Smart Agricultural Technology 12 (2025) 101594

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Dual-task multi-species network design and training using advanced augmentation techniques ☆

Jonathan Ford *, Edmund Sadgrove, David Paul

School of Science and Technology, University of New England, Armidale, Australia

ARTICLE INFO

Keywords:
Convolutional neural networks
Computer vision
Site-specific weed management
Precision agriculture
CutMix augmentation

ABSTRACT

Accurate weed localization in pasture environments is essential for effective site-specific weed management (SSWM). This study evaluates two dual-task convolutional neural networks (CNNs) - a truncated ConvNeXt (tCN) and a truncated UniStemNet (tUSN) - for simultaneous plant segmentation and spraypoint detection across four weed species in south-east Australian pastures. To assess model performance, two testing regimes were used: a pooled test, where all images were randomly split for training and validation, and condition-invariance tests, in which models were trained and evaluated on entirely separate groups of images collected under different environmental conditions. Four data augmentation strategies were compared, including advanced image blending techniques CutMix and MixUp, which had not previously been explored in this domain. In the pooled test, the tUSN model achieved a mean Intersection over Union (mIU) of up to 0.897 for plant segmentation, while the tCN model reached an F1-score of 0.880 for spraypoint detection. Under condition-invariance testing, CutMix augmentation provided the most consistent generalization, with mIU values up to 0.897 for plant segmentation, and F1-scores up to 0.954 for spraypoint detection. HistMatch normalization assisted with model generalization in weaker augmentation setups. These findings demonstrate that advanced augmentation and normalization strategies are critical for achieving robust CNN-based weed detection across diverse real-world field conditions.

1. Introduction

Efficient and accurate weed management is a critical challenge in modern agriculture, particularly in extensive pasture systems where manual weed control is impractical. Site specific weed management (SSWM) aims to target undesirable pasture species while leaving other sections of the pasture untreated, thereby reducing chemical use and limiting damage to surrounding vegetation [12]. However, the success of such systems relies on the ability to detect, localize, and discriminate between weed species under diverse environmental conditions [8].

Convolutional neural networks (CNNs) are commonly used in weed localization studies [14,3,18,5], but their large number of parameters makes them prone to overfitting, leading to reduced performance when tested on unseen samples. Although data augmentation techniques can be used to mitigate this issue [14], to date, most augmentation strategies employed in weed localization studies have been limited to combinations of spatial transformations [21,6,1], pixel intensity shifts [16] and the addition of random noise or blurring [7]. By comparison, sample mixing techniques, where two or more samples are combined to create a new one, have not yet been investigated. Such techniques, including CutMix [19] and MixUp [20] may offer promising avenues for developing more robust and diverse augmentation strategies.

Previous work has explored dual-task computer vision models for weed management, particularly where varied treatment methods, such as herbicide application or mechanical removal are required [10,11,21,6]. These models typically use convolutional neural networks (CNNs), and take an input image of pasture or crop field and output multiple heatmaps, one showing weed (or crop) locations, and another indicating the centre of each plant. While foliar herbicides or fertilizers are generally applied across the entire plant, treatments like root-absorbed herbicides or mechanical removal target the plant's centre, which may be referred to as the spraypoint or stem [10,6].

Ford et al. [6] demonstrated the feasibility of real-time, pixel-level weed detection and spraypoint localization in complex pasture environ-

☆ The first author was supported by a Commonwealth of Australia RTP scholarship, with a top-up scholarship from the "Novel Autonomous Robotic Weed Control to Maximise Agricultural Productivity" (Commonwealth Grant CRCPIX000099) CRC. The authors would like to thank the NSW Department of Primary Industries and Agent Oriented Software for providing the dataset.

* Corresponding author.
E-mail addresses: jford3@myune.edu.au (J. Ford), esadgro2@une.edu.au (E. Sadgrove), dpaul4@une.edu.au (D. Paul).

https://doi.org/10.1016/j.atech.2025.101594
Received 29 July 2025; Received in revised form 15 October 2025; Accepted 31 October 2025

Available online 6 November 2025

ments, using a dual-task CNN. Their model was a ConvNeXt-inspired architecture with separate decoders for each task, which outperformed prior approaches, including models by Lottes et al. [10] and Zhang et al. [21], across multiple experimental conditions. To improve generalization under variable imaging conditions, a histogram-based normalization method (HistMatch) was introduced, which significantly enhanced performance, particularly for models with lower representational capacity.

However, the above research was limited to a dataset containing a single weed species, serrated tussock (*Nassella trichotoma*, hereafter tussock). This left untested whether the models were sufficiently robust to handle the situation where multiple weed species were present, as is commonly the case in pastures. For this reason, this study was conducted to further test the abilities and limitations of the models in a more complex pastoral setting. Specifically, the following research questions were addressed:

- How well do the CNN models handle multispecies detection in pastures, including underrepresented weeds?
- How robust are the models to unseen growing conditions during testing?
- Can the spraypoint detector maintain accuracy with additional species present?
- Do advanced augmentation techniques improve segmentation performance in pasture settings?

2. Related work

2.1. Dual-task CNN models

Two key studies have proposed architectures for the dual task of plant segmentation and stem detection. Lottes et al. [10] designed a model with a shared encoder and two separate decoders. The encoder is built using dense blocks composed of multiple densely connected 3×3 convolutional layers, with each layer outputting a fixed number of feature maps (the growth rate). Bottleneck layers using 1×1 convolutions help reduce computational load, and downsampling is achieved via 5×5 convolutions with stride 2. The decoder arms use transpose convolutions (2×2, stride 2) for upsampling, and skip connections are implemented via filter-axis concatenation.

