

Enhanced Architecture for Non-destructive Leaf Area Estimation Based on a Semantic Segmentation Network

Caio Seixas Duarte, Luiz Maurilio da Silva Maciel, Saulo Moraes Villela, and Marcelo Bernardes Vieira(✉)

Department of Computer Science, Federal University of Juiz de Fora, Juiz de Fora, Brazil
caioseixas.duarte@estudante.ufjf.br,
{luiz.maciel,saulo.moraes,marcelo.bernardes}@ufjf.br

Abstract. The leaf area is a fundamental parameter for evaluating plant growth, development, and physiological status, playing a central role in agricultural and ecophysiological studies. Conventional measurement methods based on physical contact are often destructive, time-consuming, and inefficient, motivating non-destructive approaches based on computer vision and deep learning. However, the high computational cost and architectural limitations of convolutional-based models still represent significant challenges. In this study, architectural improvements are proposed for a deep convolutional neural network derived from an adapted version of DeepLabv3+, aiming to enhance non-destructive leaf area estimation in common bean (Phaseolus vulgaris) using a public dataset. The proposed architecture employs an encoder and two decoders for semantic segmentation and pixel-wise area estimation. The main contributions include the design of a dual-decoder DeepLabv3+-based architecture with a dedicated area estimation branch enhanced by CBAM and C3K2 blocks, along with an analysis through cross-validation and ablation studies to evaluate the impact of architectural design choices. Structural improvements were applied to the area decoder to enhance feature representation. The model was evaluated using the Relative Error Rate (RER) and compared with a reference architecture. Results show a consistent reduction in RER for leaf area estimation. Ablation studies indicate that increasing architectural complexity does not necessarily lead to performance gains. Overall, the proposed improvements increase estimation accuracy, highlighting the potential of efficient deep learning models for precision agriculture.

Keywords: leaf area estimation · deep learning · semantic segmentation

1 Introduction

The leaf is the plant organ responsible for photosynthesis and plays an essential role from both physiological and ecophysiological points of view [1]. Several

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 580–596, 2027.
https://doi.org/10.1007/978-3-032-30491-9_36

parameters associated with this organ are directly linked to plant growth and development, including size, shape, thickness, area, and leaf amount [2]. Among these attributes, leaf area stands out as one of the most relevant indicators for assessing plant growth, productivity, and physiological status [3].

The measurement of leaf area can be performed using contact and non-contact methods. Contact-based methods typically require physical manipulation of plant material, which may damage plant development, demand longer execution times, and are often destructive [4,5]. In contrast, non-contact methods are recognized for their higher speed, while also enabling reductions in resource waste and physical damage during analysis [5,6].

In this context, image-based techniques have been increasingly adopted as non-contact measurement approaches, as they allow the extraction of relevant information without interfering with plant growth and development [5]. Some studies estimate the leaf area using pixel-based methods [7], whereas others exploit images acquired by unmanned aerial vehicles to extract the leaf area index of plant canopies [8].

With the advancement of computer vision, a wide range of methods have been employed for image analysis and processing, enabling greater accuracy in tasks such as semantic segmentation [9]. In this context, Convolutional Neural Networks have stood out in performing these tasks due to their superior learning capability and performance, which is attributed to their hierarchical architecture [10]. However, there are still several relevant challenges, particularly those related to feature refinement and feature representation [11,12]. Therefore, attention mechanisms such as CBAM and convolutional structures such as C3K2 have demonstrated the ability to improve the model's ability to discriminate important features and information and improve the capture of fine details, respectively [12,13].

In this scenario, the present work proposes architectural enhancements to a deep convolutional neural network derived from a previously adapted version of DeepLabv3+ [14], aiming to improve the non-destructive estimation of the leaf area of common beans using a public dataset [15]. To this end, our main contributions comprise new structural modifications introduced into the DeepLabv3+ based architecture, including the incorporation of novel blocks such as CBAM and C3K2 to analyze their impact on model performance, and a five-fold cross-validation for a more comprehensive analysis of the results, along with an ablation study varying parameters and the internal structure of the incorporated blocks. Furthermore, we compared the performance of the proposed model with the results obtained from the reference architecture and performed an analysis of predictions by image.

2 Background

In recent years, several studies have been developed that focus on estimating leaf area through computational models [16]. This interest arises from the relevance of this metric for monitoring plant physiological parameters, such as photosynthetic

capacity, respiration, and evapotranspiration [17], as well as its importance for tracking plant growth [18] and for prediction of agricultural yield [19].

The methods used for measurement of the leaf area can be classified as direct (contact-based) and indirect (non-contact) approaches [20]. Direct methods rely on physical measurements made in the field, requiring substantial time and labor, in addition to potentially causing damage to plant development. In contrast, indirect methods are known to offer greater speed, accuracy, and flexibility, while also reducing damage to analyzed resources [4–6].

Among the tools used to implement indirect methods, Computer Vision techniques, particularly those based on semantic segmentation, have demonstrated promising results in this context [21]. A relevant example is the work proposed by Silva et al. [14], which presented an adaptation of the DeepLabv3+ network to estimate the are of bean leaves using images containing a leaf and a fiducial marker, achieving satisfactory results.

Despite its strong performance in semantic segmentation tasks, the DeepLabv3+ architecture presents important limitations, such as the high computational cost resulting from its complex feature extractor and the large number of parameters [22,23]. Furthermore, a progressive reduction of spatial resolution in the encoder can lead to a loss of information and fine details in the decoder [22].

