Evaluation of Transfer Learning in Semantic Segmentation Models for Soybean Seedlings

Emannuel Diego G. de Freitas[1,2], Pedro Henrique dos S. e Silva[1], José Luciano M. Neto[2], Haynna F. Abud[3], Danielo G, Gomes[1]

[1]Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Fortaleza-CE, Brasil.
[2]Instituto Federal de Educação Ciência e Tecnologia do Ceará, campus Iguatu, Iguatu-CE, Brasil.
[3]Image Pesquisas, Av. Humberto Monte, Bloco 310, Galpão 17, Fortaleza-CE, Brasil.

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 evaluation is fundamental in the quality control of commercial lots, as it is directly associated with the rapid and uniform emergence of seedlings and the initial performance of crops in the field. Traditional methods, although widely used, present limitations such as long execution time, dependence on the evaluator's experience, and subjectivity. In this context, systems based on Computer Vision emerge as promising alternatives for automating vigor assessment, as they enable more objective and faster analyses. However, for these systems to extract reliable information from images, precise segmentation of the regions of interest is indispensable. Segmentation separates the relevant anatomical structures, being crucial for feature extraction. This is extremely useful because, from the segmented regions, it is possible to measure the length of the seedling parts and estimate their value. Failures in this step compromise estimates such as the length of seedling parts, making it decisive for the performance of the computer vision system. Although seedling segmentation has been explored, there remains a research gap regarding the systematic evaluation of the impact of different convolutional neural network encoders (backbones) pre-trained, especially considering the asymmetry of segmentation errors for morphological measurement purposes. Given this demand, this work investigates the impact of transfer learning on the semantic segmentation of soybean seedling parts, using the U-Net architecture with systematic replacement of the encoder by 53 backbones pre-trained on ImageNet. A dataset of 80 high-resolution images (4000×3000 pixels) was used, containing 50 seedlings per image, totaling 4,000 cotyledons, 3,700 hypocotyls, and 3,800 radicles manually annotated at the pixel level by a specialist on the Roboflow platform. The results highlighted the differences among the generated models, with better performance associated with deeper models and those with advanced feature extraction strategies. It was also observed that the hypocotyl constitutes the most challenging class. In general, the best models presented Dice coefficients above 90% and IoU values around 83% to 84%, indicating high agreement with the reference annotations. Considering applications aimed at measuring the length of seedling parts, overestimation errors proved to be more critical, with DenseNet-161 being the most suitable model for presenting greater control over false positives.

Keywords
Seed vigor, Computer vision, Semantic segmentation, Soybean seedlings.

Introduction

Seed analysis plays a fundamental role in determining the quality of a seed lot and its potential for sowing. Its primary purpose is to assess the physiological quality of seeds and to express their capacity to promote rapid and uniform emergence, resulting in the establishment of normal seedlings under a wide range of environmental conditions (Baalbaki et al., 2009).

The results of this analysis are essential, as seeds with high vigor tend to exhibit superior field performance, particularly under adverse conditions, positively influencing germination rate and uniformity, seedling emergence, early seedling growth, and, consequently, crop yield (Peske; Villela; Meneghello, 2019).

Seed vigor, in turn, refers to the set of seed properties that determine the potential for rapid and uniform emergence and seedling development under a wide range of field conditions (AOSA, 2002). In this context, vigor tests consist of a set of assessments applied to seed lots with the objective of estimating their physiological performance and establishment capacity under field conditions.

Several tests are routinely employed for the evaluation of soybean seed vigor, including accelerated aging, tetrazolium, electrical conductivity, and seedling growth tests (Vieira et al., 2003). Although widely used, these methods present important limitations, particularly regarding the time required for their execution and the inherent subjectivity of the evaluation process, which may compromise the reproducibility and standardization of the results (Pinto et al., 2015).

Given this scenario, research in seed technology has directed efforts toward the development of Computer Vision-based approaches, enabling the automation of vigor assessment through Digital Image Processing and Artificial Intelligence techniques, thereby providing greater objectivity, speed, and consistency in the obtained results (da Silva, 2024).

However, for computer vision systems to extract reliable information and perform quantitative analyses from images, the correct segmentation of regions of interest is essential. This stage consists of the precise division of a digital image into its constituent parts, separating them from the background and from one another based on neighborhood similarity or the identification of discontinuities (Zhang and Li, 2014; Anand et al., 2022). Segmentation constitutes one of the most critical stages in computer vision systems, since the overall performance of the system depends directly on its quality (Xu, 2024). Therefore, accurate and consistent segmentation is an essential prerequisite for proper feature extraction and classification.

