Weed Science

www.cambridge.org/wsc

Knowledge distillation and student–teacher learning for weed detection in turf

Danlan Zhai[1,*], Teng Liu[2,*], Feiyu He[3], Jinxu Wang[2], Xiaojun Jin[4] and Jialin Yu[5]

[1]Intern, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China; [2]Research Assistant, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China; [3]Student, Department of Computer Science, Duke University, Durham, NC, USA; [4]Associate Professor, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China and [5]Professor and Principal Investigator, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China

Research Article

Cite this article: Zhai D, Liu T, He F, Wang J, Jin X, Yu J (2024) Knowledge distillation and student–teacher learning for weed detection in turf. Weed Sci. **72**: 804–815. doi: 10.1017/wsc.2024.85

Received: 20 May 2024
Revised: 21 September 2024
Accepted: 12 October 2024
First published online: 29 October 2024

Keywords:
Deep learning; neural networks; turf management; precision herbicide application

Corresponding authors: Xiaojun Jin; Email: xiaojun.jin@pku-iaas.edu.cn; Jialin Yu; Email: jialin.yu@pku-iaas.edu.cn

*These authors contributed equally to this work.

Abstract

Machine vision–based herbicide applications relying on object detection or image classification deep convolutional neural networks (DCNNs) demand high memory and computational resources, resulting in lengthy inference times. To tackle these challenges, this study assessed the effectiveness of three teacher models, each trained on datasets of varying sizes, including D-20k (comprising 10,000 true-positive and true-negative images) and D-10k (comprising 5,000 true-positive and true-negative images). Additionally, knowledge distillation was performed on their corresponding student models across a range of temperature settings. After the process of student–teacher learning, the parameters of all student models were reduced. ResNet18 not only achieved higher accuracy (ACC $\geq$ 0.989) but also maintained higher frames per second (FPS $\geq$ 742.9) under its optimal temperature condition ($T = 1$). Overall, the results suggest that employing knowledge distillation in the machine vision models enabled accurate and reliable weed detection in turf while reducing the need for extensive computational resources, thereby facilitating real-time weed detection and contributing to the development of smart, machine vision–based sprayers.

Introduction

Turfgrass is widely grown in urban landscapes, including athletic, commercial, and residential lawns; golf courses; roadsides; and parks (Pincetl et al. 2019). Turf provides various advantages, including evaporative cooling in urban areas, soil remediation, atmospheric pollutant absorption, and beautifying residential and nonresidential landscapes (El-Haggar and Samaha 2019; Stier et al. 2013). Nevertheless, weed competition is a severe constraint for turf management. Weeds compete with turfgrasses for environmental resources such as sunlight, moisture, and soil nutrients (Hamuda et al. 2016; Liu and Bruch 2020), reducing turf aesthetics and functionality (Monteiro 2017; Pincetl et al. 2019). Weed management in turfgrass landscapes traditionally relied heavily on broadcast herbicide application (McCullough et al. 2015; McElroy and Martins 2013), although weeds almost always present in nonuniform and patchy distributions (Dai et al. 2019; Yu et al. 2019a), leading to herbicide application on areas where weeds do not occur. The excessive use of synthetic herbicides poses a potential risk to human health and may result in environmental pollution (Alengebawy et al. 2021; Hasanuzzaman et al. 2020; Mennan et al. 2020; Yu et al. 2019b). For example, atrazine, a photosystem II inhibitor, is commonly used in warm-season turfgrasses, yet it is frequently detected in groundwater (Yu and McCullough 2016). Consequently, it has been classified as a restricted-use pesticide in the United States (USEPA 2023). Manual spot spraying of herbicide can reduce herbicide input but is time-consuming and labor-intensive, and thus is impractical for large landscape areas (Kakarla et al. 2022).

Machine vision–based precision herbicide application technologies offer a viable solution to minimize herbicide use and weed control costs (Jin et al. 2022b; Partel et al. 2019; Shuping et al. 2023; Upadhyay et al. 2024a, 2024b). Traditional machine learning methods analyze plant imagery, considering factors such as color (Tang et al. 2016), morphology (Perez et al. 2000), textural traits (Bakhshipour et al. 2017), and hyper- or multispectral features (Jiang et al. 2020; Pantazi et al. 2016), for the purpose of identifying target weeds or distinguishing between crops and weeds. Nevertheless, detecting and differentiating weeds within crops is inherently difficult due to their resemblances in color and morphology (Al-Badri et al. 2022; Hasan et al. 2021).

In recent years, improvements in graphics processing unit (GPU) computing capabilities have greatly advanced the development of deep convolutional neural networks (DCNNs)

© The Author(s), 2024. Published by Cambridge University Press on behalf of Weed Science Society of America.

(Krichen 2023; Tulbure et al. 2022). Many innovative concepts, such as activation functions, parameter optimization, model size, and inference architecture, have been explored to further enhance the performance of DCNNs (Khanam et al. 2024). DCNNs have shown impressive capabilities in weed detection within turf environments, as demonstrated in recent research (Jin et al. 2024; Xie et al. 2021; Yu et al. 2019b). For instance, research elucidated the efficacy of employing object detection (DetectNet) and image classification neural networks (including AlexNet, GoogLeNet, and VGGNet) to detect weeds in bermudagrass [*Cynodon dactylon* (L.) Pers.] and perennial ryegrass (*Lolium perenne* L.) turfgrasses. The findings highlighted that image classification neural networks excelled in detecting images containing broadleaf and grassy weeds within turfgrass (Jin et al. 2022a; Yu et al. 2019a, 2019c, 2020). Nevertheless, deep learning–based methods for weed detection in turf enhance accuracy at the expense of increasing computational load and decreasing detection speed, which limits their practical application. Many studies have demonstrated that models developed on high-performance computers often have excessive parameters, which complicates efficient inference on terminal devices (Chen and Ran 2019; El-Rashidy et al. 2020; Shakarami et al. 2021; Yang et al. 2022). Consequently, it is fairly challenging to develop alternative approaches for identifying weeds in turf while balancing the model's accuracy with real-time weed detection.

Knowledge distillation is a contemporary neural network technique aimed at diminishing neural network size while maintaining or enhancing performance (Hinton et al. 2015). In knowledge distillation, a small student model is typically supervised by a large teacher model (Ba and Caruana 2014; Buciluǎ et al. 2006; Hinton et al. 2015; Urban et al. 2017). Basically, a knowledge distillation system consists of three fundamental components: knowledge, a distillation algorithm, and teacher–student architecture. It involves the student model emulating the teacher model to achieve competitive or even superior performance. Existing distillation algorithms typically use a fixed temperature as a hyperparameter in the softmax layer to control the smoothness of the distribution and accurately determine the difficulty level of the loss minimization process (Li et al. 2023). In recent years, there has been growing research in the utilization of knowledge distillation in the domain of agriculture, such as crop segmentation (Angarano et al. 2023), fruit and vegetable defect detection (Cai et al. 2024; Nithya et al. 2022; Zhou et al. 2023), and plant leaf segmentation (Jung et al. 2022).