In contrast, Zhang et al. [21] constructed a model based on MobileNetV2's inverted residual modules (IRMs), which combine depth-wise and point-wise convolutions to improve efficiency while preserving accuracy. The encoder contains multiple stages of IRMs, with downsampling achieved by stride-2 depth-wise convolutions. The architecture features task-specific decoder branches that split off from different encoder resolutions. These decoder arms include intra-task connections to enhance feature flow and a cross-task fusion link from the segmentation decoder to the stem decoder, ensuring that stem localization is guided by plant structure. Each decoder arm produces an independent output, and deep supervision is applied at all outputs to strengthen training.

The model described in the preceding paragraph was further refined by Ford et al. [6] through simplification of the decoder arms, an increase in parameter count, and the incorporation of ConvNeXt-style architectural elements [9], including a "patchify" stem, ConvNeXt modules, and adjusted stage ratios. This updated model was evaluated on a joint plant segmentation and spraypoint localization task using a dataset containing serrated tussock, with ground-level images collected from south-east Australian pastures.

2.2. Data augmentation in weed localization studies

Data augmentation strategies artificially expand the size of a training dataset by generating new images from existing ones through a variety of random image transformations. This is particularly important when training CNNs with a large number of parameters, which are prone to overfitting [14]. In the field of weed classification, augmentation techniques are typically limited, and include spatial transformations such as random horizontal and vertical flips [21,13], zooms, rotations, and shears [7,6,1,14,18]. Other common approaches include adjusting pixel intensity and/or contrast [6,1,14], and adding random noise or blurring [7,16]. More specialized techniques have also been explored, such as dividing images into $n \times n$ patches and augmenting each patch independently [2], or applying style-transfer methods to simulate environmental effects like dew on plants [17].

3. Methods

3.1. Dataset

The dataset consisted of images containing four weed species: serrated tussock, horehound, nettle, and variegated thistle. The tussock images were the same as those used in [6]. The images were captured at three distinct locations — Jugiong (lat. -34.862, long. 148.319), Wallacetown (-34.947, 147.488), and Yelangello (-34.860, 149.168) — over a 20-month period from October 2021 to May 2023, spanning multiple collection dates. All images were resized to 768×576 pixels, and the corresponding segmentation maps were downsampled by a factor of 16, resulting in a final output size of 48×36 pixels. The downsampled output greatly decreases the computational complexity of the models, while maintaining sufficient resolution to allow for effective activation of herbicide spray nozzles.

Not all weed species were present at every location or collection date. The proportion of image patches covered by each species (and the negative class), as well as the number of images across the different locations and collection dates is presented in Fig. 1.

Two output segmentation maps were generated: one for plant segmentation and another for spraypoint segmentation. The spraypoint represents the centre of the plant, where root-absorbed herbicides are applied to maximize uptake and effectiveness. Spraypoints were labelled only for serrated tussock, as the other weed species are typically managed using foliar herbicides that are applied to the entire plant.

3.2. CNN models

Two models were evaluated, based on those referred to as Zhang's model and our model in [6]. Zhang's model is the UniStemNet model by Zhang et al. [21], which was adapted for the spraypoint detection task. To align with the downsampled output segmentation maps, both models were truncated by removing the decoder arms responsible for processing higher-resolution features.

For Zhang's models, this means that there is a single decoder arm for each task, at a downsampling factor of 16. This has removed the bidirectional feature map exchanges that operated between adjacent intra-task decoder arms. However, the cross-task feature fusion that operates from the plant segmentation decoder to the spraypoint decoder has been retained.

For our model, both tasks have a single decoder arm, at the 32× downsampled scale for the spraypoint task, and at the 16× downsampled scale for the plant task. For the plant task, a U-Net style connection is used to concatenate upsampled feature maps from the 32× downsampled scale to the 16× downsampled feature maps from the encoder arm. This is in contrast to the bidirectional feature map exchanges that were present in the original model. As is the case for Zhang's model, the cross-task feature fusion has been retained. The number of layers in the encoder has also been halved, so that there are 2, 6 and 2 layers at downsampling scales of 8, 16 and 32 respectively. These models will be called the truncated UniStemNet (tUSN) and truncated ConvNeXt (tCN) models in the remainder of the paper. An overview of the structure of both models used in this paper is presented in Fig. 2.

3.3. Augmentation strategies

In order to evaluate the effectiveness of different augmentation strategies, including sample-mixing techniques, four different augmen-