In light of these challenges, the search for architectural adaptations that incorporate more efficient mechanisms with lower computational costs has become increasingly common [22,24]. In this context, modules such as the Convolutional Block Attention Module (CBAM) and efficient convolutional blocks, such as C3K2, have been employed to enhance performance in various Computer Vision tasks, including semantic segmentation, by promoting the prioritization of visually relevant regions and the efficient refinement of extracted features [13,22].

The C3K2 block was introduced in YOLOv11, an architecture designed for object detection and instance segmentation, with the objective of improving feature extraction while maintaining low structural complexity. Thus, it seeks to balance the representational capacity of the model without incurring a high computational cost [25]. Additionally, this block contributes to more effective feature refinement, resulting in more discriminative visual representations [13].

In terms of its internal structure, the C3K2 module extends the conventional C3 structure by incorporating a more flexible and efficient feature processing strategy. Instead of employing large convolutional kernels, it leverages parallel convolutional paths that approximate a wider receptive field while keeping the computational cost low. This module begins with a convolution that expands the input feature channels, followed by a split operation that separates the feature map into two distinct paths. One branch bypasses most of the transformations, preserving original information, while the other is processed through a sequence of residual bottleneck units, which may incorporate the C3K structure based on the iterative configuration. Within this processing path, parallel convolutional operations are employed to approximate a larger receptive field, and their out-

puts are subsequently merged to enrich the feature representation. After these transformations, the different branches are concatenated and passed through a final convolutional layer to fuse the aggregated information and produce the output feature map. This design, combining some operations with residual processing, enables efficient feature extraction and representation while maintaining a reduced computational overhead [26].

In turn, CBAM is a lightweight attention mechanism proposed to enhance the capability of convolutional neural networks with reduced computational overhead [27]. This module sequentially applies channel and spatial attention mechanisms, allowing the network to emphasize relevant features while suppressing less informative ones [22]. By operating directly on intermediate feature maps, CBAM performs adaptive feature refinement and can be easily integrated into different convolutional architectures without requiring significant structural modifications [22,27].

Based on these considerations, this work proposes the adaptation of a DeepLabv3+ based deep neural network through modifications to the decoder responsible for estimating leaf area and the fiducial marker. These modifications aim to reduce information loss and preserve fine details, thereby increasing the efficiency of the proposed network compared to the reference model.

3 Proposed Method

For this work, a neural network based on DeepLabv3+ was employed, implemented in TensorFlow 1.15.0. The architecture was modified from the reference model proposed by Silva et al. [14] to improve the prediction of the pixels areas corresponding to leaves. Figure 1 presents an overview of the proposed architecture, highlighting the new added blocks.

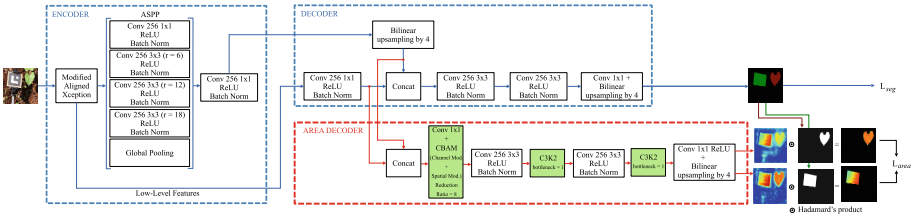


Fig. 1. Proposed model architecture. The blocks highlighted in green indicate the Convolution 1×1, CBAM and C3K2 introduced to the Area Decoder in the architecture proposed by Silva et al. [14]. (Color figure online)

3.1 Adapted Deeplabv3+

For leaf area estimation, we used a network based on the reference model. This network simultaneously performs semantic segmentation and estimation of the

per-pixel area of objects present in the input image, as illustrated in Fig. 2. The architecture is composed of an encoder and two distinct decoders, with the first responsible for semantic segmentation and the second responsible for per-pixel area regression.

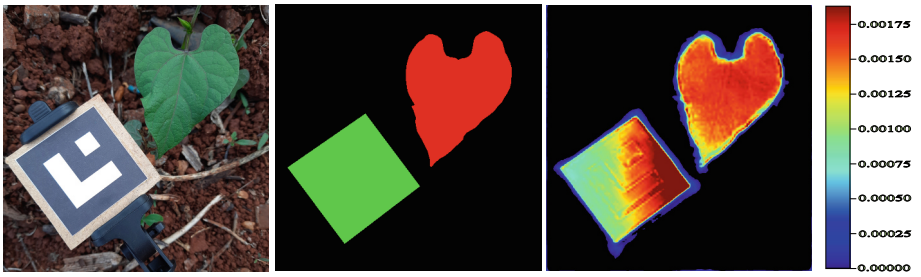


Fig. 2. From the left to right, the input sample used, the segmentation output and the area output.

The encoder consists of a convolutional backbone network, followed by an Atrous Spatial Pyramid Pooling (ASPP) module, formed by parallel separable convolutions and finalized by a 1×1 convolution with 256 channels. As in the reference work [14], we adopted the Aligned Xception as the backbone. The ASPP module is used to capture contextual information at multiple scales, while the 1×1 convolution acts to fuse information across channels, contributing to the preservation of the spatial resolution of the feature maps.

From the encoder, the feature maps are directed to the original DeepLabv3+ decoder, which is responsible for semantic segmentation. At this stage, low-level feature maps extracted directly from the backbone and processed by a 1×1 convolution are concatenated with the maps originating from the ASPP, which were previously resized through bilinear upsampling by a factor of 4. Subsequently, two separable 3×3 convolutions are applied to refine the features, and a 1×1 convolution, followed by another bilinear upsampling with the same factor, finalizes the process of generating the segmentation masks and logits.