Knowledge distillation offers a method for converting complex weed detection models into lightweight versions, facilitating deployment on resource-constrained mobile or embedded devices without sacrificing performance. This research hypothesized that applying knowledge distillation to turf weed detection could enhance weed detection performance while optimizing the use of limited time and computational resources, thereby improving the efficiency of developing effective neural network models. This approach shows significant promise for smart weeding robots, boosting their capability for real-time and precise herbicide application. A key factor in knowledge distillation is the hyper-parameter temperature (T), which plays a crucial role in balancing the knowledge transfer between the teacher and student models. Therefore, the objectives of this research were to (1) assess the performance of three teacher models in detecting weeds in turf across datasets of different scales, (2) compare the results of three student models at different temperatures after knowledge distillation to determine their respective optimal temperatures, and (3) evaluate three student models individually at their respective optimal temperatures to identify the most suitable model for practical application.

Materials and Methods

Dataset

The experimental images in this research were captured at different times from various turf landscapes containing diverse weed species. Some images were captured in spring 2021 using a Panasonic® digital camera (model DMC-ZS110) at two distinct locations in China: sod farms in Jiangning District, Nanjing City, Jiangsu Province, China (31.95°N, 118.85°E) and sod farms in Shuyang, Jiangsu Province, China (34.12°N, 118.79°E), while others were obtained in autumn 2018 using a SONY® Cyber-Shot Digital Still Camera (model DSC-HX1) from two separate locations in the United States: the University of Georgia Griffin Campus in Griffin, GA, USA (33.26°N, 84.28°W), and multiple golf courses in Peachtree City, GA, USA (33.39°N, 84.59°W). The turf species in these locations was bermudagrass, and the most commonly observed weed species were dallisgrass (*Paspalum dilatatum* Poir.), dandelion (*Taraxacum officinale* F.H. Wigg. ssp. *officinale*), doveweed [*Murdannia nudiflora* (L.) Brenan], Florida pusley (*Richardia scabra* L.), lawn pennywort (*Hydrocotyle sibthorpioides* Lam.), old world diamond flower (*Oldenlandia corymbosa* L.), purple nutsedge (*Cyperus rotundus* L.), smooth crabgrass [*Digitaria ischaemum* (Schreb.) Schreb. ex Muhl.], and white clover (*Trifolium repens* L.). The camera was configured in automatic mode for parameters such as exposure, focus, and white balance. Images were captured at a height that resulted in a ground-sampling distance of 0.05 cm pixel^{-1}, under varying lighting conditions, including clear, cloudy, and partially cloudy weather. All the images were taken in a 16:9 ratio, with a resolution of 1,920 by 1,080 pixels.

The raw images were initially cropped to dimensions of 240 by 240 pixels using Irfanview (v. 5.50, Irfan Ski jan, Jajce, Bosnia). As demonstrated in Figure 1, each resulting image block was then categorized into one of two classes: "weed," representing sub-images containing weeds, and "turf," representing sub-images without weeds. As shown in Table 1, 5,000 images per class were selected to create the small training dataset D-10k, and each class was expanded to include 10,000 images for the large training dataset D-20k. Additionally, an additional 500 images per class were set aside for the validation dataset, while another 500 images per class were allocated for the testing dataset.

Neural Network Models

In the context of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)-2015, the champion ResNet (He et al. 2016) introduced "residual blocks" to address issues of gradient vanishing and declining training set accuracy in deep neural networks. Instead of attempting to directly learn the complete underlying mapping from inputs to outputs, these blocks enable the network to focus on learning the difference (residual) between the input and the desired output. This architectural innovation shifts the network's task from fitting the entire low-level mapping to modeling the residual in relation to the original network, thereby significantly reducing training complexity.

DenseNet (Huang et al. 2017) shares a similar goal of overcoming the challenge of training deep neural networks by incorporating "skip connections" or "shortcut connections." It strongly emphasizes "dense connectivity," where every layer is densely connected to every other layer, creating a tightly interconnected network structure. This architectural approach promotes extensive feature reuse across the network, thereby

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

Table 1. Training, validation, and testing dataset specifications.[a]

	Training		Validation		Testing	
	Class		Class		Class	
Dataset[b]	Turf	Weed	Turf	Weed	Turf	Weed
	image quantity					
D-20k	10,000	10,000	500	500	500	500
D-10k	5,000	5,000	500	500	500	500

[a]Images at a resolution of 240 × 240 pixels were used for training, validation, and testing.
[b]D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.

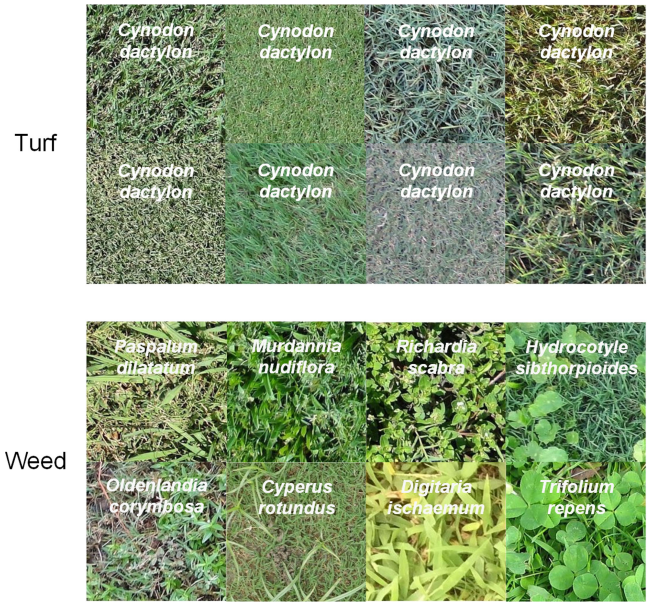


Figure 1. Representation of the two classes in the training, validation, and testing datasets. "Turf" refers to sub-images that exclusively contain bermudagrass (*Cynodon dactylon*). "Weed" refers to sub-images that contain one of the following species: *Paspalum dilatatum*, *Taraxacum officinale*, *Murdannia nudiflora*, *Richardia scabra*, *Hydrocotyle sibthorpioides*, *Oldenlandia corymbosa*, *Cyperus rotundus*, *Digitaria ischaemum*, or *Trifolium repens*. Only sub-images containing a single weed species were used for training, validation, and testing.

facilitating feature propagation and gradient flow throughout the entire model, ultimately contributing to more effective training of deep networks.