To contribute to this demand, the present study aims to evaluate the impact of transfer learning on the segmentation of soybean seedling parts using a U-Net architecture, employing the systematic replacement of 53 distinct backbones as encoders. The comparative analysis seeks to identify which pre-trained architectures achieve superior performance in the semantic segmentation of soybean seedling parts, providing support for the selection of more robust and accurate models for applications in seed technology.

Materials and Methods

Input Dataset

For model training, a dataset consisting of 80 images was used, each containing 50 soybean seedlings at three days after germination, including some non-germinated seeds, in order to represent the natural variability observed in commercial seed lots. The dataset comprises eight soybean cultivars, with 10 images per cultivar, all acquired at a resolution of 4000 × 3000 pixels

under controlled conditions to ensure visual uniformity among samples.

Image annotation was performed manually at the pixel level by a specialist using the Roboflow platform. After the annotation process, the dataset consisted of 11,500 individualized anatomical instances, distributed among 4,000 cotyledons, 3,700 hypocotyls, and 3,800 radicles, which constituted the sample effectively used for model training and evaluation.

Fig 1. Sample of the dataset used, showing (a) the original image and (b) the ground truth.

In order to increase the variability of the training set and improve the model's generalization capability, a data augmentation strategy was employed. Controlled geometric transformations were applied to both the images and their corresponding segmentation masks while preserving pixel-to-pixel correspondence, including horizontal and vertical flips and random rotations of up to ±10°. As a result, each original image generated one augmented version, totaling 160 images and 23,000 labeled instances.

Additionally, considering that the high resolution of the original images (4000 x 3000 pixels) imposed a substantial computational cost, the images were resized to 640 x 640 pixels prior to the training stages. This reduction aimed to make the experiments computationally feasible, enabling the execution of different model configurations and parameter settings under appropriate hardware conditions.

U-Net Architecture

Image segmentation is a classification task in which a label is assigned to each pixel in the image, such that pixels sharing the same label are associated through visual or semantic properties, thereby separating the object(s) of interest from the image background by means of these labels [25]. One of the most well-known segmentation methods in the field of deep learning is U-Net, developed by Olaf Ronneberger et al. [26], originally designed for biomedical image segmentation tasks through a symmetric structure composed of two parts: an encoder and a decoder.

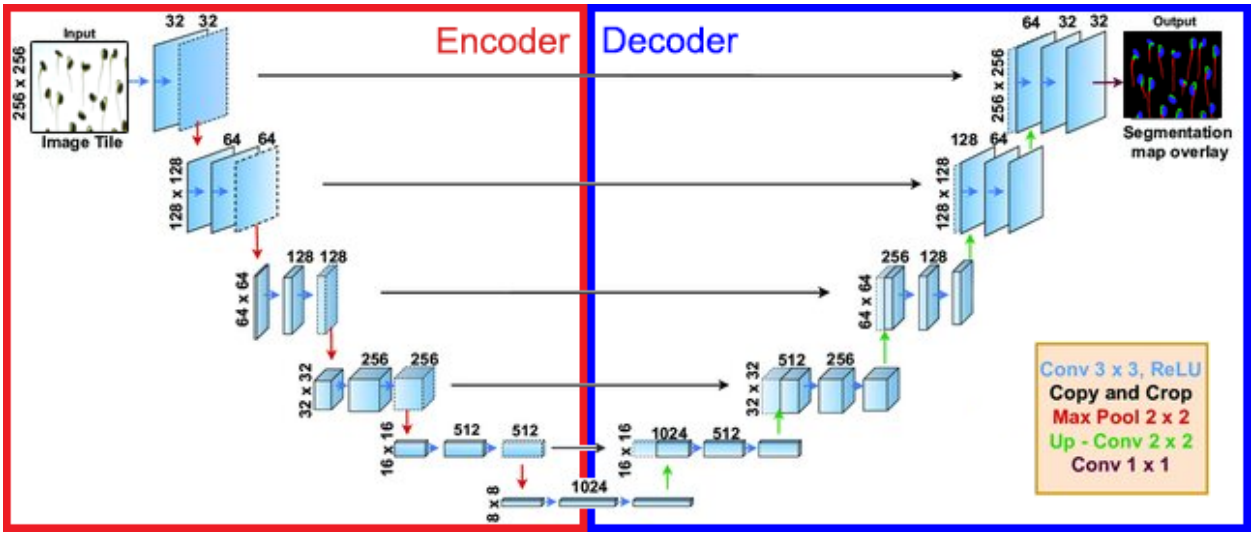

Fig 2. Diagram of the U-Net architecture.