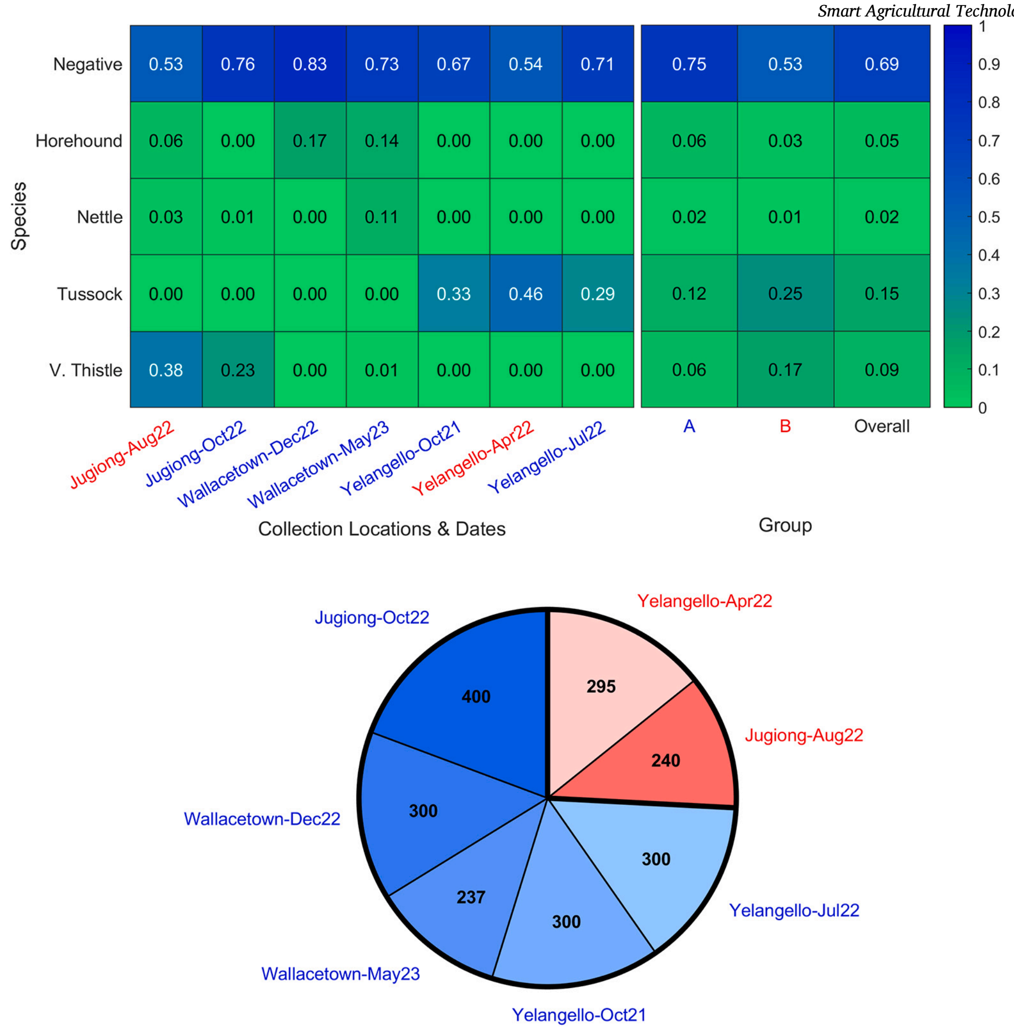

Fig. 1. The proportion of image patches covered by each species (and the negative class) and the number of images across different locations and collection dates. The images have been separated into groups A (blue), and B (red).

tation schemes were tested in this study, being Zoom/Rotate, CutMix [19], MixUp [20] and both CutMix and MixUp. For Zoom/Rotate augmentation, images were randomly zoomed with limits of −20% to 10% and rotated with limits of $\pm\frac{\pi}{8}$.

CutMix blends two images by replacing specific regions of one image with patches from another, effectively combining their content. This enhances regularization and encourages the model to learn to identify objects even when key discriminative features are partially missing. While vanilla CutMix randomly selects a rectangular region for replacement, in this study, a free-form mask was generated by applying a Gaussian filter, with $\sigma = 2$, to a randomly generated mask, followed by binarization. The threshold for binarization was set to the median pixel value of the mask prior to binarization, ensuring that each image contributed an equal area to the augmented image. The mask was chosen to have dimension 48×36, matching the size of the output masks. The process of mask generation is shown on the right side of Fig. 3.

By contrast, MixUp generates new samples by blending two images at the pixel level, which is achieved using a weighted average of their pixel values. This technique encourages smoother decision boundaries and reduces overfitting. In this study, the coefficient for the weight factor was drawn from a beta distribution with $\alpha = \beta = 2$, as this worked well in preliminary tests. In all forms of augmentation, random horizontal and vertical flips were also applied. Examples of the different augmentation strategies are provided in Fig. 3.

3.4. *Testing*

An initial pooled test was conducted in which all images were combined into a single dataset, from which an 80%/20% training/validation split was generated. This setup was used to optimize the learning rate and to evaluate the different augmentation strategies under conditions where the training and validation data shared the same growing conditions.

To evaluate how well the models generalized to previously unseen conditions, two condition-invariance tests were conducted. The training data were first divided into two groups, A and B, as indicated by text colour in Fig. 1, where blue and red represent groups A and B, respectively. In Invariance Test 1, models were trained and validated on Group A (using an 80%/20% split), and tested on Group B. In Invariance Test 2, the setup was reversed, with Group B used for training and validation, and Group A for testing.

3.5. *Training parameters*

Training was carried out in Keras/TensorFlow as per [6]. Learning rates from 0.0001 to 0.01 were tested in the pooled test. For the condition-invariance tests, a fixed learning rate of 0.001 was used, as this was found to be the most consistent in the pooled test. For the pooled test, 300 epochs of training were used, whereas for the condition-invariance tests, 700 epochs were used to ensure that the training converged to a maximum. For the spraypoint task, due to extreme class

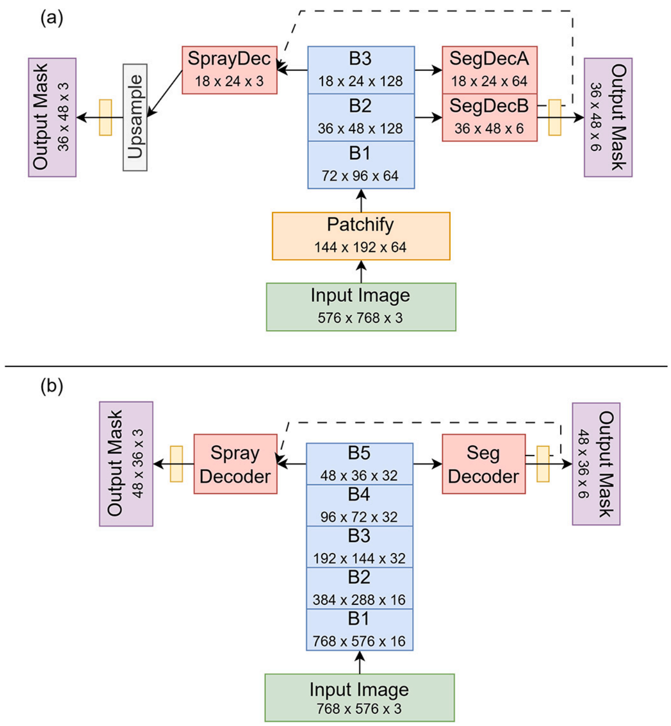

Fig. 2. The truncated ConvNeXt (a) and truncated UniStemNet (b) models used in this paper. Colours represent different network components: green for the input layer, orange for the patchify stem, blue for the encoder, red for the decoder, grey for upsampling, yellow for the loss layer, and purple for the output layer. Arrows indicate the flow of information, while dashed arrows represent cross-task feature fusion.