EfficientNet (Tan and Le 2019) is a set of eight convolutional neural network models ranging from B0 to B7. EfficientNet achieves more efficient results through uniform scaling of depth, width, and resolution while shrinking the model size. The initial phase of compound scaling involves a grid search to determine the relationships among different scaling dimensions of the baseline network under fixed resource constraints. Subsequently, appropriate scaling factors are determined and applied to scale the baseline network to the target network. The primary building block of EfficientNet is the MBConv module, consisting of a layer that first expands and then compresses channels, utilizing depth-wise separable convolutions to competently reduce the number of parameters.

Knowledge Distillation

Complex models often possess an extended parameter space, enhancing performance and generalization capabilities. Knowledge distillation (Hinton et al. 2012) leverages the knowledge acquired by complex models to guide the training of smaller models, thereby compensating for the limited expressive capacity imposed by the smaller scale of the latter. This leads to an improvement in the performance of smaller models.

Assuming a reliable teacher model is accessible, the student model can calculate the probability of each category output from the teacher model, denoted as the "soft label." In contrast, the actual image labels are considered to be the "hard label." Classification models typically utilize a softmax layer to compute the probability of each output category. The formula for this calculation is as follows, where q_i represents the output probability of class i, and z_i represents the output logit of class i.

$$q_i = \frac{\exp(z_i)}{\sum_i \exp(z_i)} \quad [1]$$

Using the softmax output of the teacher model directly as the soft label is not a practical approach. This is because when the entropy for the probability distribution of the softmax output is low, the probability of the negative category label tends to be close to 0, and as a result, its contribution to the loss function becomes negligible. Therefore, a new variable called "temperature" can be introduced, and the softmax function can be calculated using the following formula, where T represents the temperature.

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad [2]$$

After introducing the temperature factor T, the soft targets produced by the softmax classifier largely preserve the probability relationships between different sample classifications.

The application of knowledge distillation requires both a teacher model and a student model, and the final loss is composed of the cross-entropy functions of both models, calculated through linear weighting. The training process is depicted in Figure 2. The soft loss can mitigate the overfitting of hard labels by student models (Cho and Hariharan 2019), and the final loss function is represented as Equation 3, where p_j represents the output of the teacher model; p_i represents the output of the student model; y denotes the true label; CE stands for the cross-entropy function; and λ is the hyperparameter that adjusts the weighting of the loss function.

$$\text{Loss} = \lambda \text{CE}(y, p_i) + (1 - \lambda)\text{CE}(p_j, p_i) \quad [3]$$

Experimental Environment and Procedure

This study examined three teacher models: ResNet101, DenseNet201, and EfficientNetB5; and three student models: ResNet18, DenseNet121, and EfficientNetB0. Compared with the teacher models, student models had shallower and less complex architectures. Using ResNet as an example, the subsequent numerical values indicate the diverse depths of various models. ResNet101 has a greater depth, consisting of 101 convolutional layers, while ResNet18 is relatively shallower, with only 18 convolutional layers (He et al. 2016).

In this study, a total of 48 image classification neural networks were trained and tested, comprising 8 teacher models and 40 student models with varying temperatures and structures. Initially, the weights of the teacher models, which were pretrained

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

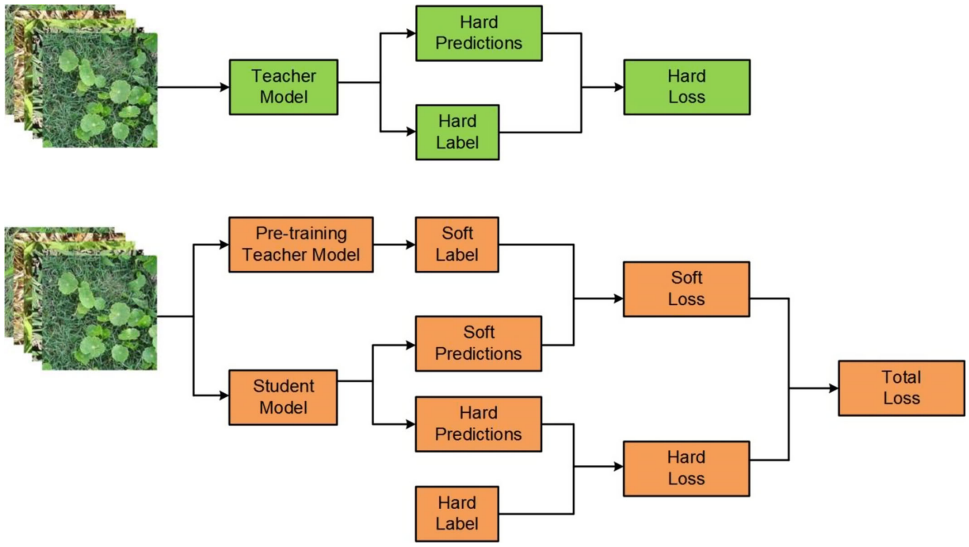


Figure 2. Flowchart of the knowledge distillation training process.

on the ImageNet dataset (Deng et al. 2009), were delivered to their corresponding model architectures using transfer learning. Three teacher models were subjected to fine-tuning to adjust their fully connected layer outputs for binary classification. To assess their performance, each teacher model was independently trained on two datasets of varying sizes: the larger dataset, designated as D-20k, and the smaller dataset, referred to as D-10k. This separate training approach allowed for a comparative analysis of the model's performance across datasets of different scales. Subsequently, a knowledge distillation approach was employed to transfer the acquired knowledge to lightweight student models. The validation set accuracy and model stability were compared under different temperature settings to determine the optimal teacher and student models. The hyperparameters used for training in different experimental setups are presented in Table 2. All models were trained and tested on the open-source PyTorch deep learning framework (v. 1.8.1, Facebook, San Jose, CA, USA), which was installed on a workstation equipped with a GeForce RTX 3080 Ti GPU (NVIDIA) and 64 GB of memory.

Evaluation

For both teacher and student image classification neural networks, the assessment results were organized in a binary classification confusion matrix encompassing four outcomes: a true positive (TP), a true negative (TN), a false positive (FP), and a false negative (FN). TP indicates the count of correctly predicted weed-free samples, whereas TN represents the count of correctly predicted samples with weeds. TP and TN are indicators reflecting the true condition of weeds. Conversely, in cases where samples are actually infested with weeds, FP signifies the incorrect prediction of samples as weed-free, and FN signifies the incorrect prediction of samples as infested with weeds. TP and TN expose instances of prediction errors in recognizing the weed condition. The performances of the neural networks were evaluated using Accuracy (ACC), precision, recall, F1 score, and the Matthews' correlation coefficient (MCC) via confusion matrices (Sokolova and Lapalme 2009).