In the first part, feature extraction is performed through convolutional layers, while image dimensionality is reduced through pooling layers, resembling a conventional CNN structure. In the second part, the opposite process takes place, where the image is progressively expanded through up-sampling layers, mapping a pixel to a region and filling it with the same value. In addition to these components, skip connections are employed, concatenating the feature maps from the encoder with the corresponding positions in the decoder, thereby increasing the number of channels and compensating for information loss during decoding, which consequently improves the accuracy of the final result.

Table 1 presents a structural summary of a typical U-Net architecture, considering an RGB input image with a resolution of 640 x 640 pixels and an output consisting of four classes.

Table 1. Structural configuration of the U-Net architecture adopted in this study.

Estágio	Módulo	Camadas	Resolução	Nº de Filtros	Nº de Parâmetros
Encoder	Block 1	2 x Conv (3x3)	640 x 640	64	38 K
	Block 2	MaxPool + 2x Conv (3×3)	320 x 320	128	221 K
	Block 3	MaxPool + 2x Conv (3×3)	160 x 160	256	885 K
	Block 4	MaxPool + 2x Conv (3×3)	80 x 80	512	3.54 M
Bottleneck	Centro	MaxPool + 2x Conv (3×3)	40 x 40	1024	14.16 M
Decoder	Bloco 1	UpConv + Concat + 2x Conv	80 x 80	512	7.08 M
	Bloco 2	UpConv + Concat + 2x Conv	160 x 160	256	1.77 M
	Bloco 3	UpConv + Concat + 2x Conv	320 x 320	128	442 K

	Bloco 4	UpConv + Concat + 2x Conv	640 x 640	64	110 K
Saída	Conv 1x1	Conv (1x1)	640 x 640	4	260

In the final stage, a 1×1 convolution is applied to map the feature maps to an output with dimensions of 640 × 640 × 4, corresponding to the four semantic classes considered. In this way, U-Net produces a segmented mask with the same spatial resolution as the input image, enabling the precise identification of the different anatomical structures of interest at the pixel level.

Segmentation Using Pre-Trained Transfer Learning Models

In general, transfer learning refers to the process by which a model developed to solve a specific problem is reused, either wholly or partially, to solve a related problem. In the context of deep learning, this approach consists of employing a neural network previously trained on a large dataset, such as ImageNet, as a starting point for training on a new task, enabling the reuse of previously learned feature representations.

In this strategy, the model intended for the target problem incorporates one or more layers from the pre-trained model, which may be kept frozen or fine-tuned during training, depending on the degree of similarity between the source and target domains. In the present study, the segmentation of soybean seedling parts was evaluated using 53 models based on different encoder architectures (backbones), all of which were pre-trained on ImageNet.

The backbones considered encompass a diverse set of deep neural network architectures belonging to different families that are well established in the literature. Models from the ResNet family (ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152) and ResNeXt (ResNeXt-50-32x4d and ResNeXt-101-32x8d) were evaluated, as well as their extensions incorporating Squeeze-and-Excitation attention mechanisms, including SE-ResNet (SE-ResNet-101 and SE-ResNet-152) and SE-ResNeXt (SE-ResNeXt-50-32x4d and SE-ResNeXt-101-32x4d). Dual Path Network (DPN) models (DPN-68, DPN-98, and DPN-131) and DenseNet family architectures (DenseNet-121, DenseNet-161, DenseNet-169, and DenseNet-201) were also considered.

Additionally, classical architectures from the VGG family (VGG-11, VGG-13, VGG-16, and VGG-19, with and without batch normalization) were included, as well as models based on the Inception architecture, specifically Inception-v4 and Inception-ResNet-v2. Efficient architectures designed to reduce computational cost were also evaluated, including EfficientNet (B0 to B5), their variants implemented in the timm library (EfficientNet-B0 to B6 and EfficientNet-Lite0 to Lite4), as well as MobileNet-V2 and Xception. Finally, models based on hierarchical Transformers, represented by the MiT family (MiT-B0 to MiT-B4), were also considered, enabling a comparative analysis between convolutional approaches and those based on attention mechanisms.