imbalance, IOU loss was used [15]. For the segmentation task, cross entropy loss was used. Other hyperparameters were set to the values used in [6], without additional optimization. In all cases, the model weights corresponding to the best performance on the validation data were restored at the end of training. Normalization was carried out on a per-image basis. For the pooled test, standard normal normalization was used. For the condition-invariance tests, default, standard normal, and HistMatch normalization techniques were compared. Further information about these normalization techniques can be found in [6].

3.6. Post-processing and metrics

To convert the model-generated heatmaps for the spraypoint task into individual predictions, the following steps were applied. First, a threshold value of 0.5 was used to binarize the heatmap. The resulting binary mask was then processed to identify connected components, corresponding to potential detections. To reduce noise and eliminate small, spurious detections, a minimum size threshold was applied to each component, with values of 2, 4, 6, 8, and 10 pixels tested.

The centroid of each connected component was used to determine the positions of both true and predicted spraypoints. These spraypoints were then matched based on proximity, with a distance threshold of 3.125 (equivalent to 50 pixels in the input image, which is the same threshold used in [6]) used to classify a predicted spraypoint as a true positive. Although this value was not rigorously selected, it is based on an estimation of the precision of the herbicide spray nozzles, and the spread of the central root system of the tussock plants. Each true spraypoint was paired with at most one predicted spraypoint, and vice versa. True spraypoints that were not matched within the required distance were counted as false negatives, while unmatched predicted spraypoints were counted as false positives.

For the plant segmentation task, the mean Intersection over Union (mIU) was computed on a per-patch basis. For the spraypoint task, evaluation was performed using the per-object F1 score.

4. Results

The mIU for the plant segmentation task, and the F1-score for the spraypoint task for the pooled test are presented in Fig. 4. For the tCN model, connected components with a size of at least 4 output pixels were counted as valid detections, whereas for the tUSN model, the threshold was set to 10 pixels, as these were the values that led to the best results (in the pooled test). The average improvement in F1 score (across all augmentation techniques) when small connected components were ignored was 0.001 for the tCN model, compared to 0.049 for the tUSN model.

The results for the condition-invariance tests are presented in Fig. 5 for test one, and Fig. 6 for test two. Each figure presents four heatmaps: the top row corresponds to the tCN model, and the bottom row to the tUSN model. The left column displays results for the plant segmentation task using the mIU metric, while the right column shows results for the spraypoint task using the F1 score. Within each heatmap, rows represent the four different augmentation strategies, and columns represent the three different normalization techniques. The colour scale is non-linear and was chosen to accentuate differences between closely ranked metric scores, particularly at the higher end of the scale. The same colour scale is used for both figures, as well as for Fig. 4, to ensure consistency and comparability.

In test one, CutMix augmentation with default preprocessing performed best for both models and tasks (Fig. 5, second row, left column of each heatmap). The tCN model performed better than the tUSN model for both tasks, with an improvement in metrics of 0.012 and 0.047 for the plant and spraypoint tasks respectively. In test two, CutMix augmentation with default preprocessing achieved the best performance for both models for the plant task (Fig. 6, left-hand heatmaps). For the spraypoint detection task, Both augmentation with HistMatch preprocessing performed best for the tCN model, while Both augmentation with default preprocessing was the top performer for the tUSN model (Fig. 6, right-hand heatmaps). In terms of model comparison, the tUSN model slightly outperformed the tCN model in plant segmentation, with a marginal improvement of 0.001. For the spraypoint detection task, the tCN model performed better, showing a more substantial improvement of 0.015.

Fig. 7 presents the normalized confusion matrix for the plant segmentation task, comparing both models on the combined test using CutMix augmentation. In both cases, confusion between weed species was minimal, with the majority of the mIU error caused by weeds being misclassified as negative.

Fig. 8 presents example outputs from condition invariance test one, using default normalization. The outputs are shown for both the Zoom/Rotate and CutMix augmentation strategies, selected to highlight differences between the two models as well as the effects of each augmentation approach. As observed (columns 3 and 5), models trained with Zoom/Rotate augmentation often produced large regions with incorrect class labels in the plant segmentation task and frequently failed to detect spraypoints.

5. Discussion and conclusion

This study aimed to evaluate the effectiveness of CNN models for the detection and segmentation of multiple weed species in South-East Australian pastures, with four species represented in the dataset. The selected CNNs were dual-task models capable of simultaneously generating a segmentation mask of the target weed and identifying the plant centre, which corresponds to the root zone and is therefore the optimal spraypoint for root-absorbed herbicides, such as Flupropanate. Spraypoint annotations were provided only for serrated tussock, however, which is commonly treated with Flupropanate. The study also placed particular emphasis on augmentation strategies, especially advanced image blending techniques, which have not previously been explored in the context of weed localization.

