



Evaluation of the DeepLab Architecture Family for the Segmentation of Internal Structures in Brachiaria Seeds Using X-ray Images

Livia Kelly G. Maia¹, Emannuel Diego G. de Freitas^{1,2}, Wagner V. Dias¹, Jose Wellington F. da Silva¹, Bruno Riccelli dos S. Silva¹, Haynna F. Abud³, Danielo G. Gomes¹

¹Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza–CE, Brazil.

²Federal Institute of Education, Science and Technology of Ceará, Iguatu Campus, Iguatu–CE, Brazil.

³Image Pesquisas, Av. Humberto Monte, Block 310, Warehouse 17, Fortaleza–CE, Brazil.

**A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th
Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil**

Abstract.

Seed vigor is an essential factor for the uniform emergence of seedlings and for the proper establishment of crops, being directly associated with the productive potential of agricultural crops. Accurate evaluation of this vigor is therefore fundamental for decision-making in seed management and production. Among the methods used for the analysis of physiological quality, the use of X-ray images stands out for allowing the non-destructive visualization of internal morphological structures of seeds, such as the endosperm and the embryo, preserving their physical integrity. However, the manual interpretation of these images is a time-consuming process, subject to variability among evaluators and high operational costs, which has motivated the development of automated approaches based on Computer Vision and Deep Learning. In this context, this work presents a comparative analysis of the performance of the DeepLabV1, DeepLabV2, DeepLabV3, and DeepLabV3+ architectures in the semantic segmentation of internal structures in X-ray images of Brachiaria seeds (Urochloa brizantha cv. Xaraés). The models were trained and evaluated using a set of images previously labeled by specialists, which ensures the reliability of the annotations. Performance evaluation was carried out using the metrics Precision, Recall, Dice, and Intersection over Union (IoU), in addition to inference time analysis, aiming to measure both the accuracy and computational efficiency of the architectures. The results obtained demonstrate high performance for all evaluated networks, with values above 0.96 in all considered metrics, indicating high precision and robustness in the segmentation of internal seed structures. The DeepLabV3+ model presented the best overall results, standing out for its higher accuracy and consistency in segmentation. On the other hand, DeepLabV1 showed competitive performance, combined with lower computational cost and shorter inference time. These findings reinforce the feasibility of using DeepLab architectures in the automated analysis of X-ray images of seeds and indicate that the choice of model should consider the balance between accuracy and efficiency, according to computational constraints and the specific objectives of the application.

Keywords.

Semantic Segmentation; X-ray Imaging; Seed Quality; Deep Learning; DeepLab.

Introduction

Seed vigor determines the potential for rapid and uniform plant emergence under diverse environmental conditions in the field (Rajjou et al., 2012; Finch-Savage, 2010) and is directly associated with greater initial growth and increased productivity in agricultural systems (Han et al., 2014).

In the context of Brazilian agriculture, characterized by wide edaphoclimatic diversity and production systems ranging from smallholder farming to large-scale agribusiness, the use of seeds with high physiological quality is essential. Climatic variability, the expansion of cultivated areas, and the need for high productivity levels make seed vigor a strategic factor in ensuring proper crop establishment and the sustainability of agricultural production in the country.

Conversely, the use of low-vigor seeds in agricultural production results in abnormal seedlings or non-germinated seeds, leading to waste of labor and financial resources (Xing et al., 2023).

Therefore, the physiological quality of seeds—shaped by genetic, physical, sanitary, and physiological factors—plays a fundamental role in crop establishment, directly affecting yield potential across plant species (Oliveira et al., 2022; Campos et al., 2022).

Among the factors influencing this quality, the internal morphological characteristics of seeds, such as endosperm and embryo development, stand out, as they are closely related to physiological potential (Basu et al., 2023; Marchi et al., 2017).

These factors can be observed and measured using X-ray imaging (“Granata, 2025”), a non-destructive analytical method that enables the assessment of several aspects of seed development, establishing a connection between internal morphology and structural integrity (Basu et al., 2023; Zang et al., 2022). However, although effective, these methods present important limitations, including subjectivity, operational cost, and dependence on specialized labor (Zhuang et al.; Zhang et al.; Oquab et al.).

