IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4138~4146
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4138-4146

Seeding precision: a mask region based convolutional neural networks classification approach for the classification of paddy seeds

Rajashree Nambiar[1,2], Ranjith Bhat[1,2], Varuna Kumara[2,3]
[1]Department of Robotics and Artificial Intelligence Engineering, NMAM Institute of Technology, NITTE (Deemed to be University), Nitte, Karnataka, India
[2]Faculty of Engineering and Technology, JAIN (Deemed to be University), Bengaluru, India
[3]Department of Electronics and Communication Engineering, Moodlakatte Institute of Technology, Kundapura, India

Article Info

Article history:

Received Feb 23, 2024
Revised Jun 25, 2024
Accepted Jun 28, 2024

Keywords:

Bounding box
Mask region based convolutional neural networks
Paddy classification
Region of interest
Synthetic data

ABSTRACT

The generation of sufficient training data that is accurately labelled for a deep neural network involves a significant amount of effort and frequently constitutes a bottleneck in the implementation process. For the purpose of this research, we are training a neural network model to perform instance segmentation and classification of crop seeds for various rice cultivars. Synthetically constructed dataset is used here. The concept of domain randomization, which offers a productive alternative to the laborious process of data annotation, serves as the basis for our methodology. We make use of the domain randomization technique in order to produce synthetic data, and the mask region-based convolutional neural network (Mask R-CNN) architecture is utilized in order to train our neural network models. A cultivar name is used to designate the seeds, and they are differentiated from one another using colors that are comparable to those used in the actual dataset of paddy cultivars. Our mission focuses on the identification and categorization of rice paddy varieties within automatically generated photographs. Farmers are able to accurately sort crop seeds from a variety of rice cultivars with the use of this approach, which is particularly useful for phenotyping and optimizing yields in laboratory settings.

Corresponding Author:

Ranjith Bhat
Department of Robotics and Artificial Intelligence Engineering, NMAM Institute of Technology NITTE (Deemed to be University)
Nitte, Karkala Taluk, Udupi, Karnataka 574110, India
Email: ranjithbhat@gmail.com

1. INTRODUCTION

Deep learning has gained popularity in both the scientific and industrial spheres. Deep-learning methods, such as convolutional neural networks (CNNs) [1], are extensively employed in computer vision for tasks like image classification, object detection, and semantic as well as instance segmentation [2]–[4]. Using these methods has also affected agriculture. According to Kamilaris and Boldú [5], image-based phenotyping detects weeds, agricultural diseases, and fruits. Deep learning complements the sector's [6] abundant high-context data. However, deep learning requires considerable labelled data preparation. As of 2012, ImageNet has 1.2 million training images and 150,000 validation/test images with hand categorization [7]. 328,000 pictures with 2.5 million tagged objects from 91 categories were used for the 2014 common objects in context (COCO) object detection task [8]. This annotating the dataset order may be challenging for a researcher. Agriculture research reveals that a grain head detection network may be trained with 52 photos averaging 400 objects per image [9] and a crop stem detection network with 822 images [10]. These case studies demonstrate that ImageNet classification and COCO detection require more data than specialized work. While domain adaptation and active learning are used in plant/bio science applications to cut labor costs, researchers find annotating unpleasant because it's like running a marathon without a target [11]–[13].

The sim2real transfer, or learning from synthetic images, reduces manual annotations. Training data for plant image analysis was prepared similarly. Using synthetic plant models, Isokane *et al.* [14] predicted branching pattern, while several researchers [15], [16] generated realistic images from generated datasets using generative adversarial network (GAN). GAN-generated images were used to train a neural network for Arabidopsis leaf counting by Giuffrida *et al.* [17]. Similar to Arsenovic *et al.* [18] StyleGAN28 created plant disease classification training pictures. However, sim2real generates nearly limitless training data. To bridge the sim2real gap, domain randomization trains deep networks with enormous variants of synthetic images with randomly selected physical attributes. Domain randomization is related to data augmentation (e.g., randomly flipping and rotating photographs), but the synthetic environment can reflect variety under numerous scenarios, unlike genuine images. The conventional approach, as shown in Figure 1, involves manually labeling photos to create the training dataset. In contrast, our suggested method eliminates this step by utilizing a synthetic dataset for the crop seed instance segmentation model.