Evaluation Metrics

The performance of the segmentation models was evaluated using metrics that quantify the quality of the correspondence between the predicted segmentations and the reference annotations (ground truth). The metrics adopted in this study were Precision, Recall, the Dice Coefficient, and Intersection over Union (IoU).

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

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

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

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

Where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) represent the classification outcomes, and A and B correspond to the sets of pixels in the predicted segmentation and the reference segmentation, respectively.

Precision measures the proportion of true positives relative to the total number of positive predictions, reflecting the model's ability to reduce false positives, which are associated with Type I errors. In contrast, Recall represents the proportion of actual positive instances that are correctly identified and is sensitive to the occurrence of false negatives, which are related to Type II errors.

The Dice Coefficient evaluates the similarity between the predicted segmentations and the reference segmentations by quantifying the degree of overlap between them, with values ranging from 0 (indicating no overlap) to 1 (corresponding to perfect segmentation). Complementarily, the IoU metric, also known as the Jaccard Index, measures the ratio between the area of intersection and the area of union of the segmented regions, with higher values indicating better segmentation quality.

Experiment Design

The experiments were conducted in the Google Colab environment, running on an infrastructure based on an Intel® Xeon® CPU @ 2.20 GHz processor with 12 logical cores, x86_64 architecture, and 52 GB of RAM. For computational acceleration, an NVIDIA L4 GPU equipped with 24 GB of dedicated memory and support for CUDA 12.4 was employed.

The implementation of the models and experimental procedures was carried out in Python, using the PyTorch framework as the basis for the development and training of the neural networks. Additionally, the Segmentation Models PyTorch (SMP) library was employed for the definition and configuration of the U-Net-based semantic segmentation architectures and their respective backbones.

Each model was trained for 100 epochs, with weight optimization performed using the AdamW algorithm, adopting a fixed learning rate of 1×10^{-4} and a batch size of 4. The dataset was partitioned into training and testing subsets using a ratio of 70% for training and 30% for testing.

A composite loss function combining multiclass cross-entropy loss and Dice loss with equal weights (0.5 and 0.5) was employed. Cross-entropy operates at the pixel level, promoting training stability and the correct discrimination of anatomical structures, whereas Dice loss operates at the regional level, reducing the impact of class imbalance and improving the segmentation of less represented regions, such as the hypocotyl. This combination balances local accuracy and global morphological consistency, making the optimization process more robust.

Results and Discussion

Figure 3 presents a global comparison of the performance of the 53 evaluated backbones, considering the average metrics obtained at the final training epoch. The graphs reveal consistent differences among the architecture families, indicating that deeper models and those employing advanced feature extraction strategies tend to exhibit more stable and superior performance.

In contrast, simpler or more classical architectures demonstrate greater variability, reflecting limitations in modeling the more complex structures of soybean seedlings. This overview establishes the context for the subsequent analyses, in which the best-performing models are examined in greater detail.