ACC measures the percentage of accurately classified samples within a specified dataset and was calculated using the following formula:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad [4]$$

Precision measures the ability of the model to accurately detect the target and was computed using the following formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad [5]$$

Recall measures the effectiveness of the neural network to correctly identify the target and was defined using the following formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad [6]$$

F1 score measures the overall performance of the neural network and represents the harmonic mean of precision and recall, which was determined using the following formula:

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [7]$$

MCC is a metric to quantify a predictive model's performance quality. It provides a more balanced assessment, yielding values between −1 and 1. A score of 1 indicates a perfect prediction, 0 represents random predictions, and −1 signals total disagreement between the model and actual outcomes. It was calculated using the following formula:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad [8]$$

Furthermore, frames per second (FPS) is a critical metric in the realm of computer graphics technology. It measures the number of individual images processed and predicted by a neural network model in a single second (Stewart et al. 2021). Higher FPS values result in faster image classification speeds, indicating more robust real-time processing performance.

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

Table 2. Hyperparameter values used for training the teacher models.

Deep learning architecture	Optimizer[a]	Base learning rate	Weight decay	Batch size	Learning rate policy	Momentum	Epochs
ResNet101	SGD	0.1	0.0001	32	step	0.9	100
EfficientNetB5	SGD	0.1	0.0001	16	step	0.9	100
DenseNet201	SGD	0.1	0.0001	32	step	0.9	100

[a]SGD, stochastic gradient descent.

Table 3. Validation results of teacher models for weed detection in bermudagrass turf.

Neural network	Training dataset[a]	ACC[b]	Precision	Recall	F1 score	MCC[c]
Resnet101	D-20k	0.979	0.984	0.974	0.979	0.958
	D-10k	0.974	0.984	0.964	0.974	0.948
EfficientNetB5	D-20k	0.989	0.986	0.992	0.989	0.978
	D-10k	0.987	0.986	0.988	0.987	0.974
DenseNet201	D-20k	0.982	0.988	0.976	0.982	0.964
	D-10k	0.976	0.978	0.974	0.976	0.952

[a]D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.
[b]ACC, accuracy.
[c]MCC, Matthews' correlation coefficient.

Table 4. Testing results of teacher models for weed detection in turf.

Neural network	Training dataset[a]	ACC[b]	Precision	Recall	F1 score	MCC[c]	Model size	FPS[d]
Resnet101	D-20k	0.987	0.994	0.98	0.987	0.974	340.8MB	564.2
	D-10k	0.987	0.994	0.98	0.987	0.974	340.8MB	554.5
EfficientNetB5	D-20k	0.993	0.998	0.988	0.993	0.986	227.9MB	380.1
	D-10k	0.993	0.998	0.988	0.993	0.986	227.9MB	375.2
DenseNet201	D-20k	0.99	1.0	0.98	0.99	0.98	146.3MB	505.6
	D-10k	0.983	0.996	0.97	0.983	0.966	146.3MB	511.4

[a]D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.
[b]ACC, accuracy.
[c]MCC, Matthews' correlation coefficient.
[d]FPS, frames per second.

Finally, model size was utilized as an important metric for comparing the parameter scale and complexity level of the teacher and student models.

Results and Discussion

Teacher Model Performance

In the present study, the three teacher models were trained on the D-20k and D-10k datasets, and the performance of weed detection was evaluated using the same validation and testing dataset, as shown in Tables 3 and 4. Additionally, Figure 3 illustrates the confusion matrices of teacher models on the testing dataset, providing a more detailed presentation of the model's classification outcomes. In general, the performances of weed detection neural networks exhibited a minor improvement on the testing dataset relative to the validation dataset. Specifically, following training on the dataset D-20k, the accuracy for ResNet101, EfficientNetB5, and DenseNet201 showed increases of 0.8%, 0.4%, and 0.8%, respectively, when evaluated on the testing dataset. Similarly, when trained on the dataset D-10k, the models demonstrated accuracy improvement of 1.3%, 0.6%, and 0.7%, respectively. The three distinct teacher models maintained consistently exceptional performance, achieving ACC values of 0.974 or higher in distinguishing between turf and weeds. This outcome could be attributed to the characteristics of the dataset. Empirical observations suggested that dataset D-10k contained a sufficiently diverse array of images, thereby facilitating the effective adaptation of the models to the data.

For the teacher model ResNet101, training on both datasets resulted in no significant differences in performance metrics. The ACC, precision, recall, F1 score, and MCC values of ResNet101 were consistent across the D-20k and D-10k datasets. Notably, the F1 score of ResNet101 on the testing dataset reached 0.987, marginally lower by 0.6% compared with that of EfficientNetB5 evaluated on the same testing dataset. These findings suggested that ResNet101 is less sensitive to dataset size and remains robust even with limited training data. Furthermore, ResNet101 exhibited a pronounced advantage in processing speed, operating at 554.5 FPS, compared with EfficientNetB5's 375.2 FPS, in identifying and distinguishing sub-images with weeds, demonstrating an approximate 1.48-fold increase in processing speed. In summary, while EfficientNetB5 achieved the highest accuracy and F1 score, ResNet101 exhibited a significantly higher FPS, demonstrating superior processing speed. This trade-off between accuracy and speed is critical for real-time applications such as precision herbicide spraying, where timely detection is essential.