To address this challenge, recent literature has explored various Computer Vision methods and Digital Image Processing techniques to automate seed quality assessment using X-ray images (Pessoa et al., 2023; Medeiros et al., 2021; Hong et al., 2023).

For Computer Vision systems to extract reliable information and perform quantitative analyses from images, proper segmentation of Regions of Interest (ROIs) is essential, as it constitutes one of the most critical stages of these systems; segmentation quality directly influences overall system performance (Xu et al., 2024).

Thus, accurate segmentation is indispensable for correct feature extraction and object classification. With advances in Deep Learning, new automated approaches have been proposed for precise segmentation of internal seed structures (Freitas et al., 2024; da Silva et al., 2024).

Following this line of research, the present study aims to perform a comparative analysis of the performance of different Convolutional Neural Network architectures from the DeepLab family (DeepLabV1, DeepLabV2, DeepLabV3, and DeepLabV3+) applied to the semantic segmentation of internal structures in X-ray images of *Brachiaria* seeds. Specifically, the objective is to segment the endosperm and the area bounded by the seed coat in order to evaluate the ability of these architectures to capture relevant morphological features with high precision. By comparing the

results obtained from the different model versions, the study seeks to identify the most suitable architecture for segmenting these Regions of Interest, thereby advancing automated methodologies for assessing seed vigor and physiological quality.

DeepLab Architecture Family

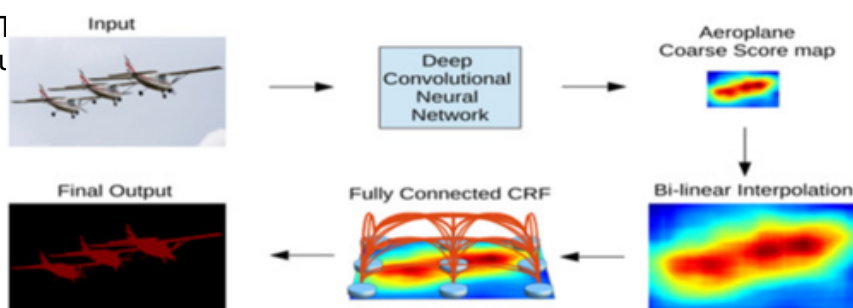
The DeepLab architecture family (Chen et al., 2014) combines Deep Convolutional Neural Networks (DCNNs) with probabilistic graphical models to achieve pixel-level classification (semantic segmentation). Conceptually, the main improvements introduced by the DeepLab architectures lie in the addition of layers to enhance multi-scale feature extraction. The characteristics of the architectures selected for this study are summarized in Table 1.

Table 1. Summary of additions to the analyzed architectures.

Architectures	Main Additions or Changes	Number of Parameters
DeepLabV1	DCNN + CRF	19,447,875
DeepLabV2	ASSF	42,674,244
DeepLabV3	Improved ASPP + CRF removal	58,626,371
DeepLabV3+	Encoder-decoder architecture	45,670,227

DeepLabV1

Complementing 1 graphically in Fig



e presented

DeepLabV2

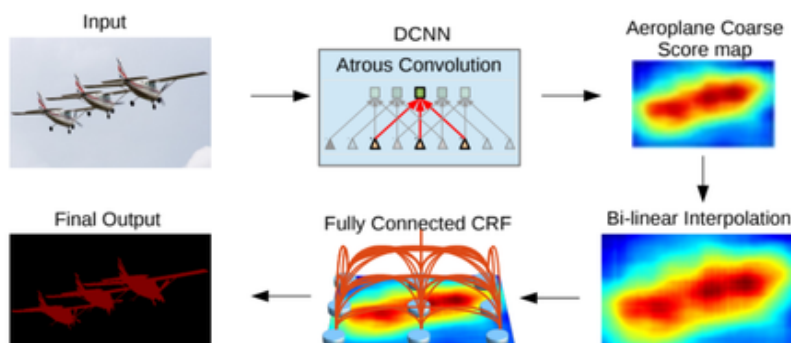


Figure 1. Image of the DeepLabV1 architecture (Chen et al., 2014).

Figure 2. Image of the DeepLabV2 architecture (Chen et al., 2016).