In parallel with the segmentation decoder, the architecture incorporates a decoder dedicated to area estimation. Similarly to the original DeepLabv3+ decoder, this structure performs the concatenation of low-level features with the maps produced by the ASPP, followed by the application of two 3×3 separable convolutions and a 1×1 convolution, accompanied by bilinear upsampling. However, this decoder differs by employing the ReLU activation function, ensuring the generation of positive area values, as well as by producing maps that represent the per-pixel estimated area of the objects [14]. Finally, a Hadamard product is performed between the segmentation masks provided by the segmentation decoder and the per-pixel area maps, enabling the separate estimation of the areas of the leaf and the fiducial marker, culminating in the generation of area logits.

3.2 Modified Decoder for Leaf Area Estimation

Based on the reference architecture, with the objective of improving the estimation of leaf area on a per-pixel basis, we implemented relevant modifications in the area decoder architecture. Starting from the premise that the depth of the network can be explored, blocks such as C3K2 and CBAM were added. These modifications are highlighted in green in Fig. 1.

We added a 1×1 convolution to the area decoder after the concatenation stage, with the purpose of performing channel fusion between low-level feature maps and those originating from the ASPP module. Following this, we added a CBAM block that is responsible for applying channel and spatial attention. This strategy aims to enable the network to emphasize the features and regions of the image that are relevant for the area estimation. Additionally, a C3K2 block was inserted after each 3×3 separable convolution, which improves the refinement of the extracted features.

The added blocks have internally adjustable parameters. The CBAM's channel attention module has a parameter called the reduction rate, which regulates the number of input tensor channels. The C3K2 block, in turn, allows for variation in the number of internal modules, which are named bottlenecks. We performed an ablation study of these parameters, described in Sect. 4.2.

The loss function used in the proposed model was maintained as defined by Silva et al. [14], which is composed of the sum of the semantic segmentation loss and the area estimation loss.

Furthermore, an ablation study of the proposed architecture was conducted, in which different configurations of the added blocks were evaluated. This study included variations of the reduction ratio in the CBAM, the number of bottlenecks in the C3K2, as well as the position of internal modules within the CBAM. These experiments were performed using fold 1 of the first cross-validation of the dataset. Additionally, complete experiments were conducted considering all five folds of the same cross-validation. In all cases, the results were compared with those obtained using the network proposed by Silva et al. [14].

4 Experiments and Results

To evaluate our method, we performed a comparative analysis with the reference network. We preserved the main training hyperparameters, including *atrous* rates, weight decay, learning policy, learning power, optimizer, and momentum to ensure a fair comparison between the approaches, as detailed on Table 1. Likewise, the weight initialization and pre-training strategies were also maintained. As performed on the network of baseline, the encoder and decoder networks were initialized with weights pre-trained on the PASCAL VOC 2012 dataset [28], with the exception of the logits layer. Separable convolution layers were initialized from a truncated normal distribution (mean = 0.0, standard deviation = 0.33). For the remaining convolutional layers, Xavier initialization [29] was adopted, drawing weights from a uniform distribution. Only the hyperparameters learning rate and batch size differed from those defined in the reference work. The batch

size was set to 8. We conducted the experiments using one NVIDIA A30 GPU with 24 GB of VRAM.

Table 1. Main hyperparameters used on the present work.

Parameter	Values
Atrous rates	[6, 12, 18]
Weight decay	0.00004
Learning policy	Polynomial
Learning power	0.9
Optimizer	Momentum
Momentum (m)	0.9

For model training, a maximum number of 150,000 steps was defined. Performance evaluation was carried out using 51 checkpoints saved at regular intervals of 100 steps, ranging from step 145,000 to step 150,000. After training the models, the checkpoints saved during this stage were evaluated in the training set, and based on these results, the best step was selected. Subsequently, during the testing phase, this selected model was evaluated in the testing set. As a performance metric, the Relative Error Rate (RER) was adopted, as defined in Eq. 1, which corresponds to the percentage ratio between the absolute difference between the estimated area (a_{est}) and the real area (a_{real}) relative to the real area [14]:

$$\text{RER}(\%) = \left| \frac{a_{\text{est}} - a_{\text{real}}}{a_{\text{real}}} \right| \cdot 100 \tag{1}$$

The selection of the best model considers the sum of the mean and standard deviation of the RER for both the leaf and the fiducial marker. Thus, the selected model was the one that presented the lowest value of $\mu_l + \sigma_l + \mu_m + \sigma_m$.

4.1 Dataset

The dataset used in this study consists of images of common bean leaves in the vegetative stage, captured together with a custom 5 × 5 cm fiducial marker from the ArUco library [30]. The dataset contains 6,981 images of 512 × 512 pixels from 612 distinct leaves [15]. The dataset is organized into a 5-5-fold distribution. Each leaf appears exclusively in the training or test set, preventing any overlap within the folds. We used the first cross-validation for our experiments. For some ablation tests, we employed only the first split. For our final configuration, all five splits were evaluated, enabling a complete analysis of the dataset. Since Silva et al. [14] used only a partial version of the dataset, our "baseline" model is an evaluation of the reference model on the complete dataset.

Table 2. Ablation study varying the CBAM reduction ratio parameter. The values of Avg RER and Std RER are shown in percentage.