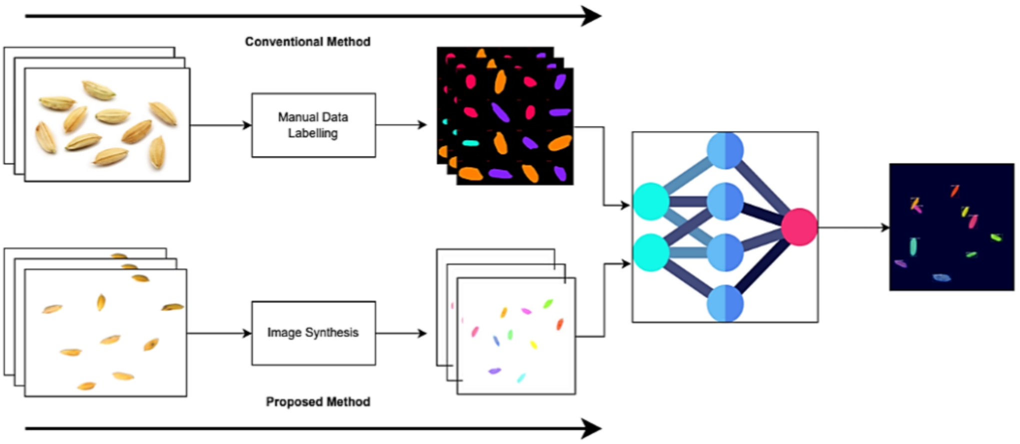

Figure 1. Overview of the suggested training procedure for seed instance segmentation

This approach involves training deep neural network models to perform the intricate task of instance segmentation, wherein individual seeds are classified and precisely localized within images. By leveraging synthetically generated datasets and randomization techniques, we can create a robust and versatile training environment for these models. The benefits of paddy seed classification using deep learning are manifolds. It not only significantly reduces the labor and time required for seed sorting but also ensures consistency and precision in the classification process. Moreover, it has the potential to improve crop management practices, as accurate cultivar-level seed data can inform decisions related to planting, fertilization, and pest control. Many studies have found that using seed width as a primary parameter increases rice output. The focus on morphological seed traits shows promise for improving agricultural productivity and promoting biological research. It is important to remember, nevertheless, that many earlier researches evaluated seed form using qualitative measures, Vernier callipers, or manually annotating images using image-processing tools. This phenotyping procedure may lead to quantification mistakes that differ amongst annotators and is often labor-intensive.

2. RELATED WORKS

Widiastuti *et al.* [19] suggests that rice seed quality is traditionally determined by human visual assessment. This method is highly subjective when comparing rice varieties with similar physical features. The research recommends flatbed scanning and digital image processing to assess rice seed purity to overcome this barrier. A field-based grow out test (GOT) validates rice seed shape analysis in this method. An analysis of the 14 morphological qualities found relationships in only six area, feret, minimum feret, aspect ratio, roundness, and solidity. Growing methods, harvesting, shipping, and post-harvest processing can affect seed purity. In addition to quality, seed certificate labels must clearly display seed purity values. The proposed method [20] improves rice seed purity testing due to its speed and cost, grow-out test dependability. It can be difficult to distinguish between seeds with the same morphology during purity testing. Molecular approaches are being studied to differentiate such seeds as a treatment. The method in Adjemout *et al*. [21], employs machine learning and image processing algorithms to categories whole and broken rice by how well they meet national rice quality standards. The objects are classified using CNN technology. The image database used in this study contains self-collected photos of Loc Troi 20 breed rice forms. The photos were taken with a Sony Z1 smartphone's 20.7 MP camera. The experiments reveal that convolutional neural networks have 99.16% precision. Son *et al.* [22] introduced deep-rice, a new rice evaluation method. It extracts distinguishing attributes from rice photo perspectives using a multi-view CNN architecture. Additionally, it uses a redesigned SoftMax loss function to optimize CNN parameters. This created a new rice-rating algorithm under deep-rice, this solves rice grading problems using deep residual networks and deep learning. Wijerathna and Ranathunga [23] describes a computer vision and image processing system for rice seed production that automatically classifies rice types. Since rice seeds from different varieties might look identical in color, shape, and texture, categorizing them correctly is difficult. The study evaluated feature extraction methods to portray rice seeds [24]. They also tested powerful classifiers' performance with these extracted attributes to select the most trustworthy classifier. The research showed that their random forest (RF) categorization technique had an average accuracy rate of 90.54 [25], [26]. The availability of diverse cultivars in different places makes data collecting for this study difficult.