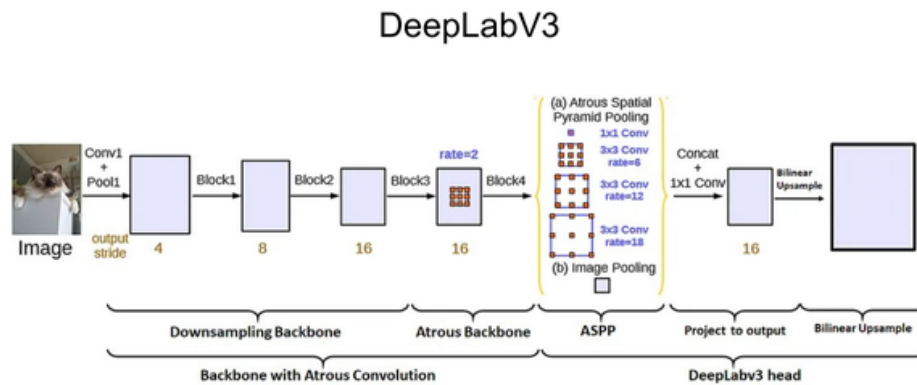


Figure 3. Image of the DeepLabV3 architecture (Chen et al., 2017), adapted by Isaac Berrios (2023).

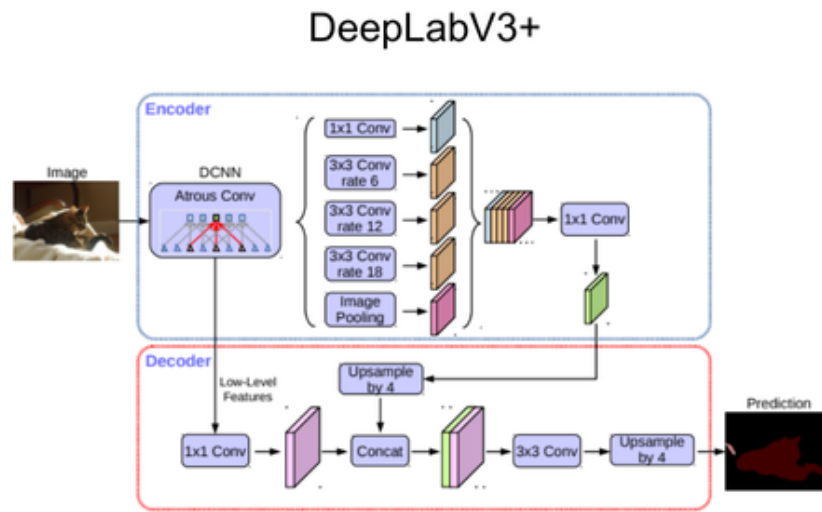


Figure 4. Image of the DeepLabV3+ architecture (Chen et al., 2018).

These architectures employ Transfer Learning, a knowledge reuse technique that leverages information acquired in one domain and transfers it to another, improving model training performance (Zhuang et al., 2020; Zhang et al., 2015; Oquab et al., 2014). In this context, specific components of CNNs such as VGG-16, ResNet-101, and MobileNet, pre-trained on the ImageNet dataset, are incorporated into the architectures.

DeepLabV1 Architecture

DeepLabV1 was the first architecture developed in the DeepLab family. It integrates two techniques into its structure: Deep Convolutional Neural Networks (DCNNs) and fully connected Conditional Random Fields (CRFs). The VGGNet backbone was incorporated into the DCNN framework for feature extraction. A visual representation of the architecture is shown in Figure 1 (Chen et al., 2014).

DeepLabV2

Architecture

DeepLabV2 differs from DeepLabV1 in two main aspects. First, it introduces Atrous Spatial Pyramid Pooling (ASPP) into the convolutional layers to capture multi-scale contextual information. This structure consists of one 1×1 convolution, three 3×3 convolutions, and an image-level feature pooling layer. Second, DeepLabV2 can be built using either the ResNet-101 or VGGNet backbone, as in DeepLabV1; however, the use of ResNet-101 improved final model performance in the original study (Chen et al., 2016). A visual representation is shown in Figure 2.

