+14.84% points) and AP_{50} from 75.55% to 91.77% (+16.22% points). This indicates that the 10MV configuration is the minimum required to achieve significantly satisfactory results in instance segmentation. When all 14 bands were combined after post-processing, the model produced the best results, achieving an mIoU of 96.59% and an AP_{50} of 96.14%. These results show smaller, yet continuous, gains compared to previous combinations.

A comparative analysis of intermediate combinations shows that including strategic bands has a greater impact on improving segmentation than adding less informative bands. For instance, increasing from 11MV to 12MV (by adding 405 nm) raises the mIoU by just 2.78% points and the AP_{50} by 3.15% points. This behaviour is likely related to the presence of non-vegetation targets in the dataset, for which the additional band (405 nm) helps distinguish objects lacking chlorophyll. The gains between 12MV, 13MV and 14MV (by adding 490 and 525 nm) are even smaller. This reinforces the idea that a large improvement is achieved by ensuring minimum coverage of critical spectral information, with additional bands mainly serving to refine segmentation.

Figure 7 shows that the class-wise segmentation error analysis revealed a systematic improvement as the number of multispectral bands increased. Higher rates of misclassification were observed in the leaf-related classes (0–24) in the 7 MV to 9MV configurations, particularly in classes 0, 1 and 4–7. From the 10 MV configuration onwards, there was a marked reduction in these errors, and between the 11MV and 14 MV configurations, the leaf classes were segmented consistently, indicating a significant improvement in spectral stability and the model’s discriminative capability. Artificial material classes, such as calibration targets (classes 25, 27, 29, 31, 35, 39, 42 and 43), exhibited a similar trend. and no errors observed beyond 10MV. This suggests that including the 685, 710 and 735 nm bands was key to enhancing the distinction between the spectra of natural and artificial surfaces.

No segmentation errors were detected for the sandpaper surfaces (26, 28 and 30) or the EVA material (32–38) across any configuration, indicating strong spectral separability and classification consistency regardless of band quantity. In contrast, classes associated with tree bark (40, 41 and 49) and citrus residues (44, 45, 47 and 48) were more sensitive to spectral variation. Errors were concentrated in the early configurations (7MV–9MV), with complete stability achieved from 13MV onwards.

Overall, the elimination of segmentation errors from 11 MV demonstrates an improvement in the model's generalisation capability. The additional bands, particularly those at 405, 490 and 525 nm, resulted in more robust and consistent segmentation performance across different surface types. These findings emphasise the cumulative benefit of multispectral information in reducing intra-class ambiguity. Moreover, they highlight that beyond 11 MV, the model reaches a level of spectral saturation where further additions yield only marginal improvements.

Across the previous analysis, it was consistently observed that larger segments tend to exhibit better relative classification performance, whereas smaller segments show greater variability and a lower proportion of correctly classified points. To validate this tendency, Figure 8 presents a detailed analysis of the relationship between the number of points per segment and the number of correctly segmented points.

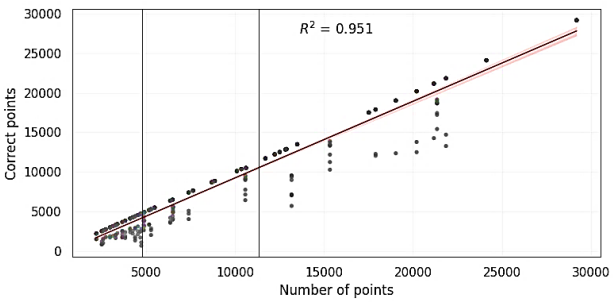

Figure 8 – OLS regression between the number of points per segment and the number of correctly segmented points. Vertical black lines indicate thresholds of the cumulative distribution, separating small, medium, and large segments based on their point counts.

The Ordinary Least Squares (OLS) regression analysis, based on 400 observations, resulted in a coefficient of determination of $R^2 = 0.951$, indicating a very strong linear association between the two variables. The regression slope was estimated as 0.9684 ($p < 0.001$), showing that, on average, each additional point in a segment contributes almost one additional correctly segmented point.