3. METHOD

Four steps are suggested in the model flow contributing to the development of a dependable mechanism for classifying seeds as shown in Figure 2. The initial paddy seed dataset comprises Gidda, Jaya, Jyothi, and M4 paddy seeds. The diverse range of data in this dataset enables our programme to accurately distinguish between different types of seeds.

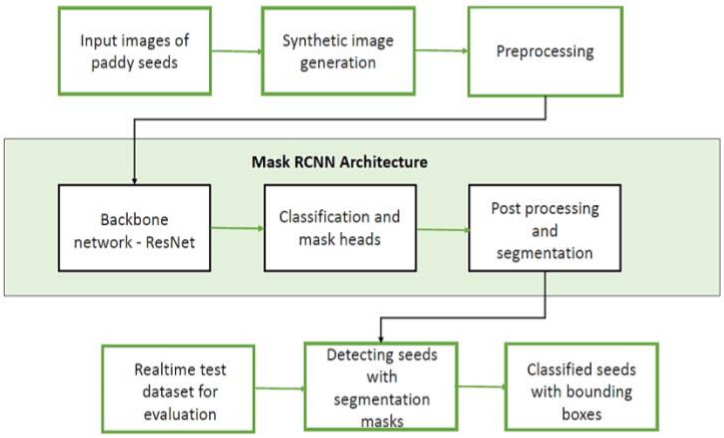

Figure 2. Proposed architecture of paddy seed classification

Creating a comprehensive database of seed images is crucial for doing further data collecting. This pool serves as the framework for synthetic images, which are an essential tool for research purposes. We employ domain randomization to generate a set of 2,000 synthetic images, with 1,400 images designated for training purposes and 600 images reserved for testing. Subsequently, the artificial dataset is employed to train the model using the mask region-based convolutional neural network (Mask R-CNN) methodology. This stage enables our model to recognize and classify seeds, providing predictions that include the seed name, as well as the bounding box and overlay color. Ultimately, the model undergoes rigorous testing to assess its efficacy and suitability in real-world scenarios. The performance of the system can be evaluated in many contexts using assessment techniques that consider both synthetic and real-world datasets. The architecture of the Mask R-CNN model is illustrated in Figure 3.

Region of interest align (RoIAlign) aims to extract a small, fixed-size feature map (like $H \times W$) from each region of interest with sub-pixel accuracy, improving upon the older RoI pooling method by avoiding quantization errors. In (1) is the representation of interpolated feature value at a specific location (x, y) within the output feature map of the RoI.

$$f(x, y) = \sum_{i,j} g(i, j).max(0, 1 - |x - i|).\ max(0, 1 - |y - i|) \quad (1)$$

Where $\sum_{i,j}$ is a summation over the neighborhood of the point (x, y) in the input feature map. And we consider the values of neighboring points (i, j) in the original feature map. $g(i, j)$ is the feature value located at (i, j) in the input feature map from which we are trying to extract the RoI. $\max(0, 1 - |x - i|)$ and $\max(0, 1 - |y - i|)$ calculate the bilinear interpolation weights. To determines the class as mentioned in (2), we use the SoftMax activation function with weight W and bias b. Here, ΔBox is the predicted offsets in (3).

$$Class = softmax(W.x + b) \quad (2)$$

$$\Delta Box = W'.x + b' \quad (3)$$

Here, in (4) outlines a common pattern in deep learning, especially in tasks related to computer vision and pattern recognition, where *x* would be a multi-dimensional array (a tensor) representing the image data and M is the convoluted output through a series of CNN layers with a sigmoid activation function.