DeepLabV3

Architecture

DeepLabV3 is structurally similar to DeepLabV2, employing a DCNN followed by ASPP. The main differences include the removal of the CRF module, enabling end-to-end learning, and enhancements to the ASPP module through the incorporation of image-level features to capture long-range contextual information, along with batch normalization parameters to facilitate training (Chen et al., 2017). A visual representation is shown in Figure 3.

DeepLabV3+

Architecture

DeepLabV3+ is the most recent version proposed by the original authors. It incorporates a simple yet effective decoder module to refine segmentation results, particularly along object boundaries. Unlike its predecessors, this architecture follows an encoder-decoder design. The encoder, based on the Xception network, extracts low- and high-level features and forwards the deeper features to the ASPP module, while the decoder combines these features to enhance boundary

precision and pixel-wise semantic classification. Additionally, the resolution of features extracted by the encoder can be adjusted through atrous convolution, allowing a balance between model accuracy and computational efficiency (Chen et al., 2018). A visual representation is shown in Figure 4.

Materials and Methods

This section describes the methodological procedures adopted throughout the study in an integrated and objective manner. It details the dataset construction and preparation process, the computational environment used, the model training strategies, and the performance evaluation criteria. The methodological design was structured to ensure experimental rigor, reproducibility, and consistent analysis of the segmentation models' ability to accurately identify structures of interest in X-ray seed images.

Dataset

A set of 30 X-ray images with a resolution of $2,368 \times 2,440$ pixels, each containing 20 *Brachiaria* seeds (*Urochloa brizantha* cv. Xaraés), composed the dataset used for training and evaluating the DeepLab models. The images were annotated by a specialist using the LabelMe tool, a Python-based framework, and labeled according to the morphological structures of the endosperm and the area bounded by the seed coat, defined as the target classes for prediction.

In addition, a data augmentation technique was applied to increase dataset size and introduce variability, thereby improving model generalization. Transformations included rotations of up to 20° , zoom variations of up to 15%, horizontal and vertical shifts of up to 20% of the image size, and shear transformations of up to 15%. Horizontal flipping was also performed to simulate different object orientations. To prevent empty regions after transformations, missing pixels were filled using the nearest-neighbor method. The final dataset comprised 150 images.

Experimental Design

The study was conducted on the Google Colab platform, a cloud-based computing environment provided by Google for code execution and computational experiments. The infrastructure consisted of a virtual machine equipped with an Intel® Xeon® processor operating at 2.20 GHz, with 12 logical cores (x86_64 architecture) and 52 GB of RAM. Computational acceleration was provided by an NVIDIA T4 GPU with 15 GB of VRAM and CUDA 12.4 compatibility.

Algorithm implementation was carried out using Jupyter notebooks in Python. The machine learning architectures were designed and trained using the PyTorch library, widely recognized in scientific literature for its flexibility, performance, and efficiency in developing Machine Learning and Deep Learning models.

The dataset was initially shuffled using a fixed random seed (42) to ensure reproducibility and experimental control. The samples were then split into three mutually exclusive subsets: 70% for training, 15% for validation, and 15% for testing. The split was performed after global shuffling to ensure random distribution and reduce potential bias from the original data order. Each image-mask pair was allocated entirely to only one subset, with no overlap between training, validation, and test sets. This ensured independence among stages, prevented data leakage, and enhanced the reliability of performance evaluation.

During training, a single set of hyperparameters was adopted for all analyzed models to ensure experimental consistency: 20 training epochs, the AdamW optimizer, a learning rate of 10^{-4} , weight decay of 10^{-4} , and a batch size of 4.

Performance Metrics

For quantitative evaluation, metrics derived from a confusion matrix were used, consisting of four fundamental values: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Precision (Equation 1) quantifies the proportion of pixels correctly identified as belonging to a Region of Interest (ROI) among those predicted as positive:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall (Equation 2) measures the proportion of actual positive pixels correctly identified:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

The next two metrics refer to regions in the image (A and B). The Dice coefficient evaluates the similarity between the region predicted by the model and the reference standard (ground truth).