Description	Train				Test			
	Leaf		Marker		Leaf		Marker	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std
Baseline	1.582	0.866	1.480	0.809	6.310	5.267	1.708	1.290
Proposed (Reduction Ratio = 2)	**1.016**	0.869	0.655	0.362	6.119	4.961	1.318	1.046
Proposed (Reduction Ratio = 8)	1.084	0.766	**0.321**	**0.226**	**5.989**	**4.937**	**1.147**	**0.949**
Proposed (Reduction Ratio = 16)	1.108	**0.709**	0.508	0.307	6.543	5.411	1.246	1.009

4.2 Ablation Study

To evaluate the individual contribution of the mechanisms incorporated into the proposed architecture, an ablation study was conducted based on variations in parameters and the internal arrangement of the added structures, while keeping the positioning of these structures within the architecture unchanged. In this way, different variations of the model were analyzed to investigate the influence of the CBAM and C3K2 modules on the performance of the leaf area estimation task. For the ablation experiments, we utilized only the first fold of the first cross-validation of the dataset.

Initially, we evaluated variations of the reduction ratio parameter of the CBAM channel attention module to analyze the impact of this module on network performance. For this purpose, we considered the values 2, 8, and 16, keeping the sequential arrangement of the channel and spatial attention modules in the CBAM, as well as the number of bottlenecks of the C3K2 set to 1. Table 2 presents the corresponding results.

Subsequently, we analyzed the effects of varying the number of bottlenecks in the C3K2 module, to evaluate the influence of this parameter on the refinement of features in the estimation of leaf and marker areas. For this experiment, we adopted the number of bottlenecks 1, 2, and 6, keeping the reduction ratio of the CBAM fixed at 8 and maintaining the same structural arrangement of this attention module adopted in the previous experiment. Table 3 shows the results.

Finally, we performed experiments by varying the internal arrangement of the CBAM modules in a manner analogous to the procedure described in the original work [12]. We evaluated sequential configurations of the channel and spatial attention modules in both orders, as well as a parallel configuration. In these experiments, we kept 1 bottleneck for C3K2 and set the reduction ratio in the attention module equal to 8. Table 4 displays the results achieved.

4.3 Results with Five-Fold Cross-Validation

To validate the results achieved in the ablation study and to verify the absence of bias associated with the use of a single cross-validation fold, we conducted

Table 3. Ablation study varying the number of bottlenecks of C3K2. The values of Avg RER and Std RER are shown in percentage.

Description	Train				Test			
	Leaf		Marker		Leaf		Marker	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std
Baseline	1.582	0.866	1.480	0.809	6.310	5.267	1.708	1.290
Proposed (Bottleneck = 1)	**1.084**	**0.766**	**0.321**	**0.226**	**5.989**	**4.937**	1.147	0.949
Proposed (Bottleneck = 2)	1.104	0.814	0.470	0.292	6.135	4.959	1.241	1.019
Proposed (Bottleneck = 6)	1.254	0.782	0.626	0.411	6.867	5.731	**1.144**	**0.897**

Table 4. Ablation study varying the arrangement of internal structures of the CBAM. The values of Avg RER and Std RER are shown in percentage.

Description	Train				Test			
	Leaf		Marker		Leaf		Marker	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std
Baseline	1.582	0.866	1.480	0.809	6.310	5.267	1.708	1.290
Channel → Spatial	**1.084**	0.766	**0.321**	**0.226**	**5.989**	**4.937**	1.147	0.949
Spatial → Channel	1.255	0.825	0.496	0.314	6.443	5.412	**1.136**	**0.941**
Channel || Spatial	1.261	**0.760**	0.905	0.574	6.194	5.077	1.403	1.031

experiments using both the baseline network and the model proposed in this work, considering all five folds of the first cross-validation of the dataset. For this evaluation, we computed the RER individually for each test image across all folds, accumulated the obtained values and subsequently divided by the total number of images in the dataset. Thus, the final RER values for the leaf and the marker represent the average error considering all 6,981 images in the dataset, ensuring a global and unbiased evaluation of the model. Table 5 shows the results.

4.4 Prediction by Image

In this section, we present the results of image-level predictions generated by the adapted network proposed in this study, considering the five folds of the first cross-validation procedure.

Figure 3 illustrates the relationship between the predicted leaf areas and the reference areas through scatter plots generated for each fold. In these plots, the estimated values are compared with the ground truth, and a fitted linear regression line is also presented, along with the corresponding coefficient of determination (R^2). Across all folds, the predictions follow a consistent linear trend, with R^2 values of 0.858, 0.811, 0.820, 0.803, and 0.747, indicating strong agreement between the estimated and actual areas in most cases, particularly on fold 1, which achieved the best performance. However, the presence of vertical group-

Table 5. RER per image of dataset for leaf and marker. The values of Avg RER and Std RER are shown in percentage.

Description	Leaf		Marker	
	Avg	Std	Avg	Std
Baseline	6.787	5.547	1.547	1.215
Proposed model	**6.578**	**5.516**	**1.155**	**0.941**

ings of points can be observed, suggesting variability in the predictions for the same leaf under different conditions.