4.5 Discussion

To examine the effect of segment size in more detail, the observations were divided into three categories based on the distribution of points per segment. Thresholds were defined at the 33.3% and 66.7% points of the cumulative distribution: segments

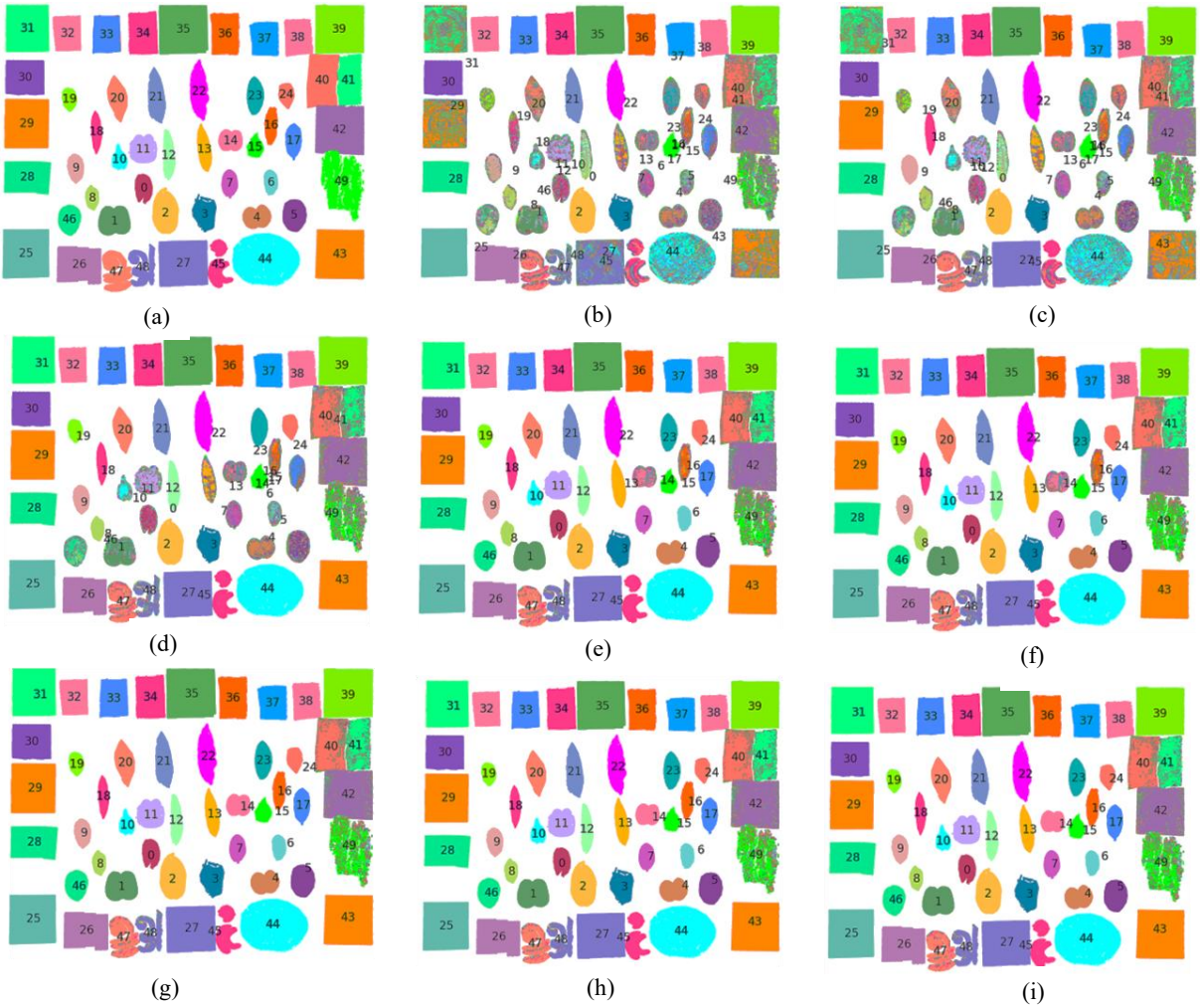

Figure 7 – Manual segmentation of the point cloud performed in (a) and the segmentation results for datasets (b) 7 MV, (c) 8 MV, (d) 9 MV, (e) 10 MV, (f) 11 MV, (g) 12 MV, (h) 13 MV, and (i) 14 MV

with 4787 points or fewer were classified as small, those between 4787 and 11,350 as medium, and segments exceeding 11,350 points as large. Vertical black lines in Figure 8 indicate these thresholds, providing a clear visual separation between the categories.