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

IoU measures the overlap between the region predicted by the model and the ground truth.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

Results and Discussion

The evaluation sought to highlight not only prediction quality but also the balance between accuracy, generalization capability, and efficiency.

Performance curves as a function of loss show a clear trend of stability and convergence throughout training. Training and validation loss curves display decreasing and closely aligned behavior, indicating consistent learning and no significant overfitting. This proximity also reinforces good generalization to unseen data, increasing model reliability. In this context, DeepLabV3 and DeepLabV3+ stand out for presenting particularly stable and well-defined curves, with minimal differences between training and validation. From a segmentation quality perspective, this behavior indicates greater robustness during training, making them especially suitable for the proposed task.

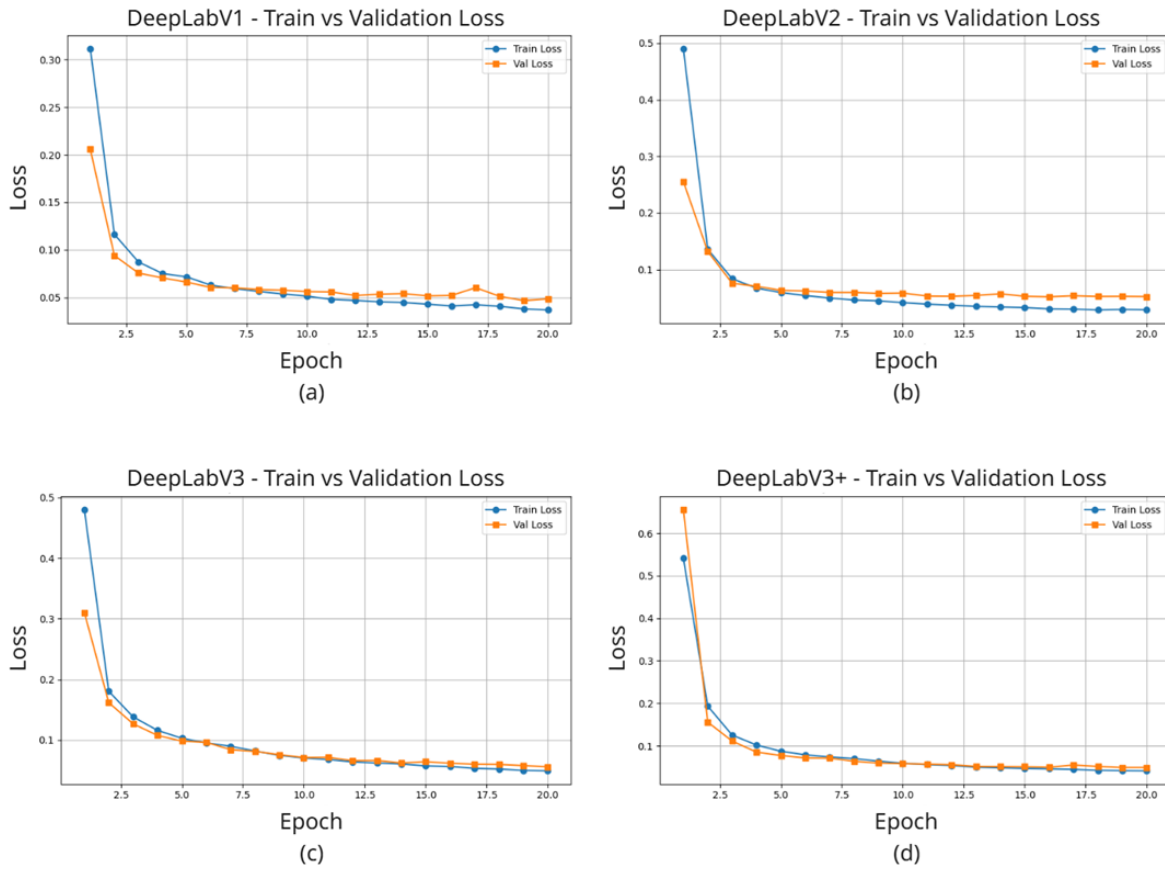


Fig. 5. Loss function graph of the DeepLab models. (a) DeepLabV1 model. (b) DeepLabV2 model. (c) DeepLabV3 model. (d) DeepLabV3+ model.