$$M = \sigma(CNN(x)) \quad (4)$$

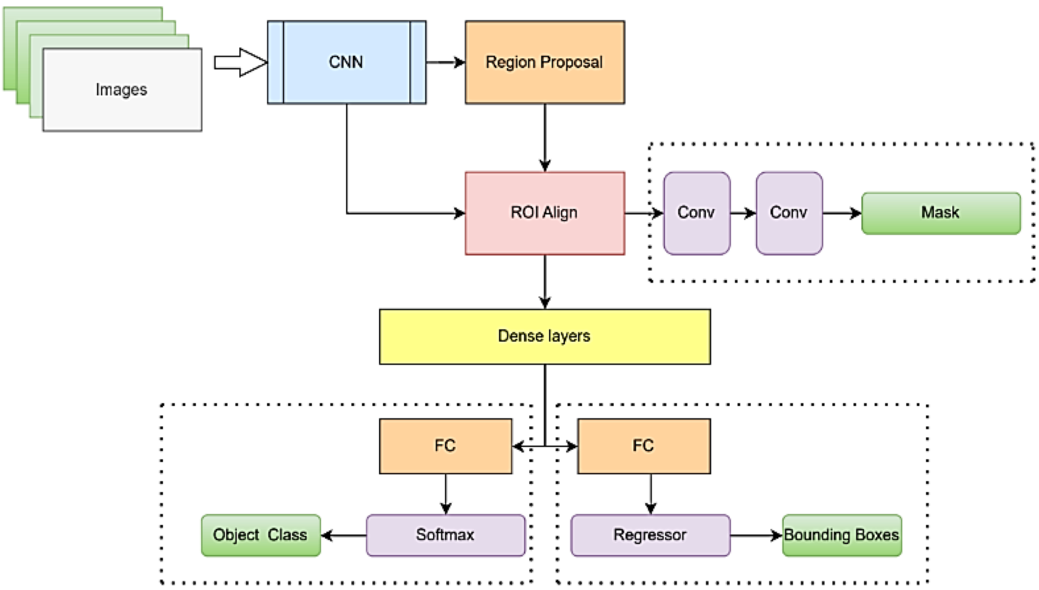

Figure 3. Mask R-CNN model structure

3.1. Collecting paddy seeds for dataset

We carefully collected a dataset of four paddy seed classes to segment crops. These classes represent Karnataka paddy seed varieties Gidda, Jaya, Jyothi, and M4. Our segmentation model will be trained on this carefully curated dataset to reliably identify and categories paddy seed classes in agricultural photography.

3.2. Synthetic image generation, preprocessing and training

We applied cutting-edge domain randomization to optimize our Mask R-CNN model for paddy seed classification via synthetic picture synthesis. This method uses four rice seed types, a varied seed pool, and resizing the photographs to 1024×1024 pixels. Starting with this seed pool, we created a huge dataset of 2,000 meticulously created synthetic photos for training and testing our model. Domain randomization is used to train a neural network classifier that equals the performance of current models trained just on actual datasets, demonstrating its versatility and efficacy. Our area of randomization experiment showed that subject variety is more relevant than secondary criteria like illumination and texturing in determining model correctness. Mask R-CNN with Keras or TensorFlow was employed for seed classification. The repository setup network designs and loss functions were employed. Features were extracted using ResNet101, a residual network initialized using MS COCO dataset weights [27]. Next, we fine-tuned our counterfeit seed picture dataset using 10 training epochs with 100 steps per epoch and 0.001 learning rate. 1,400 images were used for training and among the 400 images from the 600 in the synthetic dataset, were used for validation and 200 for testing purposes. It is noteworthy that we avoided using picture enhancement when training. The artificial training data maintained a 1024×1024 picture size constantly.

3.3. Realtime dataset for model evaluation

We put the Mask R-CNN model in inference mode and validated it using our validation dataset to appropriately assess its performance. A comprehensive validation approach lets us assess the model's accuracy and durability in real-world situations. We selected a unique dataset of 10 images of seeds from 4 paddy rice kinds for real-world testing. Real-world pictures are always 1024×1204 pixels and follow standard proportions. Our real-world dataset has 20 images with 10 seeds each. Our system accurately predicts and labels each seed with its cultivar name and color-codes each seed variety in the photo. Our model's final test is this real-time dataset, which proves its efficacy and reliability in real-world situations.

4. RESULTS AND DISCUSSIONS

Understanding the features needed to successfully replicate real-world datasets is essential to understand synthetic data's value in deep learning. Our major foundation was that the neural network must learn to detect and separate randomly inserted or overlapping seeds into objects during seed instance segmentation. While designing our synthetic picture collection, we prioritized seed orientations over seed textures. The number of images in the training dataset and the resolution and variance of the seed images used to produce synthetic images were expected to significantly affect model performance. Providing exact bounding boxes and masks for each seed item allowed our model to correctly detect instances in the supplied photographs and segment each seed. To train machine learning models for computer vision applications like image categorization, object recognition, and picture synthesis, many synthetic images are needed. Synthetic images generated as in Figure 4 are created by a model or other means rather than using real-world data.