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

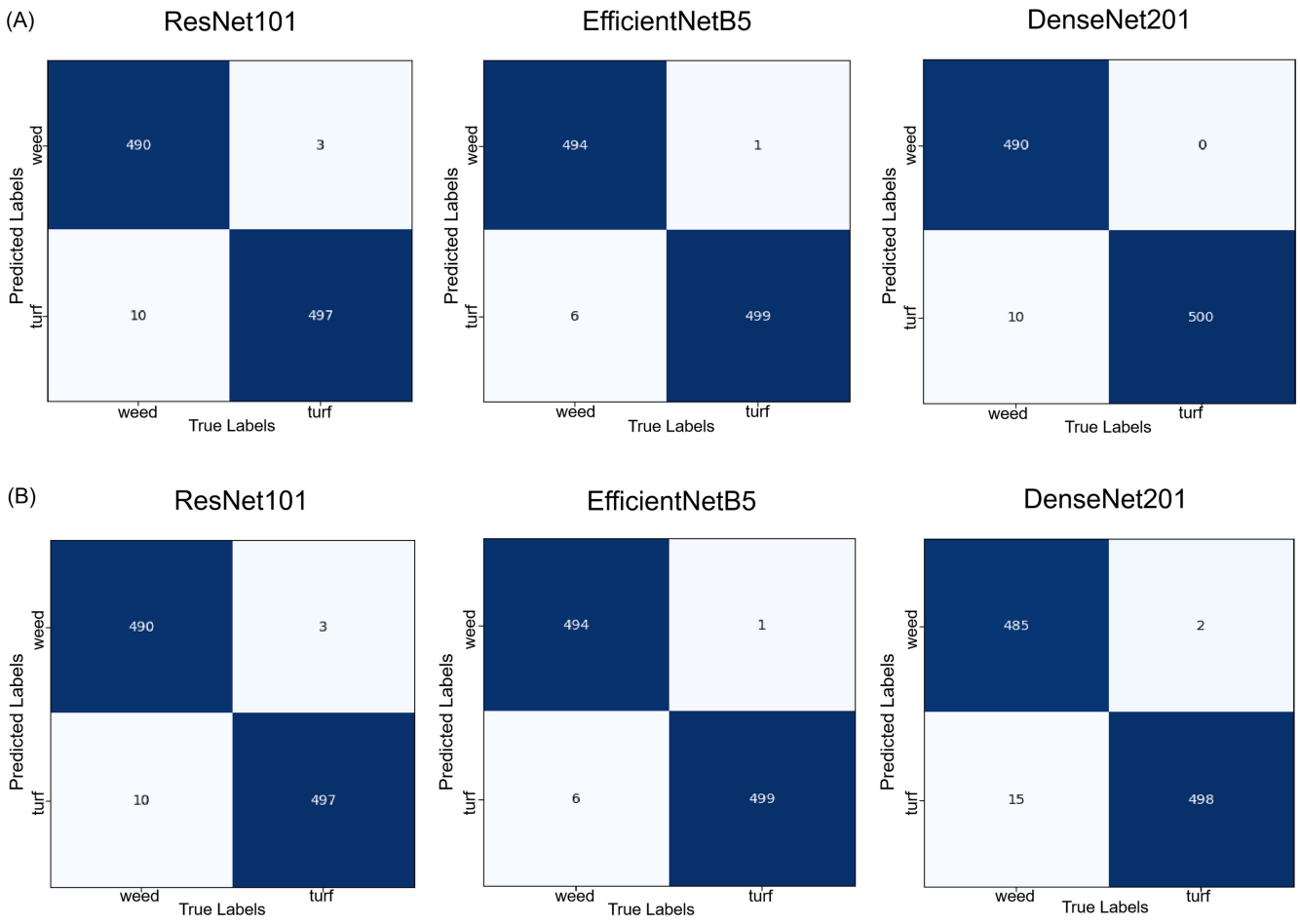


Figure 3. Confusion matrices of teacher models on the testing dataset. Training of teacher models on (A) the D-20k dataset and (B) the D-10k dataset. D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.

Figure 4 depicts the progression of ACC for each teacher model trained on datasets of varying sizes. Initially, all models demonstrated high ACC levels as a result of their pretraining on the ImageNet database (Deng et al. 2009), which endowed them with substantial generalization capabilities. Upon initiation of the learning process and adjustment of their weights, a rapid increase in the ACC curve was observed during the initial stage. Subsequently, after 30 epochs, the models exhibited marginal fluctuations before stabilizing at a level exceeding 95% accuracy. Notably, across both datasets D-20k and D-10k, the ACC values of the teacher models consistently surpassed 80%, with certain models even exceeding 98%.

Comparing Figure 4A and 4B, it is evident that the teacher model ResNet101 exhibited a consistently rising trajectory in its ACC curve when trained on both D-20k and D-10k datasets. Additionally, after 100 training epochs, the EfficientNetB5 model showed a 0.7% and 1% increase in ACC over the DenseNet121 and ResNet101 models, respectively, on the dataset D-20k. When trained on the D-10k dataset, the EfficientNet model showed an increase in ACC by 1.1% and 1.3% compared with DenseNet121 and ResNet101, respectively. During the stable period, the ACC curve of EfficientNetB5 consistently outperformed those of the DenseNet and ResNet models, potentially owing to its utilization of compound scaling. This method can effectively balance the model's width, depth, and resolution, thereby maximizing the utilization of computational resources.

Student Model Performance

The outcomes of student models on the validation dataset after knowledge distillation under different T settings are illustrated in Table 5. Significantly, applying temperature was intended to balance soft and hard target losses (Cho and Hariharan 2019). The optimal temperature for knowledge distillation was found to be 1 for the ResNet, DenseNet, and EfficientNet models, on both datasets D-20k and D-10k. These findings suggested that these models possess inherent complexity alongside robust generalization abilities, thereby enabling them to successfully perform classification tasks on turf–weed datasets. Therefore, there is no need for additional temperature factor adjustments in the student models to balance the capability and complexity. In certain instances, elevated temperature settings may even have a negative impact on model performance (Wei et al. 2022).

The optimal temperature of 1 indicated that during the process of knowledge distillation, knowledge transfer between the teacher and student models occurred with a high degree of confidence. The student models endeavored to precisely replicate the prediction probability distribution of the teacher models, rigorously adhering to the decisions made by the teacher models. Overall, the optimal temperature of 1 observed across all three models signified that the student models effectively inherited and leveraged the knowledge from the teacher models, facilitating model deployment and application.

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

Table 5. Validation results of student models across different temperature settings.

Neural network	Training dataset[a]	Temperature	ACC[b]	Precision	Recall	F1 score	MCC[c]
ResNet18	D-20k	1	0.986	0.99	0.982	0.986	0.972
		2	0.984	0.992	0.976	0.984	0.968
		3	0.982	0.988	0.976	0.982	0.964
		4	0.98	0.988	0.972	0.98	0.96
		5	0.974	0.988	0.96	0.974	0.948
EfficientNetB0	D-20k	1	0.983	0.988	0.978	0.983	0.966
		2	0.982	0.992	0.972	0.982	0.964
		3	0.972	0.978	0.966	0.972	0.944
		4	0.976	0.986	0.966	0.976	0.952
		5	0.98	0.986	0.974	0.98	0.96
DenseNet121	D-20k	1	0.983	0.988	0.978	0.983	0.966
		2	0.979	0.986	0.972	0.979	0.958
		3	0.975	0.982	0.968	0.975	0.95
		4	0.977	0.988	0.966	0.977	0.954
		5	0.981	0.992	0.97	0.981	0.962
ResNet18	D-10k	1	0.981	0.986	0.976	0.981	0.962
		2	0.974	0.988	0.96	0.974	0.948
		3	0.97	0.978	0.962	0.97	0.94
		4	0.968	0.981	0.954	0.967	0.936
		5	0.964	0.979	0.948	0.963	0.928
EfficientNetB0	D-10k	1	0.979	0.988	0.97	0.979	0.958
		2	0.974	0.986	0.962	0.974	0.948
		3	0.971	0.98	0.962	0.971	0.942
		4	0.966	0.989	0.942	0.965	0.933
		5	0.968	0.988	0.948	0.968	0.937
DenseNet	D-10k	1	0.972	0.974	0.97	0.972	0.944
		2	0.967	0.976	0.958	0.967	0.934
		3	0.971	0.984	0.958	0.971	0.942
		4	0.954	0.958	0.95	0.954	0.908
		5	0.957	0.977	0.936	0.956	0.915