This analysis of the loss function behavior is corroborated by the quantitative results presented in Figure 2, which illustrates the loss function graph of the models.

Table 2. presents the validation performance metrics for each model.

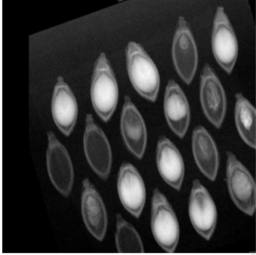
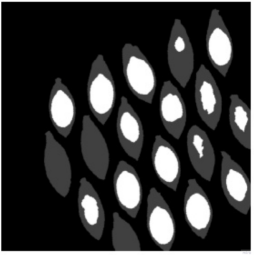
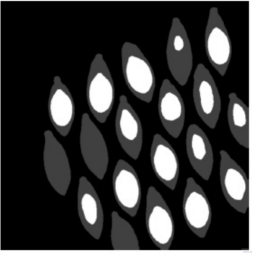
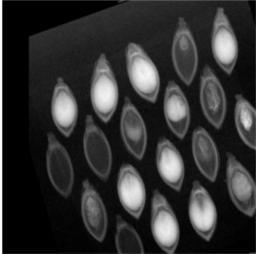
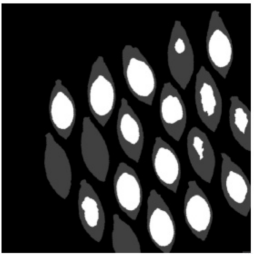
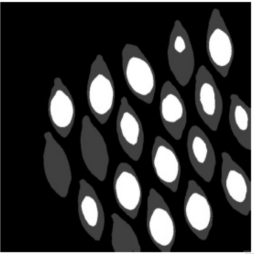
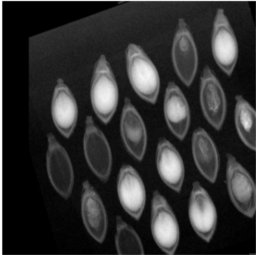
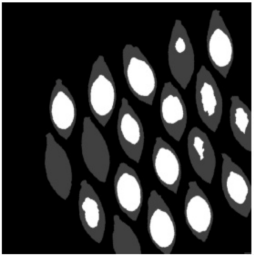
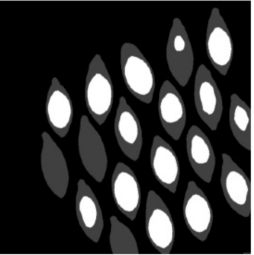
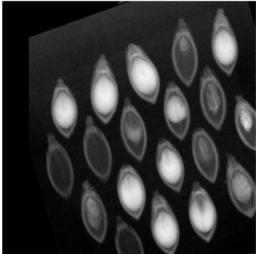
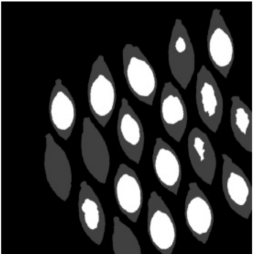
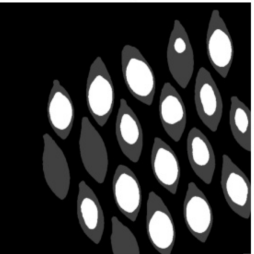
Architectures	Precisão	Recall	Dice	IoU
DeepLabV1	0.9681	0.9707	0.9692	0.9411
DeepLabV2	0.9693	0.9686	0.9690	0.9407
DeepLabV3	0.9672	0.9695	0.9683	0.9395
DeepLabV3+	0.9705	0.9708	0.9706	0.9437

Additionally, when analyzing the performance metrics, the presented results show subtle differences among their values. In general, the four models achieved Precision and Recall values above 0.96, suggesting a good balance between identifying the Regions of Interest and minimizing false positives. Among them, DeepLabV3+ stood out with the best overall results, achieving Precision of 0.9705, Recall of 0.9708, IoU of 0.9437, and Dice of 0.9706, indicating greater overlap between the predictions and the ground truth. However, DeepLabV1, despite being the oldest version, showed competitive performance compared to DeepLabV3+, with an IoU of 0.9411 and a Dice of 0.9692, values very close to those obtained by the more recent versions. This result suggests that, for the dataset used, simpler architectures can still achieve satisfactory performance.