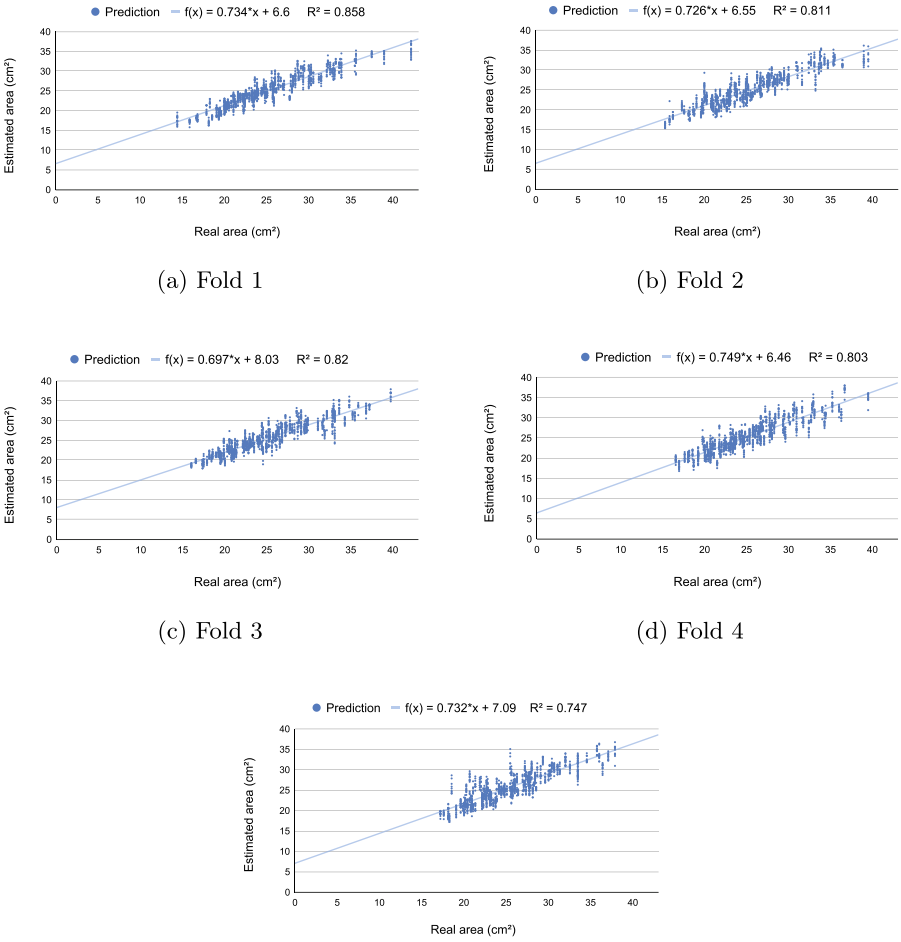
The behavior of the prediction errors is illustrated in Fig. 4, which present the error distributions for each fold. The errors, calculated as the difference between the predicted and reference values, reveal a bias toward underestimation in most folds, with the exception of folds 2 and 4, which present a greater balance between positive and negative prediction errors. In fold 1, the error values for leaf range from -5.50 to $9.50\,\mathrm{cm}^2$, with a higher concentration in the interval between -1.04 and $-0.64\,\mathrm{cm}^2$, corresponding to 139 out of 1387 images. In the remaining folds, an increase in the error range is observed, with intervals of -9.50 to $9.00\,\mathrm{cm}^2$, -7.00 to $9.00\,\mathrm{cm}^2$, -7.50 to $8.00\,\mathrm{cm}^2$, and -10.50 to $8.00\,\mathrm{cm}^2$, respectively, indicating greater dispersion in the achieved errors.

4.5 Qualitative Results

Figure 5 presents qualitative results obtained by the proposed model for representative leaf samples. Each example includes the input image, the segmentation output, the estimated area output, and the corresponding ground truth. These visual results allow a qualitative evaluation of the behavior of the proposed architecture under different leaf and marker perspectives, geometries, and positions.

The qualitative results indicate that the proposed architecture was able to accurately segment both the reference marker and the leaf region under different acquisition conditions, while also producing area estimations visually consistent with the ground truth maps. Among the evaluated samples, Fig. 5(a) achieved the best overall correspondence, with predicted area maps presenting homogeneous distributions and well-defined object boundaries for both the marker and the leaf. Figure 5(b) also presented strong agreement with the ground truth, although with slightly smoother transitions near the object contours. These results suggest that the model preserves geometric consistency even under moderate variations in object orientation and illumination.

In contrast, Fig. 5(c) exhibited larger discrepancies between the estimated area maps and the ground truth, particularly in regions affected by perspective distortion and non-uniform illumination. Although the segmentation masks remained coherent, the estimated area distribution became less homogeneous, especially in the marker regions. Nevertheless, the proposed method preserved the overall object shape and spatial correspondence, indicating robustness and consistency in leaf area estimation.



(a) Fold 1

(b) Fold 2

(c) Fold 3

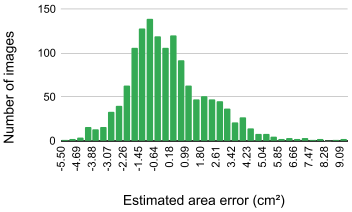
(d) Fold 4

(e) Fold 5

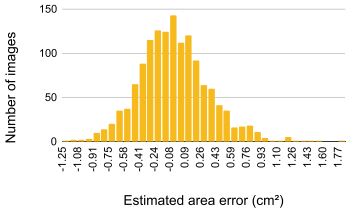
Fig. 3. Scatter plots of the estimated leaf areas for all folds in Cross-Validation 1.