Regression analysis by segment category revealed that the medium and large segments exhibited the strongest linear associations, with respective R^2 values of 0.832 and 0.836 ($p < 0.001$). These results suggest a nearly one-to-one relationship between the number of points and the number of correct points, indicating that increases in point density within these categories lead to proportional gains in accuracy. By contrast, small segments showed a less pronounced relationship (slope = 0.6525; $R^2 = 0.267$; $p < 0.001$), suggesting their contribution is more limited and susceptible to variability. Overall, the high coefficients of determination for the medium and large segments confirm that these categories predominantly shape the general regression tendency, likely due to the greater number of points per object, which improves model learning and reduces uncertainty.

5. Conclusions

This work presents a feasibility analysis of instance segmentation using multispectral point clouds. It demonstrates that, when selected appropriately, spectral bands can produce highly effective segmentation results. Our results show that a minimum of 10 bands is required to achieve satisfactory results, in this study case, and that post-processing can further enhance performance. The substantial improvement from 9MV to 10MV emphasises the significance of capturing essential spectral information, facilitating stable and precise segmentation across various surfaces and notably enhancing the identification of leaf-related and calibration-target classes.

A progressive inclusion of additional bands up to the full 14MV combination further improved the quality of segmentation, achieving the best results with an mIoU of 96.59% and an $\boldsymbol{AP_{50}}$ of 96.14%. While gains beyond 10MV were smaller, they were important for fine-tuning more challenging features. Leaf classes (0–24) and tree bark classes (40, 41 and 49) exhibited the highest misclassification rates in fewer-band configurations. In contrast, sandpaper surfaces (26, 28 and 30) and EVA materials (32–38) were consistently segmented without error. This shows that features with distinct spectral signatures can be reliably segmented and that additional bands help to resolve subtle or spectrally similar classes.

Finally, the results indicate that, while carefully selected multispectral bands already achieve high segmentation performance, the integration of full XYZ information may further improve feature discrimination and accuracy. Multispectral point

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-279-2026 | © Author(s) 2026. CC BY 4.0 License.

clouds thus provide a robust foundation for precise instance segmentation in complex 3D environments, while practical deployment in robotic platforms highlights the importance of real-time processing. Future work will investigate field-based multispectral colourisation, alternative network architectures, refinements to graph-based post-processing, and the contribution of geometric information.

References

Abbasi, R.; Martinez, P.; Ahmad, R. 2022. The digitization of agricultural industry – a systematic literature review on agriculture 4.0. *Smart Agricultural Technology* 2: 100042.

Agrowing Development Team. 2025. (https://agrowing.com/).

Aksu, G.; Güzeller, C.O.; Eser, M.T. 2019. The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model. *International Journal of Assessment Tools in Education* 6: 170–192.

Arrizza, S.; Marras, S.; Ferrara, R.; Pellizzaro, G. 2024. Terrestrial Laser Scanning (TLS) for tree structure studies: a review of methods for wood-leaf classifications from 3D point clouds. *Remote Sensing Applications: Society and Environment* 36: 101364.

Chanchí Golondrino, G.E.; Ospina Alarcón, M.A.; Saba, M. 2023. Vegetation Identification in Hyperspectral Images Using Distance/Correlation Metrics. *Atmosphere* 14: 1148.

Charisis, C.; Argyropoulos, D. 2024. Deep learning-based instance segmentation architectures in agriculture: A review of the scopes and challenges. *Smart Agricultural Technology* 8: 100448.

Coll-Ribes, G.; Torres-Rodríguez, I.J.; Grau, A.; Guerra, E.; Sanfeliu, A. 2023. Accurate detection and depth estimation of table grapes and peduncles for robot harvesting, combining monocular depth estimation and CNN methods. *Computers and Electronics in Agriculture* 215: 108362.

Cui, G.; He, Q.; Xia, X.; Chen, F.; Gu, T.; Jin, H.; et al. 2023. Demand Response in NOMA-based Mobile Edge Computing: A Two-phase Game-theoretical Approach.