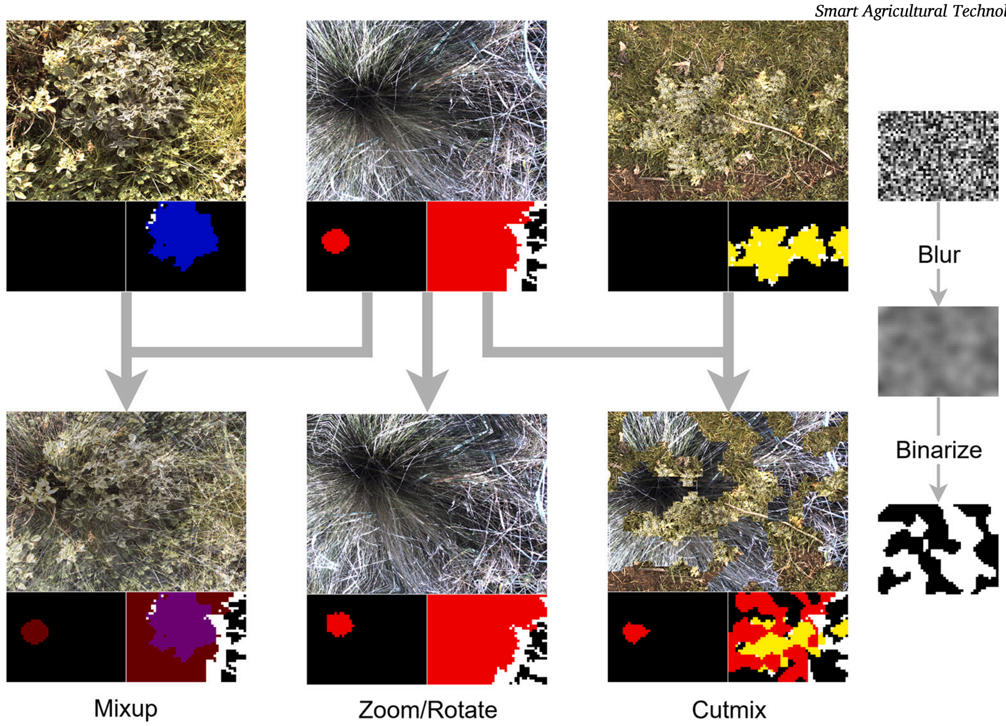

Fig. 3. Examples of the augmentation methods used. The top row shows original samples from the dataset, while the bottom row shows corresponding augmented samples. The masks are shown below the image with the spraypoint mask on the left, and the segmentation mask on the right. The method of mask generation for CutMix is visualized on the right side of the figure.

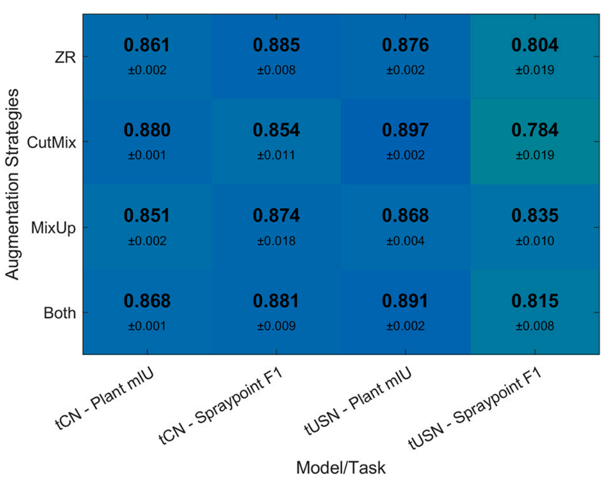

Fig. 4. Results are presented for the pooled test, where all images were combined into a single dataset from which training and testing sets were sampled. Reported values represent the mean ± standard deviation across seven test iterations. Metrics include the mean Intersection over Union (IoU) for the segmentation task and the F1 score for the spraypoint detection task, evaluated for both the truncated ConvNeXt (tCN) and truncated UniStemNet (tUSN) models.

When images from all collection locations and dates were pooled (Fig. 4), both models performed well on the multispecies plant segmentation task, achieving mIU scores between 0.851 and 0.897. On average, the tUSN model outperformed the tCN model by 0.018 across different augmentation strategies. For the spraypoint detection task, F1-scores ranged from 0.784 to 0.880, with the tCN model performing 0.064 better than the tUSN model on average. These results indicate that both models were effective in a multispecies context and that the spraypoint detector remained robust even with an increased number of plant species present.

The dataset used in this study was highly imbalanced, with the negative class covering 69% of the total image area, while individual weed species accounted for between 2% (nettle) and 15% (tussock), as illustrated in Fig. 1. As shown in the example confusion matrices for both models (Fig. 7), weed species with lower representation (horehound and nettle) tended to exhibit reduced classification accuracy. Nevertheless, overall accuracy remained relatively high, with all species achieving scores of at least 0.90. These results highlight the robustness of the selected CNN models, even when faced with significant class imbalance in the dataset.

When the dataset was split based on collection locations and dates, robust model performance could still be achieved, with mIU scores of up to 0.897 for the plant segmentation task, and F1-scores as high as 0.962 for the spraypoint detection task (Figs. 5 and 6). These results demonstrate that the models could generalize well to unseen growing conditions, even in a multispecies context.

However, in this case, the choice of augmentation strategies, and normalization techniques became critically important, with model performance being significantly impacted in the worst cases. In most cases, Zoom/Rotate augmentation performed poorly in the condition invariance tests, with mIU scores dropping as low as 0.468 for plant segmentation and F1-scores as low as 0.160 for spraypoint detection. This poor performance is highlighted in Fig. 8. While these results could be improved somewhat using HistMatch normalization, they were still significantly worse than those using other augmentation techniques.

Overall, CutMix proved to be the most consistent augmentation strategy for condition invariance, irrespective of model, test, task or normalization technique. While CutMix is generally considered to be better adapted to localization tasks, as compared to classification tasks, in which MixUp is superior [4], it is notable that this distinction only became apparent in the condition invariance experiments conducted in this study. The strength of this augmentation strategy, compared to others, may stem from both the approach itself and the specific method of mask generation, which was tailored to reflect the common scenario in pastures where plant overlap leads to partial occlusion of target species.