4.6 Discussion

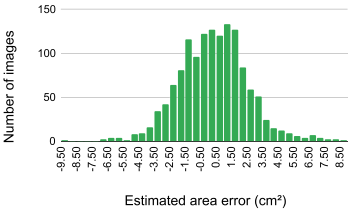
In this subsection, we discuss the contribution and impact of the mechanisms incorporated into the proposed architecture, as well as the influence of their parameters on the task of estimating the leaf area, compared to the baseline network. Based on the results obtained, it is observed that the modifications introduced in the model, in most of the evaluated configurations, consistently contributed to increasing the network accuracy. In particular, considering the results obtained on the leaf test set, the best configuration of the proposed model was composed of the CBAM with a reduction ratio equal to 8 and with a sequential internal arrangement with a channel attention module and a spatial



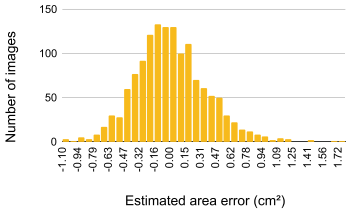
(a) Leaf prediction errors for fold 1.



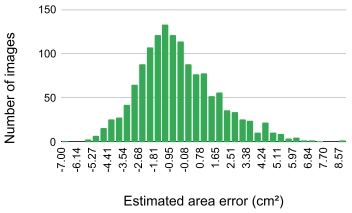
(b) Marker prediction errors for fold 1.



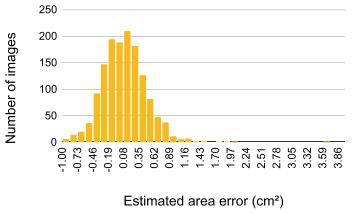
(c) Leaf prediction errors for fold 2.



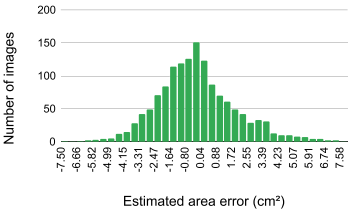
(d) Marker prediction errors for fold 2.



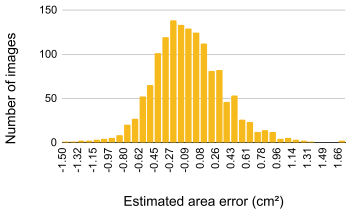
(e) Leaf prediction errors for fold 3.



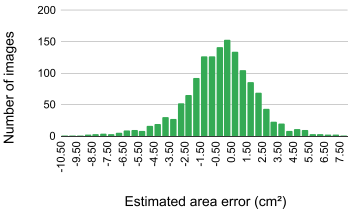
(f) Marker prediction errors for fold 3.



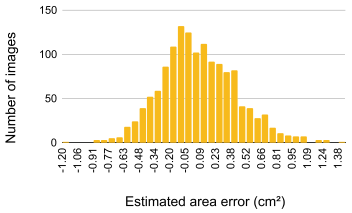
(g) Leaf prediction errors for fold 4.



(h) Marker prediction errors for fold 4.

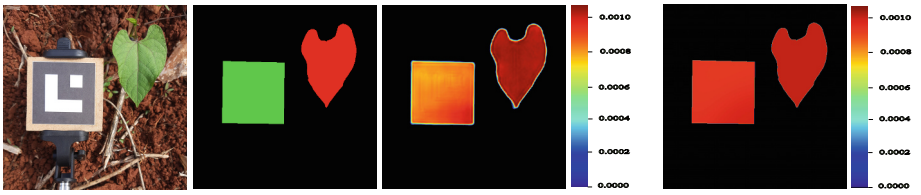


(i) Leaf prediction errors for fold 5.

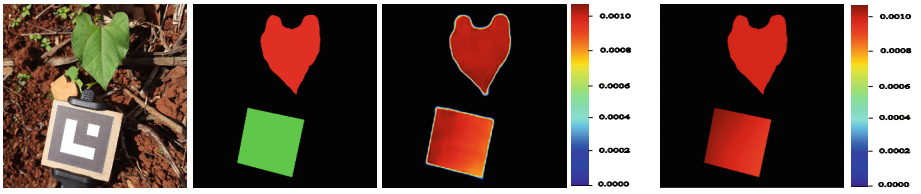


(j) Marker prediction errors for fold 5.

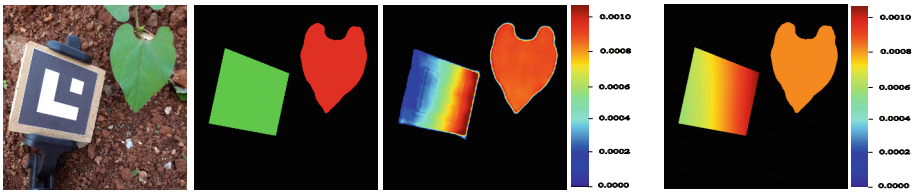
Fig. 4. Histogram of leaf and marker prediction errors achieved by the best model in the test set across all folds.



(a) First example.



(b) Second example.



(c) Third example.

Fig. 5. Qualitative results for some leaf samples. From left to right: input sample, segmentation output, area output and ground truth area image.

module, and with a bottleneck quantity equal to 1 in the C3K2. This architecture presented significant reductions compared to the baseline, both in mean and standard deviation of the RER, in the training and test sets, for the estimation of leaf area, which constitutes the main objective of this work, as well as for fiducial marker area estimation, although, in other configurations implemented in the ablation study, the network presented slightly superior performance in the training for the leaf and in the marker test.

It should be noted that the marker exhibited a more expressive gain compared to the leaf when comparing the results of the proposed model with those obtained using the baseline. This behavior can be attributed to the more uniform and well-defined geometry of the marker (plane), which tends to facilitate the network's learning process. However, since the results were consistent for both objects, it is possible to infer that the proposed model is capable of learning diverse representations, regardless of the geometric complexity of the region of interest.