Feldbauer, R.; Rattei, T.; Flexer, A. 2020. scikit-hubness: Hubness Reduction and Approximate Neighbor Search. *Journal of Open Source Software* 5: 1957.

Gu, W.; Bai, S.; Kong, L. 2022. A review on 2D instance segmentation based on deep neural networks. *Image and Vision Computing* 120: 104401.

Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. 2018. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval* 7: 87–93.

Hagberg, A.A.; Schult, D.A.; Swart, P.J. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX.: 11–15.

Handique, B.K.; Khan, A.Q.; Goswami, C.; Prashnani, M.; Gupta, C.; Raju, P.L.N. 2017. Crop discrimination using multispectral sensor onboard unmanned aerial vehicle. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 87: 713–719.

He, H.; Garcia, E.A. 2009. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering* 21: 1263–1284.

Heydarian, M.; Doyle, T.E.; Samavi, R. 2022. MLCM: Multi-Label Confusion Matrix. *IEEE Access* 10: 19083–19095.

Ioffe, S.; Szegedy, C. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.

Jia, S.; Liu, C.; Yue, H.; Huan, W.; Zhou, X.; Qi, Y. 2025. Semantic-instance-relationship understanding of urban environments using multisource aggregated point clouds. *Building and Environment* 281: 113226.

Kang, H.; Chen, C. 2020. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. *Computers and Electronics in Agriculture* 171: 105302.

Kang, H.; Zhou, H.; Wang, X.; Chen, C. 2020. Real-Time Fruit Recognition and Grasping Estimation for Robotic Apple Harvesting. *Sensors* 20: 5670.

Kashongwe, O.; Kabelitz, T.; Ammon, C.; Minogue, L.; Doherr, M.; Silva Boloña, P.; et al. 2024. Influence of Preprocessing Methods of Automated Milking Systems Data on Prediction of Mastitis with Machine Learning Models. *AgriEngineering* 6: 3427–3442.

Katz, S.; Tal, A.; Basri, R. 2007. Direct visibility of point sets. *ACM SIGGRAPH 2007 papers* 26: 11.

Kingma, D.P.; Ba, J. 2017. Adam: A Method for Stochastic Optimization.

Mahlayeye, M.; Darvishzadeh, R.; Jepkosgei, C.; Mlawa, K.A.; Nelson, A. 2024. DESIS Hyperspectral Satellite Data for Cropping Pattern Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 17917–17929.

Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.

Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. 2017b. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.

Qi, C.R.; Liu, W.; Wu, C.; Su, H.; Guibas, L.J. 2018. Frustum PointNets for 3D Object Detection from RGB-D Data.

Reji, J.; Nidamanuri, R.R. 2023. Deep learning based fusion of LiDAR point cloud and multispectral imagery for crop classification sensitive to nitrogen level. *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*: 1–4.

Schor, N.; Berman, S.; Dombrovsky, A.; Elad, Y.; Ignat, T.; Bechar, A. 2017. Development of a robotic detection system for greenhouse pepper plant diseases. *Precision Agriculture* 18: 394–409.

Teixeira, I.; Morais, R.; Sousa, J.J.; Cunha, A. 2023. Deep learning models for the classification of crops in aerial imagery: a review. *Agriculture* 13: 965.

Verdouw, C.; Tekinerdogan, B.; Beulens, A.; Wolfert, S. 2021. Digital twins in smart farming. *Agricultural Systems* 189: 103046.

Wang, X.; Kang, H.; Zhou, H.; Au, W.; Chen, C. 2022. Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. *Computers and Electronics in Agriculture* 193: 106716.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. 2019. Dynamic Graph CNN for Learning on Point Clouds.

Xie, P.; Ma, Z.; Du, R.; Yang, X.; Jiang, Y.; Cen, H. 2024. An unmanned ground vehicle phenotyping-based method to generate three-dimensional multispectral point clouds for deciphering spatial heterogeneity in plant traits. *Molecular Plant* 17: 1624–1638.

Xu, M.; Ding, R.; Zhao, H.; Qi, X. 2021. PAConv: Position Adaptive Convolution with Dynamic Kernel Assembling on Point Clouds.

Zhao, Z.-Q.; Zheng, P.; Xu, S.; Wu, X. 2019. Object Detection with Deep Learning: A Review.