Table 5 presents the inference results obtained for each model evaluated in this study. This presentation allows for a qualitative analysis of model performance by comparing the predicted masks with their respective reference masks, enabling an understanding of model behavior in a

production scenario. In general, a progressive improvement can be observed in the performance of the more recent versions, which produce more precise segmentations with better contour definition. DeepLabV1 shows greater imprecision at the boundaries and small segmentation errors in some objects; DeepLabV2 demonstrates improvement in structural delineation; DeepLabV3 reduces noise and presents greater consistency in segmented areas; while DeepLabV3+ consolidates this progress by producing sharper contours, better spatial refinement, and greater adherence to the reference mask, demonstrating superior qualitative performance and greater robustness for production-context applications.

Table 5. Qualitative results of the DeepLab family models.

Architectures	Image	Ground Truth Mask	Prediction
DeepLabV1			
DeepLabV2			
DeepLabV3			
DeepLabV3+			

Complementing the qualitative analysis, consideration of inference time and number of parameters reveals even more pronounced differences among the architectures. Simpler models, such as DeepLabV1, have lower computational complexity, resulting in reduced inference times and fewer parameters, while maintaining competitive performance. This balance makes them more efficient alternatives for real-time applications or scenarios with hardware constraints. On the other hand, DeepLabV3 and DeepLabV3+ require greater computational cost, reflected in increased inference time and number of parameters, a direct consequence of deeper and structurally more sophisticated architectures.

Table 3. Model inference time.	
Architectures	Time (ms)
DeepLabV1	1.97
DeepLabV2	40.02
DeepLabV3	55.63
DeepLabV3+	52.38

Conclusion

Em síntese, os resultados indicam que o DeepLabV3+ apresenta o melhor desempenho global. In summary, the results indicate that DeepLabV3+ presents the best overall performance in segmentation metrics and is more suitable for scenarios in which accuracy is the primary requirement. In contrast, when the focus is on computational efficiency, DeepLabV1 stands out as an attractive alternative, achieving results close to those of more recent models at significantly lower computational cost.

Thus, the choice of the most appropriate model should consider the balance between performance and available resources, as well as the specific demands of the application. It is concluded that DeepLabV3+ is the most suitable option when the primary objective is to maximize segmentation quality, since it achieved superior performance in the main evaluated metrics. On the other hand, DeepLabV1 represents an efficient and balanced solution, offering competitive results at a significantly lower computational cost.

For future work, the investigation of model optimization and compression techniques, such as pruning and quantization, is recommended in order to reduce inference time without significantly compromising performance. Furthermore, evaluating these architectures on more diverse datasets and in real-world application scenarios may provide a broader analysis of their generalization capacity and practical applicability.

References

- Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., & Job, D. (2012). Seed germination and vigor. *Annual Review of Plant Biology*, 63(1), 507–533. doi:10.1146/annurev-arplant-042811-105550
- Finch-Savage, W. E., Clay, H. A., Lynn, J. R., & Morris, K. (2010). Towards a genetic understanding of seed vigour in small-seeded crops using natural variation in *Brassica oleracea*. *Plant Science: An International Journal of Experimental Plant Biology*, 179(6), 582–589. doi:10.1016/j.plantsci.2010.06.005
- Han, Z., Ku, L., Zhang, Z., Zhang, J., Guo, S., Liu, H., ... Chen, Y. (2014). QTLs for seed vigor-related traits identified in maize seeds germinated under artificial aging conditions. *PloS One*, 9(3), e92535. doi:10.1371/journal.pone.0092535
- Xing, M., Long, Y., Wang, Q., Tian, X., Fan, S., Zhang, C., & Huang, W. (2023). Physiological alterations and nondestructive test methods of crop seed vigor: A comprehensive review. *Agriculture*, 13(3), 527. doi:10.3390/agriculture13030527
- Oliveira, G. R. F. de, Mastrangelo, C. B., Hirai, W. Y., Batista, T. B., Sudki, J. M., Petronilio, A. C. P., ... Amaral da Silva, E. A. (2022). An approach using emerging optical technologies and artificial intelligence brings new markers to evaluate peanut seed quality. *Frontiers in Plant Science*, 13, 849986. doi:10.3389/fpls.2022.849986