[a]D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.
[b]ACC, accuracy.
[c]MCC, Matthews' correlation coefficient.

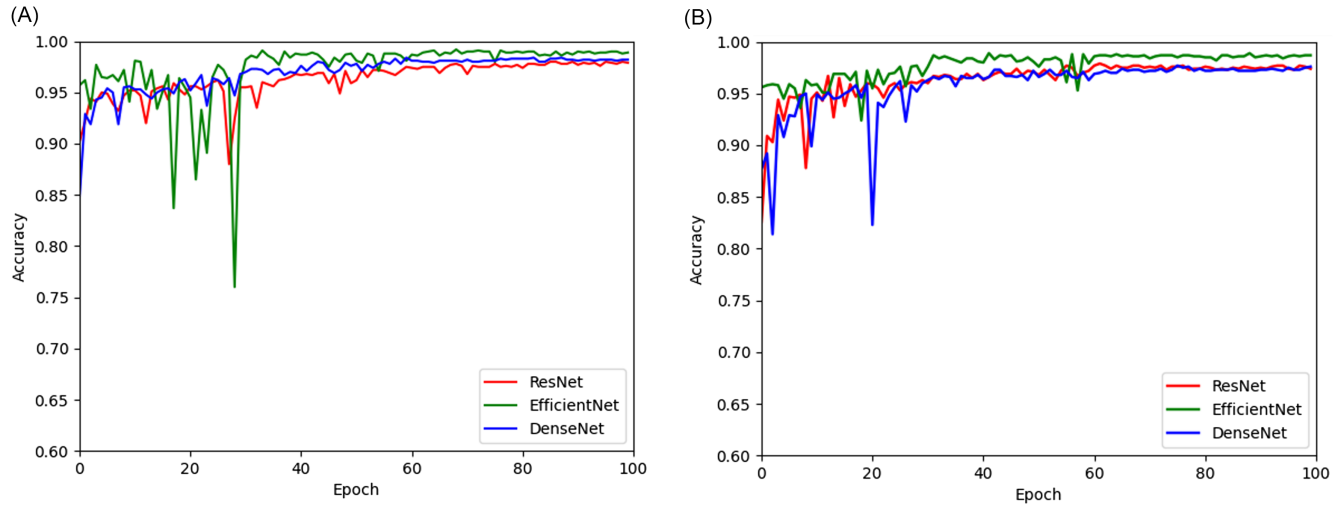


Figure 4. Fluctuations in accuracy (ACC) values during the training of teacher models on datasets of different sizes. Training of teacher models on (A) the D-20k dataset and (B) the D-10k dataset. D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.

The fluctuations in ACC values during the training process of student models are depicted in Figure 5. At the same time, the teacher models (represented by the red lines) are also compared with relevant student models at different T settings. Considering Figure 5 from the perspective of varying T settings for the same model, it can be observed that under conditions $T = 4$ and $T = 5$, each student model exhibited notable fluctuations during the initial stages of training. Upon entering the stabilization phase, the ACC metrics were observed to be lower compared with those under alternative temperature settings. Additionally, the ResNet model displayed the highest initial ACC values, exceeding 0.85 with minimal curve fluctuation, indicating its superior efficacy in knowledge distillation.

The evaluation metrics of each student model at the corresponding optimal temperature on the testing dataset are