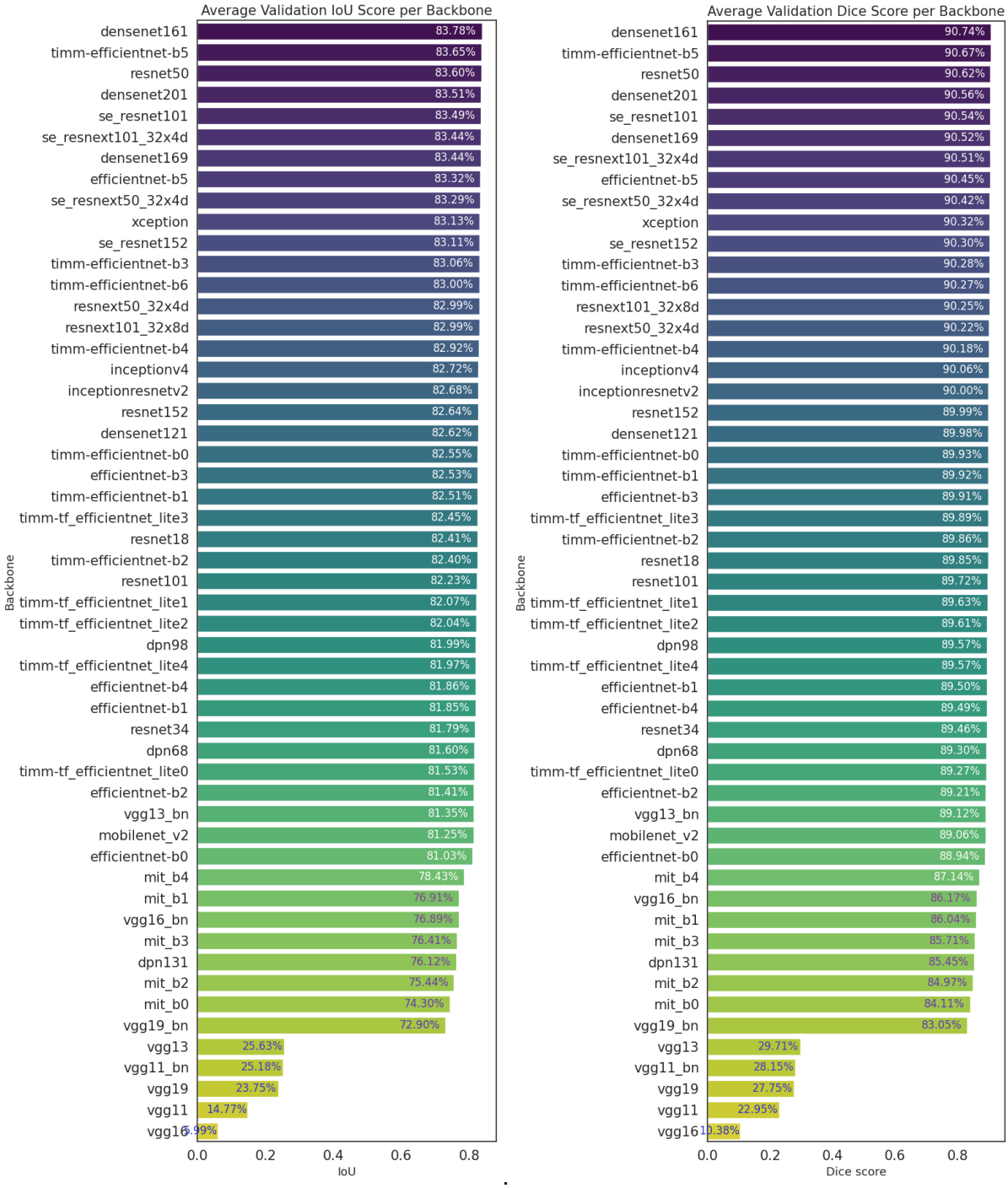

Fig 3. Mean values of (a) IoU and (b) Dice for the 53 evaluated backbones.

IoU tends to be more conservative and more sensitive to false positives and false negatives, providing a more rigorous assessment of the spatial accuracy of the segmentation. In contrast, the Dice metric is less penalizing and may overestimate overlap, especially in small regions, frequently resulting in higher values.

Due to these characteristics, the IoU metric was adopted as the primary criterion for selecting the best model for soybean seedling part segmentation, as it provides a more discriminative evaluation of segmentation quality.

Ten backbones were selected based on a percentage-based threshold criterion relative to the

global mean, considering only models whose IoU performance exceeded the overall average by at least 10%. The selected backbones were: resnet50, se_resnet101, se_resnext50_32x4d, se_resnext101_32x4d, densenet169, densenet201, densenet161, efficientnet-b5, xception, and timm-efficientnet-b5, ensuring the inclusion of the best-performing representatives for comparative analysis.

As shown in Figure 4, Class 3 represents the greatest challenge in the dataset, consistently presenting the lowest IoU values across all evaluated backbones. This behavior is associated with the greater morphological complexity and visual variability of this class, factors that directly affect the performance of segmentation models.