With respect to normalization strategies, it was found that HistMatch normalization greatly improved results in cases where the models were not robust enough to adapt to changes in growing conditions, as was the

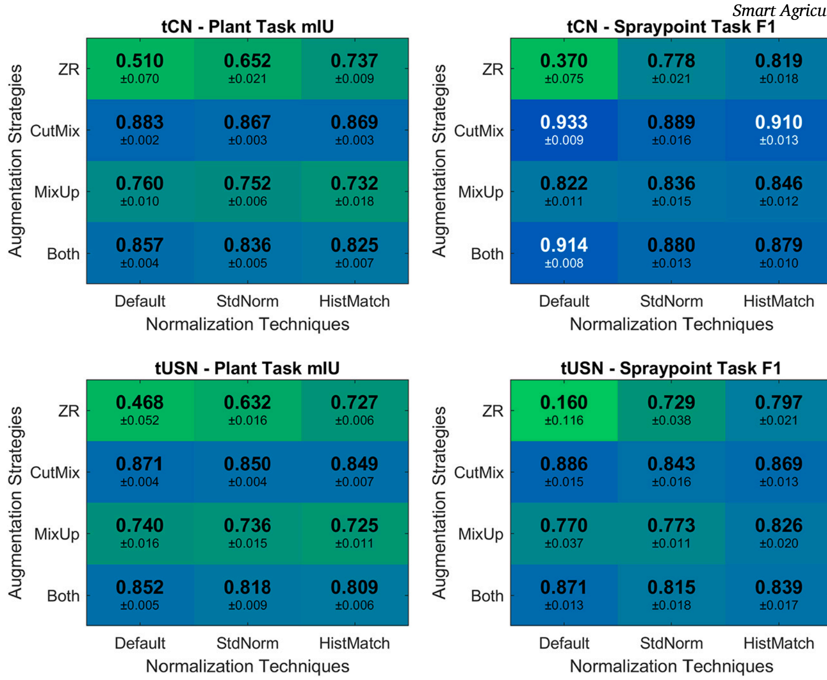

Fig. 5. Results are presented for invariance tests 1, in which group A was used for training and validation, and group B used for testing. Reported values are the same as those presented in Fig. 4, evaluated for default, standard norm and HistMatch normalization, as discussed in [6].

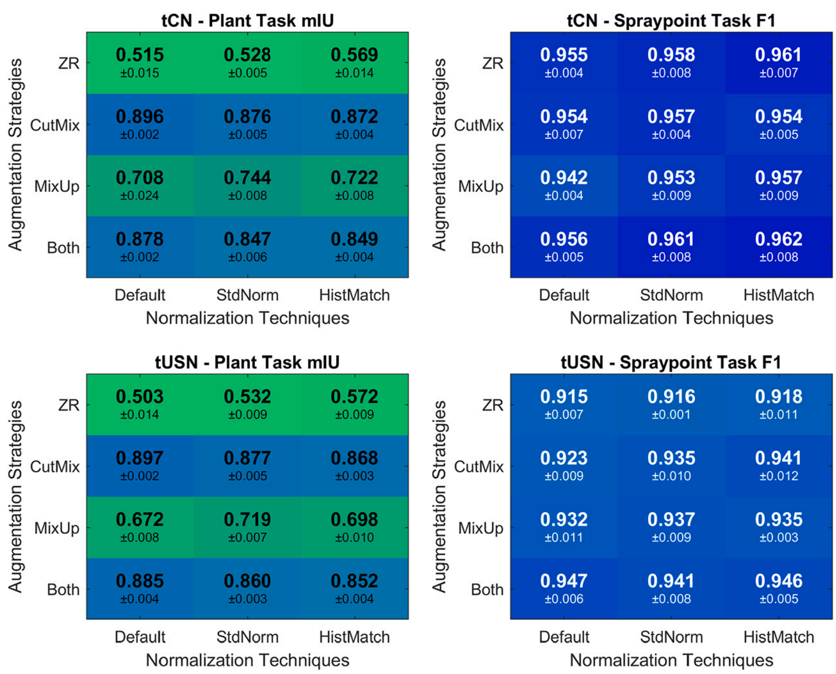

Fig. 6. Results are presented for invariance test 2, in which group B was used for training and validation, and group A used for testing. Reported values are the same as those presented in Fig. 5.

case when Zoom/Rotate augmentation was used. In the case of CutMix, where models already performed strongly with default normalization, applying HistMatch led to reduced performance, which may be due to a loss of image descriptiveness or due to confusion introduced during training by enforcing a standard pixel intensity distribution. Finally, where model performance was strongest, the choice of normalization strategy appeared to have no discernible effect. This is exemplified in the right-hand column of Fig. 6, showing the spraypoint results for invariance test two.

In the pooled test, the parameter-rich ConvNeXt-style tCN model outperformed the tUSN model on the spraypoint task, achieving F1-score improvements ranging from 0.039 to 0.081 depending on the augmentation strategy. Conversely, the tUSN model delivered superior results on the segmentation task, with mIU gains between 0.015 and 0.023. In the condition invariance tests—particularly focusing on the top-performing CutMix augmentation, the tCN model outperformed the tUSN model by an average of 0.008 in mIU for segmentation and 0.033 in F1-score for spraypoint detection.