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

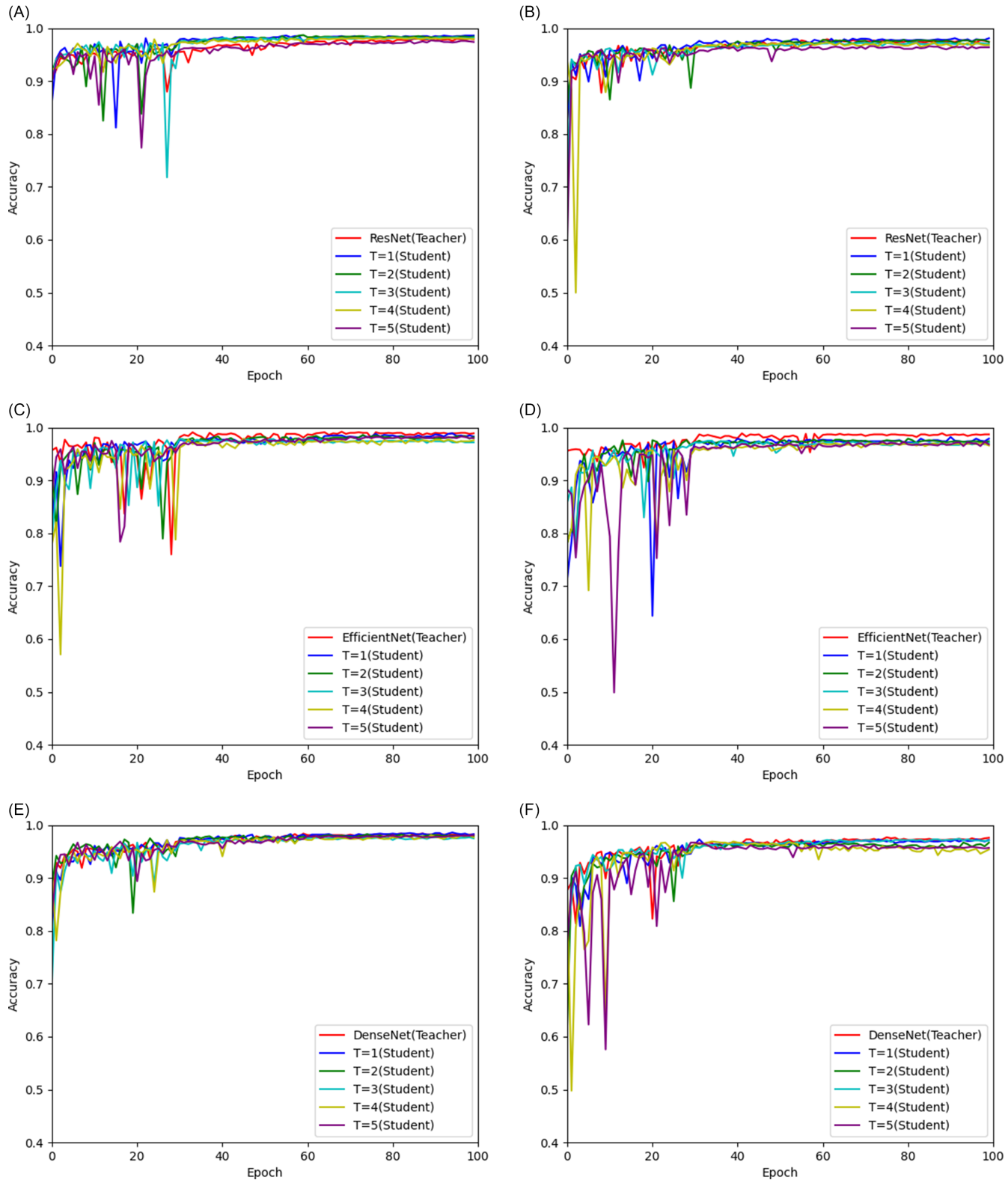


Figure 5. Fluctuations in accuracy (ACC) values during the training of student models across different *T* settings and sizes of datasets. Student–teacher learning with (A) the ResNet model on the D-20k dataset; (B) ResNet model on the D-10k dataset; (C) the EfficientNet model on the D-20k dataset; (D) the EfficientNet model on the D-10k dataset; (E) the DenseNet model on the D-20k dataset; and (F) the DenseNet model on the D-10k dataset. D-20k indicates the training dataset with 10,000 “Turf” class images and 10,000 “Weed” class images. D-10k indicates the training dataset with 5,000 “Turf” class images and 5,000 “Weed” class images.

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

Table 6. Testing results of student models at their optimal temperature for weed detection in turf.

Neural network	Training dataset[a]	Optimal temperature	ACC[b]	Precision	Recall	F1 score	MCC[c]	Model size	FPS[d]
Resnet18	D-20k	1	0.991	1.0	0.982	0.991	0.982	260.2 MB	742.9
	D-10k	1	0.989	1.0	0.978	0.989	0.978	260.2 MB	762.4
EfficientNetB0	D-20k	1	0.982	0.998	0.966	0.982	0.964	146.9 MB	696.8
	D-10k	1	0.987	0.998	0.976	0.987	0.974	146.9 MB	690.7
DenseNet121	D-20k	1	0.985	1.0	0.97	0.985	0.97	130.1 MB	580.3
	D-10k	1	0.98	0.988	0.972	0.98	0.96	130.1 MB	562.3

[a]D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.
[b]ACC, accuracy.
[c]MCC, Matthews' correlation coefficient.
[d]FPS, frames per second.

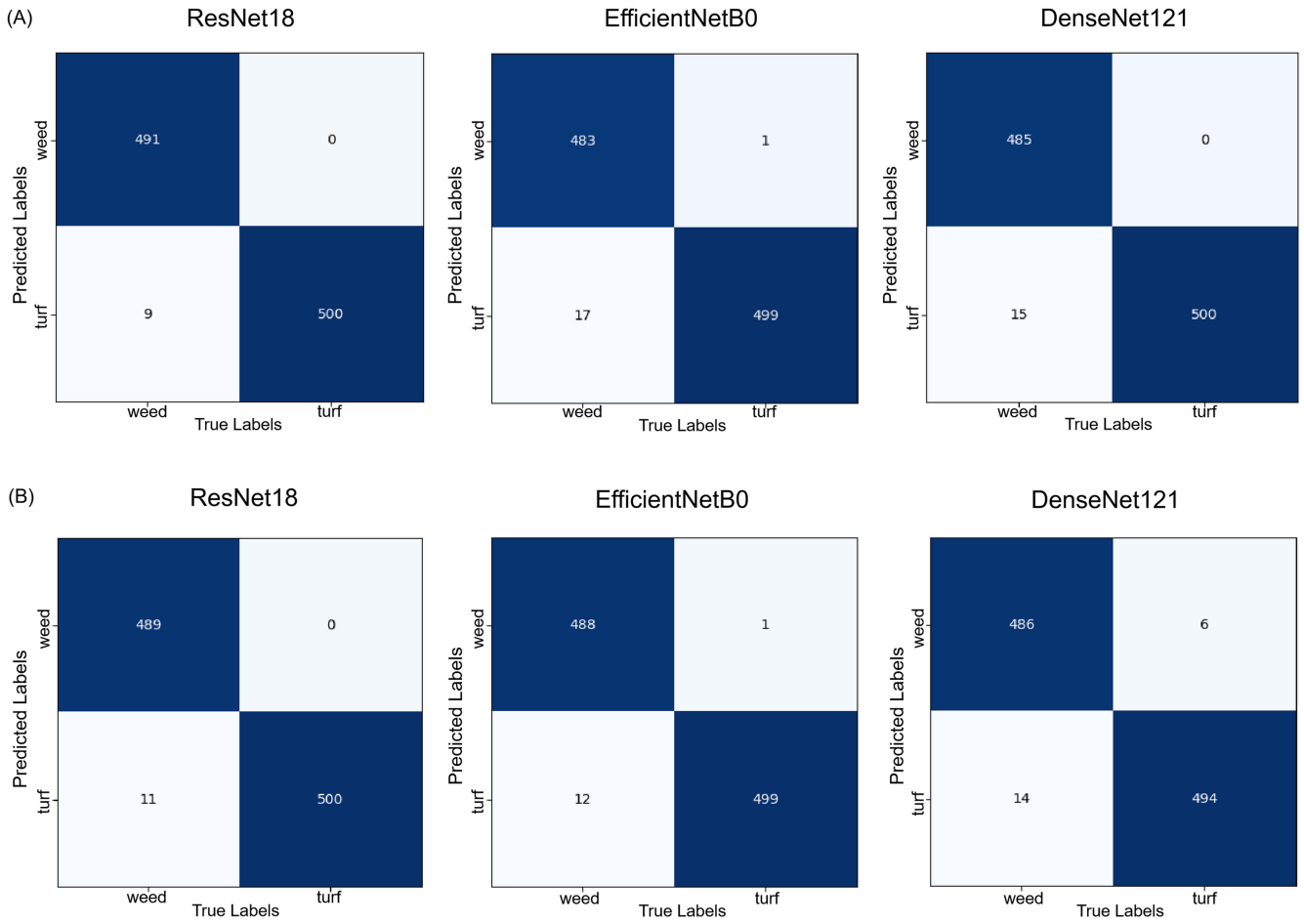


Figure 6. Confusion matrices of student models at their optimal temperature on the testing dataset. Training of student models on (A) the D-20k dataset and (B) the D-10k dataset. D-20k indicates the training dataset with 10,000 "Turf" class images and 10,000 "Weed" class images. D-10k indicates the training dataset with 5,000 "Turf" class images and 5,000 "Weed" class images.