Figure 4. Synthetic image generation using seed image pool

Mask R-CNN segments paddy seeds precisely. The masks clearly identified photo seed regions. This shows how the model accurately displays all 4 types of seeds. Accuracy around 99%, for all seed varietals as shown in Figure 5. Form and size of seeds (grains) affect crop quality and production. Our workflow allows us to phenotype many seeds without considering orientation during image acquisition.

Figure 5. Realtime samples and the visualized raw output showing the accuracy

A comprehensive analysis of training and validation losses was performed in the paddy classification study for Jaya, Gidda, Jyothi, and M4, using 1,176, 1,159, 1,157, and 1,152 samples distributed across an 80:20% train-test split. Train/box_loss, train/seg_loss, train/dfl_loss, train/cls_loss, and val/box_loss, val/seg_loss, val/dfl_loss, and val/cls_loss was evaluated. The results provided intriguing model performance insights. Our experimental investigation used Mask R-CNN as the fundamental method for picture segmentation, benchmarking it against a variety of segmentation models in Table 1. To evaluate each model's ability to segment complicated images, the structural similarity index measure (SSIM), accuracy, precision, recall, and F1-score were assessed. Mask R-CNN achieved an SSIM score of 0.90, demonstrating its ability to maintain structural similarity between segmented pictures and ground truth. Mask R-CNN surpassed its competitors with 0.95 accuracy, 0.94 precision, 0.94 recall, and 0.94 F1-score, demonstrating its resilience in detecting and outlining objects in images.

Table 1. Comparative analysis of image segmentation models based on SSIM and other performance metrics presenting an overview of the performance of various segmentation models across multiple metrics such as SSIM, accuracy, precision, recall, and F1-score

Model	SSIM	Accuracy	Precision	Recall	F1-Score	Remarks
U-Net [28]	0.85	0.92	0.90	0.89	0.89	High precision in biomedical image segmentation.
FCN [29]	0.83	0.90	0.88	0.87	0.87	Good for general purposes, versatile.
DeepLab (v3+) [30]	0.88	0.93	0.91	0.92	0.91	Captures multiscale information effectively.
PSPNet [31]	0.86	0.91	0.89	0.90	0.89	Effective global context information.
SegNet [32]	0.82	0.89	0.87	0.86	0.86	Efficient, suitable for real-time applications.
RefineNet [33]	0.87	0.92	0.90	0.91	0.90	High-resolution imagery, fine-grained segmentation.
Enet [34]	0.80	0.88	0.85	0.84	0.84	Optimized for speed, real-time processing.
HRNet [35]	0.89	0.94	0.92	0.93	0.92	Maintains high-resolution representations
Mask R-CNN [36]	0.90	0.95	0.94	0.94	0.94	Superior for instance segmentation with high detail.

Here, the Table 2 shows class correctness and Figure 6 illustrates confusion matrix. These results demonstrate Mask R-CNN's remarkable instance segmentation capabilities, especially in high-precision and detail settings. Our findings demonstrate Mask R-CNN's crucial role in image segmentation technologies, giving new insights for researchers and practitioners using deep learning for complicated image processing applications.

Table 2. Accuracy prediction for the separate 4 classes Gidda, Jaya, Jyothi, and M4

Ground truth	Mask Color	Predicted Name	Accuracy
Jaya	Yellow	Jaya	0.983
Jyothi	Pink	Jyothi	0.998
Gidda	Cyan	Gidda	1.00
Jaya	Violet	Jaya	0.992
Gidda	Blue	Gidda	0.997
M4	Yellow	M4	0.999
Jaya	Orange	Jaya	0.985

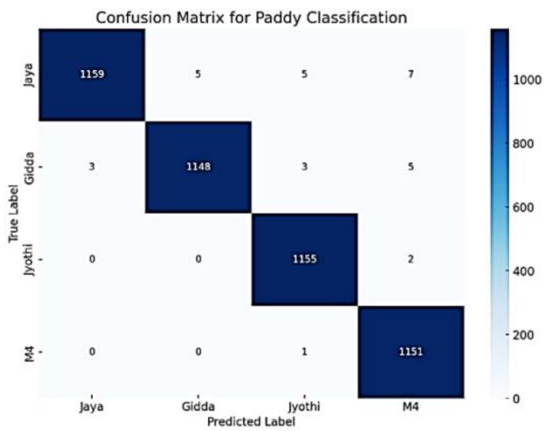

Figure 6. Visualizing the accuracy of classifying Jaya, Gidda, Jyothi, and M4 using confusion matrix