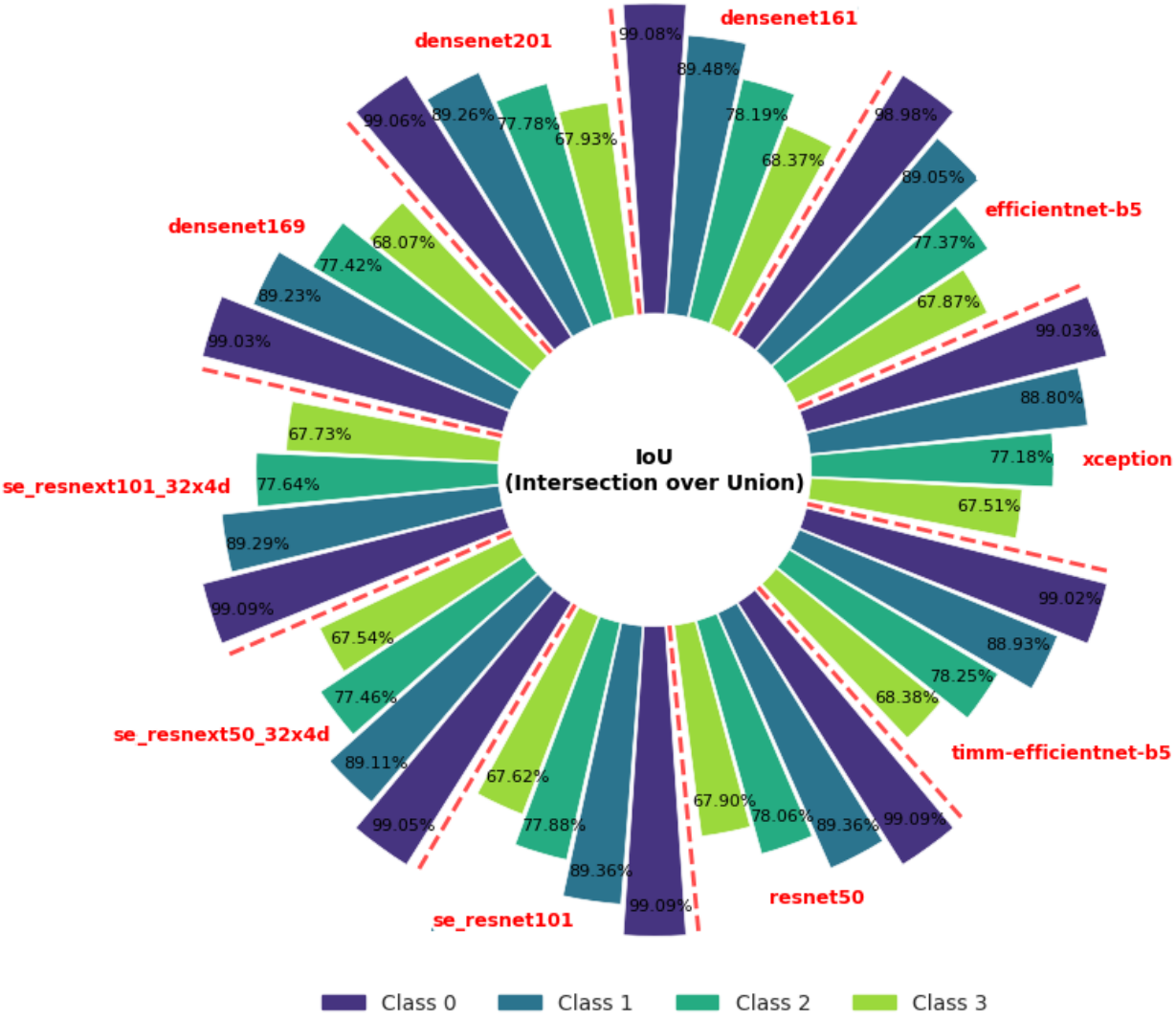

Fig 4. IoU values for each class obtained by the 10 best-performing selected backbones.

In this context, although the absolute differences in IoU among the backbones are relatively small, prioritizing performance on the most complex class becomes essential, as it ensures greater reliability in the segmentation of critical seedling structures. Based on this criterion, the timm-EfficientNet-B5 (IoU = 68.38%), DenseNet-161 (IoU = 68.37%), and DenseNet-169 (IoU = 68.07%) backbones stood out, exhibiting slightly superior performance compared with the other evaluated models.

Figure 5 presents examples of segmentations generated by models using the timm-efficientnet-

b5 backbone (c), the densenet169 backbone (d), and the densenet161 backbone (e).