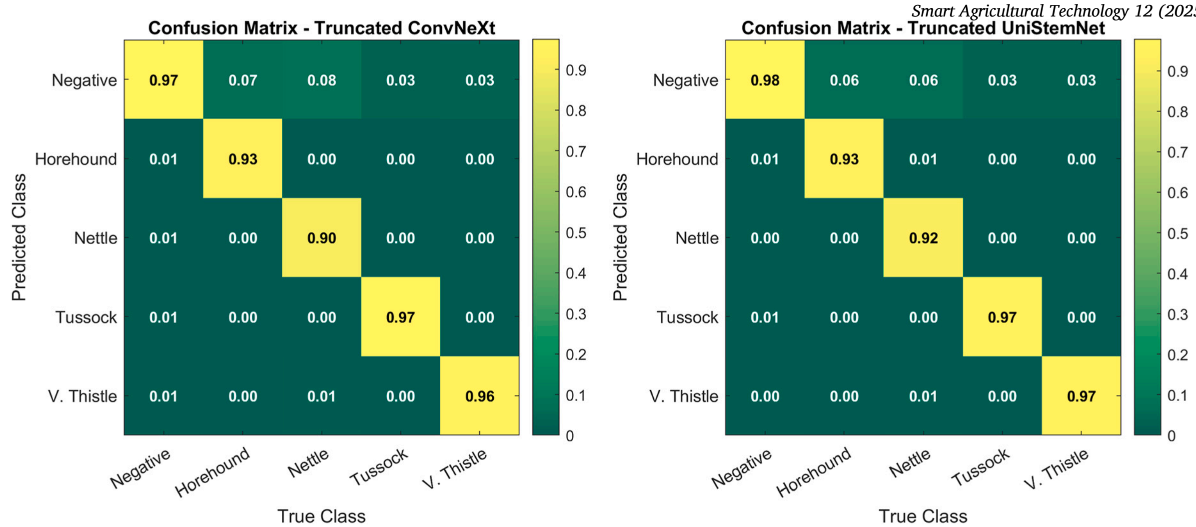

Fig. 7. Normalized confusion matrices for the plant segmentation task using the truncated ConvNeXt and truncated UniStemNet models. Results are shown for the pooled test with CutMix augmentation.

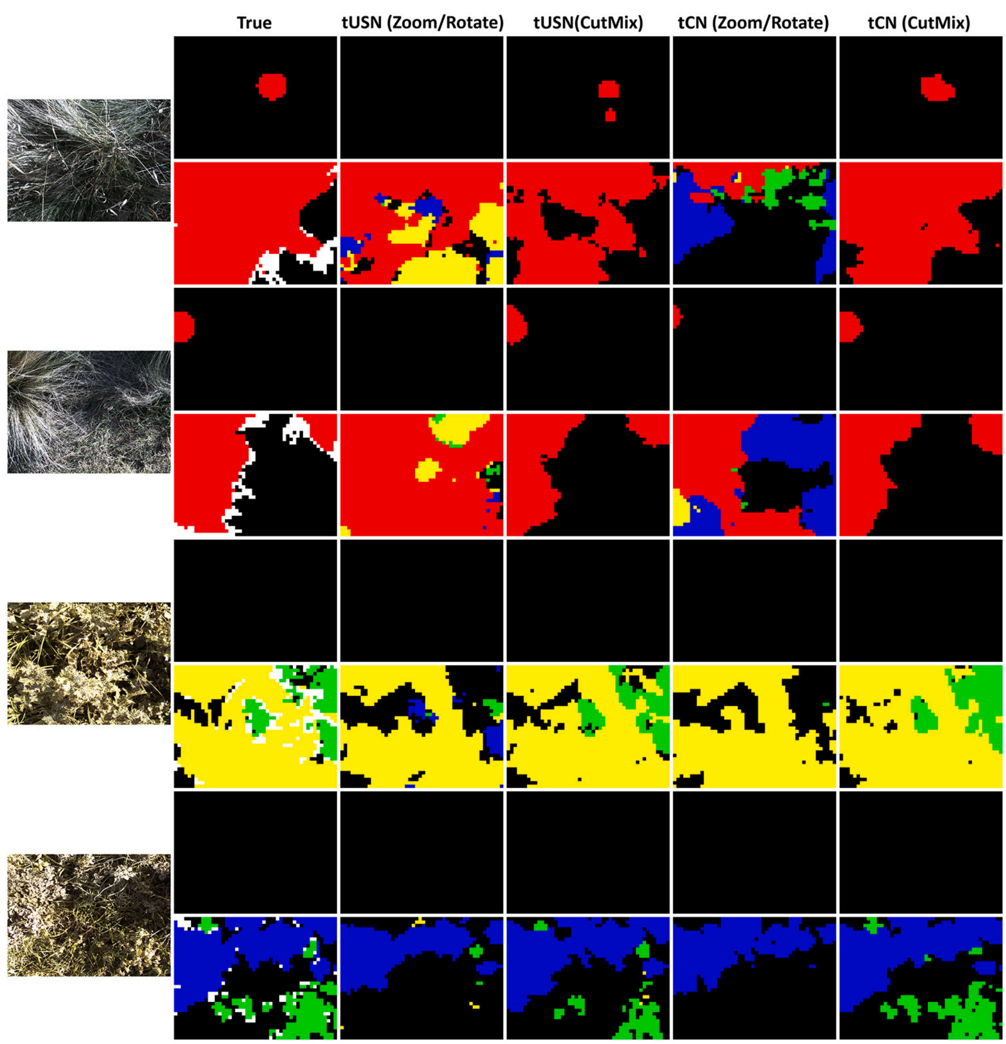

Fig. 8. Example outputs from condition invariance test one using default normalization. Left to right: input image, ground truth masks, and predicted masks under different model and augmentation settings. For each output, the spraypoint mask is shown above the plant segmentation mask. Colour legend—Black: Negative, Red: Tussock, Blue: Horehound, Yellow: Variegated Thistle, Green: Nettle, White: Difficult.

The choice of normalization technique has a significant impact on model performance, especially when evaluating condition invariance. However, the optimal normalization method appears to depend on the specific training regimen, highlighting the need for further research to better understand the factors that influence normalization effectiveness.

In the context of weed detection in Australian pastures, this study has demonstrated the potential of CNN models. However, further research is necessary to assess their effectiveness across a wider range of weed species that better represent the diversity of these environments. In particular, future work is needed to evaluate the models' ability to distinguish between morphologically similar long-bladed grass species, which may be either invasive weeds or beneficial native grasses.