Across the training phase, the model demonstrated a consistent decrease in both segmentation (seg_loss) and classification (cls_loss) losses. This downward trend in losses indicates that the model effectively learned to differentiate between the classes and segment the paddy images accurately. Notably, the box loss (box_loss) also exhibited a similar decreasing trend, highlighting the model's proficiency in localizing and precisely delineating the paddy areas within the images. During validation, the observed trends in losses were relatively stable, albeit with minor fluctuations. The validation losses closely mirrored the training losses, affirming the model's generalization ability and robustness in recognizing and classifying paddy classes unseen during training. The marginal fluctuations in validation losses might indicate a slight overfitting tendency or the complexity of distinguishing certain classes within the validation set. Overall, the model's performance showcases promising capabilities in accurately segmenting and classifying different paddy varieties. The consistent reduction in losses during training, coupled with validation losses aligning closely with training losses, signifies the model's competency in learning the distinctive features of each class.

4.1. Metrics evaluation

4.1.1. Binary classification metrics

Precision (B) and recall (B) metrics were assessed to measure the model's performance in differentiating between binary classes. Precision (B) signifies the accuracy of positive class predictions, while recall (B) gauges the model's ability to capture all positive instances within the dataset. All the plots are shown in the Figure 7.

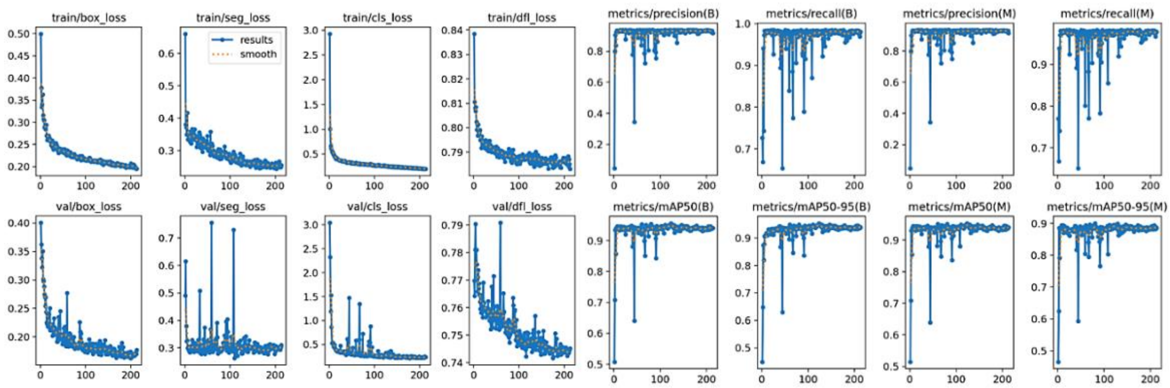

Figure 7. Plot of loss, precision and recall during training and validation for our dataset

4.1.2. Mean average precision metrics

The evaluation measured the mean average precision (mAP) at 50% intersection over union (mAP50) for both binary (B) and multiclass (M) situations. These metrics evaluate the model's precision in identifying and categorising objects at different intersection over union thresholds. The achieved mAP50 scores for both binary and multiclass scenarios demonstrated consistent and high values, indicating the model's accuracy in localising and classifying objects at various thresholds. The plot axes of are represented on the top each graph obtained.

5. CONCLUSION

The model's robust performance in differentiating paddy types is demonstrated by binary and multiclass classification metrics in the proposed work. The model's high precision and recall ratings for binary and multiclass classifications show its ability to accurately identify specific classes while balancing positive cases across the dataset. To solve this challenge, we created synthetic datasets to train the model and test it using a validation dataset using domain randomization. The model can segment these seeds into instance segments from the validation dataset, which comprises synthetically created seeds with appropriate precision and low error. Additionally, the model's strong mAP metrics at varied intersection over union thresholds demonstrate its ability to localise and categorise paddy data across changing object overlap. These comprehensive evaluations and high-performance metrics demonstrate the model's paddy classification efficacy, demonstrating its potential for real-world applications in reliably recognising and categorising varied rice kinds. Refinement and optimisation could improve the model's performance and usefulness in agriculture or automated crop monitoring systems.