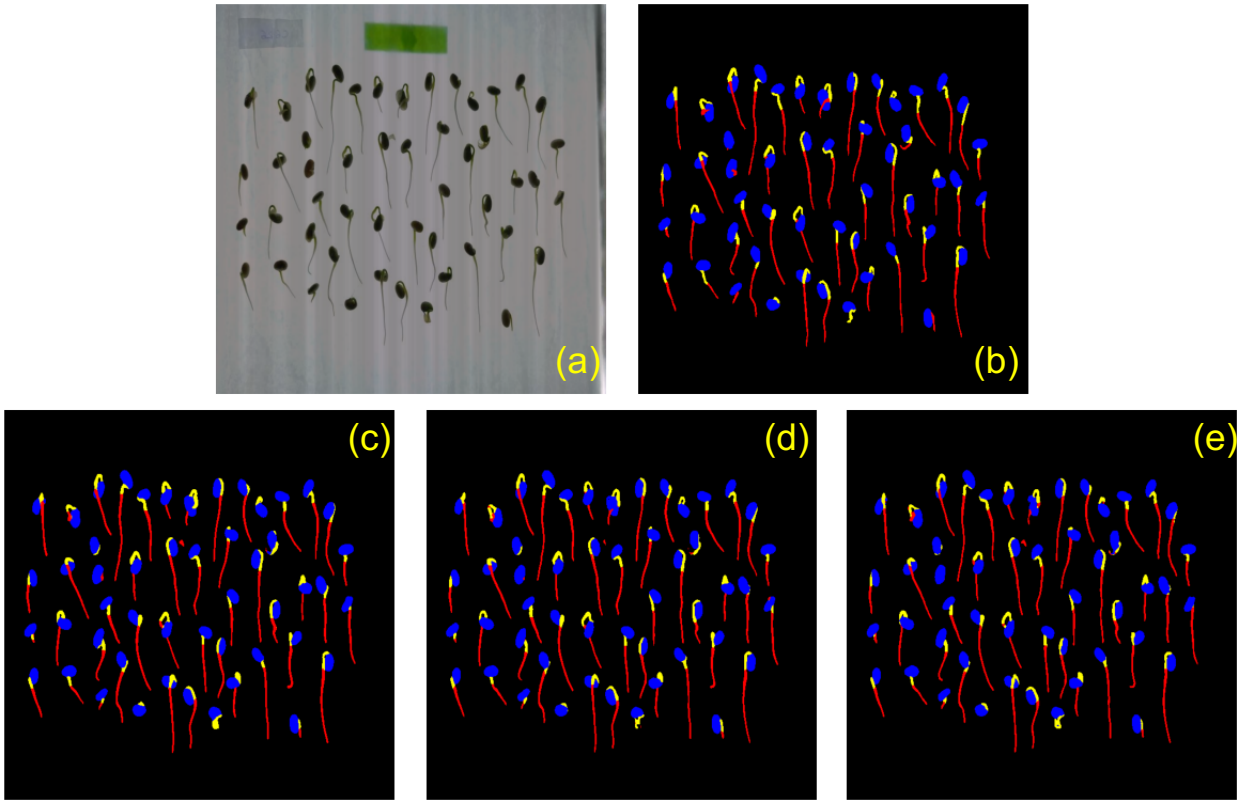

Fig 5. (a) original image, (b) ground truth, (c) segmentation generated by the model with the timm-efficientnet-b5 backbone, (d) segmentation generated by the model with the densenet169 backbone, and (e) segmentation generated by the model with the densenet161 backbone.

Considering that, in potential subsequent applications, the segmentation results may be used to measure the length of different seedling parts, Type I errors (false positives), which result in the overestimation of segmented regions, are particularly critical, as they may lead to an overestimation of structure length and, consequently, to the incorrect inference of higher seedling vigor. From a commercial perspective, this type of error is more detrimental than Type II errors (false negatives), which tend to underestimate length and therefore suggest lower vigor in a more conservative manner. In light of this criterion, the combined analysis of Precision and Recall becomes essential (Table 2).

Table 2. Precision and Recall values obtained during the validation of the best-performing backbones.

Backbone	Precision	Recall
timm-efficientnet-b5	0.8843	0.9307
densenet169	0.8907	0.9205
densenet161	0.8992	0.9159

Although the model based on timm-EfficientNet-B5 achieved the highest Recall value (93.07%), indicating a lower occurrence of false negatives, DenseNet-161 stands out by presenting the highest Precision (89.92%), reflecting greater rigor in the prediction of positive pixels and a lower incidence of false positives.