This study further highlights the effectiveness of joint plant and spraypoint CNN detectors in multi-species environments. However, achieving robust performance under varying growing conditions requires careful consideration of normalization methods and data augmentation strategies. In this context, CutMix augmentation proved particularly useful, outperforming simpler approaches such as zooms and rotations. As a result, future weed segmentation studies, and in particular those involving test images from diverse locations or time points, should consider using CutMix augmentation during training. Moreover, while this study focused on joint plant segmentation and spraypoint detection, the augmentation strategies developed are equally applicable to single-task problems, which are more commonly addressed in the literature.

CRediT authorship contribution statement

Jonathan Ford: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Edmund Sadgrove:** Writing – review & editing, Supervision, Methodology, Conceptualization. **David Paul:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used chatGPT in order to improve the readability of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Shyam Prasad Adhikari, Heechan Yang, Hyongsuk Kim, Learning semantic graphics using convolutional encoder–decoder network for autonomous weeding in paddy, Front. Plant Sci. 10 (2019).

[2] Anderson Brilhador, Matheus Gutoski, Leandro Takeshi Hattori, Andrei de Souza Inácio, André Eugênio Lazzaretti, Heitor Silvério Lopes, Classification of weeds and crops at the pixel-level using convolutional neural networks and data augmentation, in: 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE, 2019, pp. 1–6.

[3] Brendan Calvert, Alex Olsen, Bronson Philippa, Mostafa Rahimi Azghadi, Autoweed: detecting harrisia cactus in the goondiwindi region for selective spot-spraying, in: Proceedings of the 1st Queensland Pest Animal and Weed Symposium, Weed Society of Queensland Pty. Ltd., Toowoomba, QLD, Australia, 2019, p. 52.

[4] Chengtai Cao, Fan Zhou, Yurou Dai, Jianping Wang, Kunpeng Zhang, A survey of mix-based data augmentation: taxonomy, methods, applications, and explainability, ACM Comput. Surv. 57 (2) (2024) 1–38.

[5] Alessandro dos Santos Ferreira, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, Marcelo Theophilo Folhes, Weed detection in soybean crops using ConvNets, Comput. Electron. Agric. 143 (2017) 314–324.

[6] Jonathan Ford, Edmund Sadgrove, David Paul, Joint plant-spraypoint detector with convnext modules and histmatch normalization, Precis. Agric. 26 (1) (2025) 24.

[7] A.S.M. Mahmudul Hasan, Ferdous Sohel, Dean Diepeveen, Hamid Laga, Michael G.K. Jones, Weed recognition using deep learning techniques on class-imbalanced imagery, Crop Pasture Sci. (2022) 628–644.

[8] Kun Hu, Zhiyong Wang, Guy Coleman, Asher Bender, Tingting Yao, Shan Zeng, Dezhen Song, Arnold Schumann, Michael Walsh, Deep learning techniques for in-crop weed recognition in large-scale grain production systems: a review, Precis. Agric. 25 (1) (2024) 1–29.

[9] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.

[10] Philipp Lottes, Jens Behley, Nived Chebrolu, Andres Milioto, Cyrill Stachniss, Joint stem detection and crop-weed classification for plant-specific treatment in precision farming, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 8233–8238.

[11] Philipp Lottes, Jens Behley, Nived Chebrolu, Andres Milioto, Cyrill Stachniss, Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming, J. Field Robot. 37 (1) (2020) 20–34.

[12] Wouter H. Maes, Kathy Steppe, Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture, Trends Plant Sci. 24 (2) (2019) 152–164.

[13] S. Imran Moazzam, Umar S. Khan, Waqar S. Qureshi, Tahir Nawaz, Faraz Kunwar, Towards automated weed detection through two-stage semantic segmentation of tobacco and weed pixels in aerial imagery, Smart Agric. Technol. 4 (2023) 100142.

[14] Alex Olsen, Dmitry A. Konovalov, Bronson Philippa, Peter Ridd, Jake C. Wood, Jamie Johns, Wesley Banks, Benjamin Girgenti, Owen Kenny, James Whinney, Brendan Calvert, Mostafa Rahimi Azghadi, Ronald D. White, DeepWeeds: a multiclass weed species image dataset for deep learning, Sci. Rep. 9 (1) (2019) 2058.

[15] Md Atiqur Rahman, Yang Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, in: International Symposium on Visual Computing, Springer, 2016, pp. 234–244.

[16] T. Sarvini, T. Sneha, G.S. Sukanya Gowthami, S. Sushmitha, et al., Performance comparison of weed detection algorithms, in: 2019 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2019, pp. 843–847.

[17] Søren Kelstrup Skovsen, Morten Stigaard Laursen, Rebekka Kjeldgaard Kristensen, Jim Rasmussen, Mads Dyrmann, Jørgen Eriksen, René Gislum, Rasmus Nyholm Jørgensen, Henrik Karstoft, Robust species distribution mapping of crop mixtures using color images and convolutional neural networks, Sensors 21 (1) (2020) 175.

[18] Lyndon N. Smith, Arlo Byrne, Mark F. Hansen, Wenhao Zhang, Melvyn L. Smith, Weed classification in grasslands using convolutional neural networks, in: Michael E. Zelinski, Tarek M. Taha, Jonathan Howe, Abdul A. Awwal, Khan M. Iftekharuddin (Eds.), Applications of Machine Learning, SPIE, San Diego, California, 2019, p. 42.

[19] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo, Cutmix: regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6023–6032.

[20] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, mixup: beyond empirical risk minimization, in: Proceedings of the Sixth International Conference on Learning Representations, 2018.

[21] Xiaoguang Zhang, Nan Li, Luzhen Ge, Xuan Xia, Ning Ding, A unified model for real-time crop recognition and stem localization exploiting cross-task feature fusion, in: 2020 IEEE International Conference on Real-Time Computing and Robotics (RCAR), 2020, pp. 327–332.