REFERENCES

[1] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning,” *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[3] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[4] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[5] A. Kamilaris and F. X. P. -Boldú, “Deep learning in agriculture: a survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[6] Y. Kaneda, S. Shibata, and H. Mineno, “Multi-modal sliding window-based support vector regression for predicting plant water stress,” *Knowledge-Based Systems*, vol. 134, pp. 135–148, Oct. 2017, doi: 10.1016/j.knosys.2017.07.028.

[7] O. Russakovsky *et al.*, “ImageNet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[8] Y. Aytar and A. Zisserman, “Immediate, scalable object category detection,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2014, pp. 2385–2392, doi: 10.1109/CVPR.2014.305.

[9] W. Guo *et al.*, “Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy,” *Frontiers in Plant Science*, vol. 9, Oct. 2018, doi: 10.3389/fpls.2018.01544.

[10] X. Jin, S. Madec, D. Dutartre, B. de Solan, A. Comar, and F. Baret, “High-throughput measurements of stem characteristics to estimate ear density and above-ground biomass,” *Plant Phenomics*, vol. 2019, Jan. 2019, doi: 10.34133/2019/4820305.

[11] S. Ghosal *et al.*, “A weakly supervised deep learning framework for sorghum head detection and counting,” *Plant Phenomics*, vol. 2019, Jan. 2019, doi: 10.34133/2019/1525874.

[12] A. L. Chandra, S. V. Desai, V. N. Balasubramanian, S. Ninomiya, and W. Guo, “Active learning with point supervision for cost-effective panicle detection in cereal crops,” *Plant Methods*, vol. 16, no. 1, Dec. 2020, doi: 10.1186/s13007-020-00575-8.

[13] T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis, “Using DeepLabCut for 3D markerless pose estimation across species and behaviors,” *Nature Protocols*, vol. 14, no. 7, pp. 2152–2176, Jul. 2019, doi: 10.1038/s41596-019-0176-0.

[14] T. Isokane, F. Okura, A. Ide, Y. Matsushita, and Y. Yagi, “Probabilistic plant modeling via multi-view image-to-image translation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 2906–2915, doi: 10.1109/CVPR.2018.00307.

[15] C. Lazo, “Segmentation of skin lesions and their attributes using generative adversarial networks,” in *LatinX in AI at Neural Information Processing Systems Conference 2019*, Dec. 2019, doi: 10.52591/lxai201912083.

[16] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2242–2251, doi: 10.1109/CVPR.2017.241.

[17] M. V. Giuffrida, H. Scharr, and S. A. Tsaftaris, “ARIGAN: synthetic arabidopsis plants using generative adversarial network,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, IEEE, Oct. 2017, pp. 2064–2071, doi: 10.1109/ICCVW.2017.242.

[18] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, “Solving current limitations of deep learning based approaches for plant disease detection,” *Symmetry*, vol. 11, no. 7, Jul. 2019, doi: 10.3390/sym11070939.

[19] M. L. Widiastuti, A. Hairmansis, E. R. Palupi, and S. Ilyas, “Digital image analysis using flatbed scanning system for purity testing of rice seed and confirmation by grow out test,” *Indonesian Journal of Agricultural Science*, vol. 19, no. 2, pp. 49-56, Dec. 2018, doi: 10.21082/ijas.v19n2.2018.p49-56.

[20] K. S. Jamuna, S. Karpagavalli, M. S. Vijaya, P. Revathi, S. Gokilavani, and E. Madhiya, “Classification of seed cotton yield based on the growth stages of cotton crop using machine learning techniques,” in *2010 International Conference on Advances in Computer Engineering*, IEEE, Jun. 2010, pp. 312–315, doi: 10.1109/ACE.2010.71.

[21] O. Adjemout, K. Hammouche, and M. Diaf, “Automatic seeds recognition by size, form and texture features,” in *2007 9th International Symposium on Signal Processing and Its Applications*, IEEE, Feb. 2007, pp. 1–4, doi: 10.1109/ISSPA.2007.4555428.