- Campos, L. V., Rodrigues, A. A., Sales, J. de F., Rodrigues, D. A., Vasconcelos Filho, S. C., Rodrigues, C. L., ... Rubio Neto, A. (2022). Radiographic imaging as a quality index proxy for *Brachiaria brizantha* seeds. *Plants*, 11(8), 1014. doi:10.3390/plants11081014
- Basu, S., & Groot, S. P. C. (2023). Seed vigour and invigoration. In *Seed Science and Technology* (pp. 67–89). doi:10.1007/978-981-19-5888-5_4
- Marchi, J. L. de, & Gomes Junior, F. G. (2017). Use of image analysis techniques to determine the embryo size of *Senna multijuga* (Rich.) seeds and its relation to germination and vigor. *Journal of Seed Science*, 39(1), 13–19. doi:10.1590/2317-1545v39n1165423
- Granata, A., Pérez, M., Steinbrecher, T., Capozzi, F., Kondrotas, K., Zhang, L., ... Leubner-Metzger, G. (2025). Application of X-ray sorting and priming improves the germination performance of low-quality seed fractions of the *Papaccella* pepper landrace. *Scientia Horticulturae*, 350(114370), 114370. doi:10.1016/j.scienta.2025.114370
- Zhang, S.; Zeng, H.; Ji, W.; Yi, K.; Yang, S.; Mao, P.; Wang, Z.; Yu, H.; Li, M. Non-destructive Testing of Alfalfa Seed Vigor Based on Multispectral Imaging Technology. *Sensors* 2022, 22, 2760.
- FREITAS, Emannuel Diego Gonçalves de; ABUD, Haynna Fernandes; MARTINS NETO, José Luciano; GOMES, Danielo Gonçalves. Segmentação semântica de imagens de raios-X de sementes de *Brachiaria brizantha* utilizando deep learning e transfer learning para avaliação de vigor. *Informativo Abrates*, Londrina, v. 30, n. 1, edição especial, p. 308, 2024. *Anais do 22º Congresso Brasileiro de Sementes*, 10–13 set. 2024, Foz do Iguaçu, PR. ISSN 2319-0728. Disponível em: https://cbsementes.com/files/IA_XXII%20CBSEmentes.pdf.
- Pessoa, H. P., Copati, M. G. F., Azevedo, A. M., Dariva, F. D., Almeida, G. Q. de, & Gomes, C. N. (2023). Combining deep learning and X-ray imaging technology to assess tomato seed quality. *Scientia Agricola*, 80(e20220121). doi:10.1590/1678-992x-2022-0121
- Medeiros, A. D. de, Bernardes, R. C., da Silva, L. J., de Freitas, B. A. L., Dias, D. C. F. dos S., & da Silva, C. B. (2021). Deep learning-based approach using X-ray images for classifying *Crambe abyssinica* seed quality. *Industrial Crops and Products*, 164(113378), 113378. doi:10.1016/j.indcrop.2021.113378
- da Silva, D. d. A., de Freitas, E. D. G., Abud, H. F., & Gomes, D. G. (2024). Applying YOLOv8 and X-ray Morphology Analysis to Assess the Vigor of *Brachiaria brizantha* cv. Xaraés Seeds. *AgriEngineering*, 6(2), 869-880. <https://doi.org/10.3390/agriengineering6020050>
- Hong, S.-J., Park, S., Lee, C.-H., Kim, S., Roh, S.-W., Nurhisna, N. I., & Kim, G. (2023). Application of X-ray imaging and convolutional neural networks in the prediction of tomato seed viability. *IEEE Access: Practical Innovations, Open Solutions*, 11, 38061–38071. doi:10.1109/access.2023.3265998
- Xu, Y. (2024). Application of Image Segmentation Algorithms in Computer Vision. *Frontiers in Computing and Intelligent Systems*, 7(3), 17–20. doi:10.54097/gq1s6737