The analysis of the results presented in Table 3 shows that the C3K2 module achieves better performance when used in a more compact manner, that is, with fewer bottlenecks. This behavior suggests that increasing the depth of the

module does not necessarily lead to improvements in feature refinement for the area estimation task. In contrast, deeper configurations tend to introduce greater complexity into the model, which can hinder the learning process and result in marginal gains or even performance degradation. Therefore, the results indicate that C3K2 operates more effectively when used as a lightweight refinement mechanism, highlighting the importance of balancing representational capacity and computational cost.

In a complementary manner, the results obtained from the CBAM ablation study indicate that the inclusion of this module tends to contribute positively to the performance of the proposed model, especially when smaller values are adopted for the reduction ratio parameter. As shown in Table 2, configurations with reduced values of this parameter exhibited superior performance, possibly preserving a greater amount of information within the channel attention module. This behavior may be related to a more adequate representation of features relevant to the leaf area estimation task, which depends on details associated with texture and edges.

Additionally, the results show that variations in the arrangement of the internal CBAM modules, different from the sequential configuration composed of the channel attention module followed by the spatial attention module, did not result in significant performance improvements. This finding is in agreement with the arrangement originally proposed by Woo et al. [12]. Thus, the results reinforce that the appropriate choice of parameters and the internal organization of the CBAM modules exert a direct influence on the gains provided by the attention mechanisms in the proposed architecture.

The results presented in Table 5 indicate that the proposed model shows consistent performance in estimating the area of both leaf and the fiducial marker. The analysis of the image-level RER across the five folds suggests good generalization capability, without dependence on a specific portion of the data. In addition, the proposed architecture outperformed the reference network presented by Silva [14], highlighting the effectiveness of the architectural modifications adopted.

The image-level prediction analysis also reinforces the consistency of the proposed model across the evaluated folds. The scatter plots demonstrated a tendency of linear relationship between the predicted and reference leaf areas, with satisfactory R^2 values in all folds, indicating that the model presents high correlation between the estimated and real values. In addition, although the error histograms revealed a slight tendency toward underestimation in most folds, the error distributions remained concentrated around small values, suggesting that the proposed architecture maintains consistent predictions even under different image conditions.

It is possible to observe that the obtained results demonstrate the practical relevance of incorporating lightweight attention and refinement mechanisms into architectures dedicated to agricultural computer vision tasks. By improving estimation accuracy with a moderate increase in computational cost, the proposed approach may contribute to the development of more reliable systems for leaf

area estimation. This is particularly important in real-world scenarios, in which a robust area estimation may assist in crop monitoring, plant growth analysis, and precision agriculture applications. Furthermore, the results suggest that the proposed architectural modifications favor the performance and generalization capability of the model.

5 Conclusion

In this work, we introduced CBAM and C3K2 blocks into a semantic segmentation based network for leaf area estimation. The experiments conducted indicate that the integration of these blocks enhances the performance of semantic segmentation networks for leaf area estimation. The results demonstrate that the modified network outperformed the reference model based on DeepLabv3+, achieving higher accuracy in estimating the area of both leaf and fiducial marker. However, it was observed that the gain was more expressive for the marker, which can be attributed to its more uniform and well-defined geometry. These results highlight the potential of deep neural networks for agricultural applications, particularly in the non-destructive analysis of plant morphological features. The incorporation of C3K2 and CBAM modules significantly contributed to both improved performance, an aspect often cited as a challenge in the adoption of deep models in practical scenarios. Despite these promising results, future studies could explore different hyperparameter configurations to further optimize model performance, as well as strategies to reduce the difference in gain between the analyzed objects. Additionally, as a way to address a limitation of this work, it may be interesting to use the proposed model on more than one dataset, aiming to evaluate the generalization potential of the proposed network.

Acknowledgments. The authors would like to thank the Minas Gerais State Research Foundation - FAPEMIG, Brazil. This study was financed in part by the Federal University of Juiz de Fora - UFJF. This work has been supported by UFJF's High-Speed Integrated Research Network (RePesq). https://www.repesq.ufjf.br/.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. White, A.C., Rogers, A., Rees, M., Osborne, C.P.: How can we make plants grow faster? A source-sink perspective on growth rate. J. Exp. Bot. **67**(1), 31–45 (2016). https://doi.org/10.1093/jxb/erv447
2. Huang, W., Ratkowsky, D.A., Hui, C., Wang, P., Su, J., Shi, P.: Leaf fresh weight versus dry weight: which is better for describing the scaling relationship between leaf biomass and leaf area for broad-leaved plants? Forests **10**(3), 256 (2019). https://doi.org/10.3390/f10030256
3. Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G.: An overview of global leaf area index (LAI): methods, products, validation, and applications. Rev. Geophys. **57**(3), 739–799 (2019). https://doi.org/10.1029/2018RG000608