[22] N. H. Son and N. Thai-Nghe, “Deep learning for rice quality classification,” in *2019 International Conference on Advanced Computing and Applications (ACOMP)*, IEEE, Nov. 2019, pp. 92–96, doi: 10.1109/ACOMP.2019.00021.

[23] P. Wijerathna and L. Ranathunga, “Rice category identification using heuristic feature guided machine vision approach,” in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, Dec. 2018, pp. 185–190, doi: 10.1109/ICIINFS.2018.8721396.

[24] S. Khunkhett and T. Remsungnen, “Non-destructive identification of pure breeding rice seed using digital image analysis,” in *The 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE)*, IEEE, Mar. 2014, pp. 1–4, doi: 10.1109/JICTEE.2014.6804096.

[25] H.-T. Duong and V. T. Hoang, “Dimensionality reduction based on feature selection for rice varieties recognition,” in *2019 4th International Conference on Information Technology (InCIT)*, IEEE, Oct. 2019, pp. 199–202, doi: 10.1109/INCIT.2019.8912121.

[26] Y. Wu, Z. Yang, W. Wu, X. Li, and D. Tao, “Deep-Rice: deep multi-sensor image recognition for grading rice,” in *2018 IEEE International Conference on Information and Automation (ICIA)*, IEEE, Aug. 2018, pp. 116–120, doi: 10.1109/ICInfA.2018.8812590.

[27] T.-Y. Lin *et al.*, “Microsoft COCO: common objects in context,” *Computer Vision–ECCV 2014: 13th European Conference,* Zurich, Switzerland, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[28] O. Ronneberger, “Invited talk: U-Net convolutional networks for biomedical image segmentation,” *Bildverarbeitung für die Medizin*, Berlin, Heidelberg: Springer, 2017, doi: 10.1007/978-3-662-54345-0_3.

[29] M. Goyal, M. Yap, and S. Hassanpour, “Multi-class semantic segmentation of skin lesions via fully convolutional networks,” in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, SCITEPRESS - Science and Technology Publications, 2020, pp. 290–295, doi: 10.5220/0009380300002513.

[30] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[31] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[32] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[33] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: multi-path refinement networks for high-resolution semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 5168–5177, doi: 10.1109/CVPR.2017.549.

[34] W. Bai, "Enet semantic segmentation combined with attention mechanism," *Research Square*, 2021, doi: 10.21203/rs.3.rs-425438/v1.

[35] J. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686.

[36] M. Gajja, "Brain tumor detection using mask R-CNN," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 8, pp. 101–108, Jul. 2020, doi: 10.5373/JARDCS/V12SP8/20202506.

BIOGRAPHIES OF AUTHORS

Rajashree Nambiar holds a Masters of Technology degree from Nitte University, India in 2014. She also received his Bachelor of Engineering from Visvesvaraya Technological University, Belagavi, India. She is currently an Assistant Professor at Department of Robotics and Artificial Engineering at NMAM Institute of Technology, NITTE (Deemed to be University), Nitte, India. She is currently a research scholar at the JAIN (Deemed to be University), Bengaluru. Her research includes artificial intelligence, machine learning, deep learning, image, and signal processing. She can be contacted at email: raji24oct@gmail.com or rajashree.n@nitte.edu.

Ranjith Bhat holds a Masters of Technology degree from Nitte University, India in 2011. He also received his Bachelor of Engineering from Visvesvaraya Technological University, Belagavi, India. He is currently an Assistant Professor at Department of Robotics and Artificial Engineering at NMAM Institute of Technology, NITTE (Deemed to be University), Nitte, India. He is currently a research scholar at the JAIN (deemed to be) university, Bengaluru. His research includes artificial intelligence, machine learning, deep learning, network security, and computer networks. He can be contacted at email: ranjithbhat@gmail.com or ranjith.bhat@nitte.edu.

Varuna Kumara is a Research Scholar in the Department of Electronics Engineering at JAIN Deemed to be University, Bengaluru, India. He also received his B.E. and M.Tech. from Visvesvaraya Technological University, Belagavi, India in 2009 and 2012 respectively. He is currently Assistant Professor at Electronics and Communication Engineering in Moodlakatte Institute of Technology, Kundapura, India. His research interests are in artificial intelligence, signal processing, and control systems. He can be contacted at email: vkumarg.24@gmail.com.