DenseNet-169, in turn, exhibits an intermediate balance between Precision (89.07%) and Recall (92.05%), representing a well-balanced solution. However, considering specifically the asymmetric penalization of errors, in which overestimation is more critical than underestimation,

DenseNet-161 emerges as the most suitable model for commercial applications, combining competitive IoU performance on the most complex class with greater control over Type I errors, thereby aligning with the practical requirements of seedling vigor assessment.

Conclusion

This study systematically evaluated 53 ImageNet pre-trained backbones for the semantic segmentation of anatomical parts of soybean seedlings, encompassing classical convolutional architectures, efficient models, and attention-based approaches. The results indicated that the hypocotyl class constitutes the main challenge within the dataset, highlighting the influence of morphological complexity and visual variability on model performance and the need to prioritize this class during the model selection process.

Model selection was based on objective quantitative criteria, focusing on performance in the most challenging class, resulting in the selection of ten backbones for detailed analysis, among which timm-EfficientNet-B5, DenseNet-161, and DenseNet-169 stood out.

Considering the practical application of the system, in which segmentation is used for measuring the length of seedling parts and for vigor inference, Type I errors (false positives), associated with the overestimation of segmented regions, were considered more critical than Type II errors. The analysis of Precision and Recall metrics indicated that, although timm-EfficientNet-B5 exhibited greater detection capability, DenseNet-161 stood out due to its greater rigor in positive predictions.

Therefore, DenseNet-161 was selected as the most suitable model for operational applications, as it combines consistent segmentation performance with greater control over false positives, meeting the practical requirements of seedling vigor assessment.

As future work, the direct integration of the segmentation stage with morphological measurement and vigor classification algorithms is recommended, further expanding the potential application of the proposed methodology in both industrial and research contexts.

Acknowledgments

This study was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil – Finance Code 001. Danielo G. Gomes acknowledges the support of the National Council for Scientific and Technological Development (CNPq) under grant numbers 311845/2022-3 and 403996/2025-2.

References

Baalbaki, R. Z. et al. (2009). Seed vigor testing handbook, *handbook on seed testing*. (p. 341) AOSA, Ithaca, NY, USA.

Peske, S. T.; Villela, F. A.; Meneghello, G. E. (2019). Sementes: fundamentos científicos e tecnológicos. Pelotas: UFPel, (pp. 355- 405).

ASSOCIATION OF OFFICIAL SEED ANALYSTS [AOSA] (2002). *Seed vigortesting handbook*. (Contribution, 32): Stillwater, OK.

Vieira, R.D., Bittencourt, S.R.M. & Panobianco, M. (2003). *Seed vigour - an important component of seed quality in Brazil*. ISTA News Bulletin, n. 126, (pp. 21-22). https://www.seedtest.org/upload/cms/user/STI126Oct2003.pdf

Pinto, C. A. G., Carvalho, M. L. M., Andrade, D. B., Leite, E. R. & Chalfouns, I. (2015) Image analysis in the evaluation of the physiological potential of maize seeds. *Revista Ciência Agronômica*, v. 46, n. 2, (p. 319-328). doi: https://doi.org/10.5935/1806-6690.20150011.

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), (pp. 869-880). doi: https://doi.org/10.3390/agriengineering6020050

Zhang, H., & Li, D. (2014). Applications of computer vision techniques to cotton foreign matter inspection: A review. *Computers and Electronics in Agriculture*, 109, (pp. 59–70). doi: https://doi.org/10.1016/j.compag.2014.09.004

Anand, R., Ritesh Kumar Mishra, & Khan, R. (2022). Plant diseases detection using artificial intelligence. *Elsevier EBooks*, (pp. 173–190). Doi: https://doi.org/10.1016/b978-0-323-90550-3.00007-2

Xu, Y. (2024). Application of Image Segmentation Algorithms in Computer Vision. *Frontiers in Computing and Intelligent Systems*, 7(3), (pp. 17–20). doi: https://doi.org/10.54097/gq1s6737

Swarnendu Ghosh, Nibaran Das, Ishita Das, and Ujjwal Maulik. (2019). Understanding Deep Learning Techniques for Image Segmentation. *ACM Computing Surveys* (CSUR). doi: https://doi.org/10.1145/3329784

Ronneberger, O.; Fischer, P.; Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241), Springer: Cham, Germany.