4. Kim, D.C., Anderson, N.O.: Comparative analysis of laboratory freezing methods to establish cold tolerance of detached rhizomes and intact crowns in garden chrysanthemums (dendranthema× grandiflora tzvelv.). Scientia horticulturae **109**(4), 345–352 (2006). https://doi.org/10.1016/j.scienta.2006.05.014
5. Mohammadi, V., Minaei, S., Mahdavian, A.R., Khoshtaghaza, M.H., Gouton, P.: Estimation of leaf area in bell pepper plant using image processing techniques and artificial neural networks. In: 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 173–178. IEEE (2021). https://doi.org/10.1109/ICSIPA52582.2021.9576778
6. Xing, M., et al.: Physiological alterations and nondestructive test methods of crop seed vigor: a comprehensive review. Agriculture **13**(3), 527 (2023). https://doi.org/10.3390/agriculture13030527
7. Doğan, A., et al.: Grapevine leaf area measurements by using pixel values. Comptes rendus de l'Académie bulgare des sciences: sciences mathématiques et naturelles **71** (2018). https://doi.org/10.7546/crabs.2018.06.07
8. Roth, L., Aasen, H., Walter, A., Liebisch, F.: Extracting leaf area index using viewing geometry effects—a new perspective on high-resolution unmanned aerial system photography. ISPRS J. Photogramm. Remote. Sens. **141**, 161–175 (2018). https://doi.org/10.1016/j.isprsjprs.2018.04.012
9. Lei, L., Yang, Q., Yang, L., Shen, T., Wang, R., Fu, C.: Deep learning implementation of image segmentation in agricultural applications: a comprehensive review. Artif. Intell. Rev. **57**(6), 149 (2024). https://doi.org/10.1007/s10462-024-10775-6
10. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: a survey. Comput. Electron. Agric. **147**, 70–90 (2018). https://doi.org/10.1016/j.compag.2018.02.016
11. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 833–851. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49
12. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
13. He, L.h., Zhou, Y.z., Liu, L., Cao, W., Ma, J.h.: Research on object detection and recognition in remote sensing images based on YOLOv11. Sci. Rep. **15**(1), 14032 (2025). https://doi.org/10.1038/s41598-025-96314-x
14. Silva, K.G.F., Silva Maciel, L.M., Villela, S.M., Pedrini, H., Morais, L.E., Vieira, M.B.: Non-destructive leaf area estimation based on a semantic segmentation deep neural network. Comput. Electron. Agric. **237**, 110551 (2025). https://doi.org/10.1016/j.compag.2025.110551
15. Silva, K.G.F., et al.: Bean leaf image dataset annotated with leaf dimensions, segmentation masks, and camera calibration. Data Brief **59**, 111328 (2025). https://doi.org/10.1016/j.dib.2025.111328
16. Mthembu, N., Lottering, R., Kotze, H.: Forest, crop and grassland leaf area index estimation using remote sensing: a review of current research methods, sensors, estimation models and accomplishments. Appl. Sci. **13**(6), 4005 (2023). https://doi.org/10.3390/app13064005
17. Du, L., et al.: Estimating leaf area index of maize using UAV-based digital imagery and machine learning methods. Sci. Rep. **12**(1), 15937 (2022). https://doi.org/10.1038/s41598-022-20299-0

18. Dong, Y., et al.: A method for retrieving coarse-resolution leaf area index for mixed biomes using a mixed-pixel correction factor. IEEE Trans. Geosci. Remote Sens. **61**, 1–17 (2023). https://doi.org/10.1109/TGRS.2023.3235949
19. Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B.: Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. **30**(5) (2003). https://doi.org/10.1029/2002GL016450
20. Zheng, G., Moskal, L.M.: Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. Sensors **9**(4), 2719–2745 (2009). https://doi.org/10.3390/s90402719
21. Luo, Z., Yang, W., Yuan, Y., Gou, R., Li, X.: Semantic segmentation of agricultural images: a survey. Inf. Process. Agric. **11**(2), 172–186 (2024). https://doi.org/10.1016/j.inpa.2023.02.001
22. Fu, H., et al.: DSC-deeplabv3+: a lightweight semantic segmentation model for weed identification in maize fields. Front. Plant Sci. **16**, 1647736 (2025). https://doi.org/10.3389/fpls.2025.1647736
23. Liu, Y., Bai, X., Wang, J., Li, G., Li, J., Lv, Z.: Image semantic segmentation approach based on deeplabv3 plus network with an attention mechanism. Eng. Appl. Artif. Intell. **127**, 107260 (2024). https://doi.org/10.1016/j.engappai.2023.107260
24. Chen, H., Qin, Y., Liu, X., Wang, H., Zhao, J.: An improved deeplabv3+ lightweight network for remote-sensing image semantic segmentation. Complex Intell. Syst. **10**(2), 2839–2849 (2024). https://doi.org/10.1007/s40747-023-01304-z
25. He, L., Zhou, Y., Liu, L., Zhang, Y., Ma, J.: Application of the YOLOv11-seg algorithm for AI-based landslide detection and recognition. Sci. Rep. **15**(1), 12421 (2025). https://doi.org/10.1038/s41598-025-95959-y
26. Rani, M.U., Rajesh, A., Rao, V.V., Prasanna, K.L., Vasanthi, P.: Extreme small object detection with transformer-based local and global context attention. Infrared Phys. Technol. 106251 (2025). https://doi.org/10.1016/j.infrared.2025.106251
27. Liu, L., et al.: A lightweight deeplab v3+ network integrating deep transitive transfer learning and attention mechanism for burned area identification. Sci. Rep. **15**(1), 15969 (2025). https://doi.org/10.1038/s41598-024-66060-7
28. Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The Pascal visual object classes challenge: a retrospective. Int. J. Comput. Vision **111**(1), 98–136 (2014). https://doi.org/10.1007/s11263-014-0733-5
29. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
30. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recogn. **47**(6), 2280–2292 (2014). https://doi.org/10.1016/j.patcog.2014.01.005