documented in Table 6. Additionally, the confusion matrices for these models on the testing dataset are presented in Figure 6. It can be observed that the primary cause for model errors was the incorrect classification of sub-images belonging to the "weed" category as those devoid of weeds, with only minimal instances of cases where the sub-images containing turf only were erroneously identified as containing weeds. The result shows that the student models could reliably detect weeds growing on turf.

A comparative analysis between the data presented in Tables 4 and 6 reveals that the teacher model, ResNet101, had a model size of 340.8 MB. Conversely, following knowledge distillation, the student model ResNet18 had a reduced size of 260.2 MB. Additionally, after knowledge distillation, the EfficientNet model's size decreased from 227.9 MB to 146.9 MB, and the DenseNet model's size decreased from 146.3 MB to 130.1 MB. In all three cases, the model size was reduced, indicating a reduction in model parameters and a decrease in model complexity, demonstrating the effectiveness of knowledge distillation.

Relative to the teacher models, ResNet18, EfficientNetB0, and DenseNet121 exhibited a substantial increase in FPS on D-10k,

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

with enhancements of 207.9, 315.5, and 55.9, respectively. This suggests improvements in model light-weighting, enhanced image processing speed, and improved computational efficacy. ResNet18 still had the highest FPS at 762.4, signifying faster inference rates and surpassing the other neural networks in real-time classification. Moreover, among the three student models evaluated at their respective optimal temperatures, ResNet18 exhibited superior performance. On the large dataset D-20k, the ACC, F1 score, and MCC values for ResNet18 were 0.991, 0.991, and 0.982, respectively. These values exceeded those of EfficientNetB0 by 0.9%, 0.9%, and 1.8% and surpassed the values for DenseNet121 by 0.6%, 1.1%, and 2.2%, respectively. On the small dataset D-10k, the ACC, F1 score, and MCC values for ResNet18 were 0.989, 0.989, and 0.978, respectively. These values surpassed those of EfficientNetB0 by 0.2%, 0.2%, and 0.4% and exceeded those of DenseNet121 by 0.9%, 0.9%, and 1.8%, respectively. Overall, the distilled student model, ResNet18, achieved a balance between ACC and efficiency, which was more appropriate for the binary classification task of turf–weed images.

Considering our results in comparison with other studies, Ghofrani and Toroghi (2022) leveraged the knowledge distillation technique to improve the accuracy of a small client-side model in plant disease recognition, achieving a 97.58% ACC. Similarly, Wei et al. (2022) applied knowledge distillation to the neural network training process, resulting in a 98.7% ACC on the Oxford102 flower dataset. On the other hand, Zhou et al. (2023) developed a surface defect detection system for carrot (*Daucus carota* L.) combine harvesting based on multistage knowledge distillation, achieving an accuracy of only 90.7%. In contrast, the distilled student model used in this research, ResNet18, demonstrated competitive capabilities with an ACC of 98.9%. This model effectively balanced ACC and efficiency, making it particularly well suited for the binary classification task of turf–weed images.

Various herbicides, such as synthetic auxins (e.g., 2,4-D, dicamba, and MCPP) (McElroy and Martins 2013; Reed et al. 2013), acetyl-CoA carboxylase inhibitors (e.g., clethodim, sethoxydim, and fenoxaprop-*P*-ethyl) (McCullough et al. 2016; Tate et al. 2021), as well as protoporphyrinogen oxidase inhibitors (e.g., sulfentrazone) (Brosnan et al. 2020; Yu et al. 2018) are used for weed control in turf. Precise application can significantly reduce herbicide usage, lowering costs and mitigating adverse environmental impacts. In a recent study, Jin et al. (2023) evaluated a smart sprayer prototype designed for precision herbicide application in turf. Significantly, their study employed DCNN models, which incorporated a vast array of parameters, and succeeded in achieving an F1 score exceeding 0.989 in identifying weeds in turf. However, the research did not extend to the real-time implementation of precision spraying, primarily due to the slow inference speed.

To the best of our knowledge, no previous research has explored the impact of knowledge distillation on the development of lightweight and efficient weed detection models. In the present study, our results suggest that the knowledge distillation approach from teacher models to student models offers three key advantages: (1) feasibility with a relatively small training dataset containing 5,000 images per class, (2) a substantial reduction in model size, and (3) the capability for real-time weed detection. Further research is needed to evaluate the performance of employing knowledge-distilled models in the machine vision subsystem of smart sprayers for real-time weed detection and precision herbicide spraying in turfgrass landscapes.

In summary, knowledge distillation can achieve superior weed detection performance in turf while balancing accuracy and efficiency. All three teacher models displayed no significant difference on different scales of datasets, including both the D-20k and the D-10k. Each of the three student models achieved the best performance when $T = 1$, indicating their reliable identification capabilities for weed detection in turf. Moreover, ResNet18 achieved higher ACC of ≥0.989 and MCC values of ≥0.978 and maintained higher FPS rates of ≥742.9. Both ACC and FPS metrics are essential in real-world scenarios for achieving accurate and efficient weed detection. Therefore, we conclude that the ResNet18 model delivered superior results and was better suited for succeeding deployment on resource-constrained devices, although the EfficientNetB0 and DenseNet121 models had smaller sizes. Compared with DCNN models with substantial computational workload, knowledge distillation can reduce model size and delay through teacher–student learning, thereby facilitating real-time weed detection and precision spraying. Additional research is ongoing to optimize the distillation algorithm and match different model structures for teacher–student models.

Funding statement. This work was supported by the Key R&D Program of Shandong Province, China (ZR202211070163), the National Natural Science Foundation of China (Grant No. 32072498), the Taishan Scholar Program, and the Weifang Science and Technology Development Plan Project (Grant No. 2024ZJ1097).

Competing interests. The authors declare no conflicts of interest.

References

Al-Badri AH, Ismail NA, Al-Dulaimi K, Salman GA, Khan A, Al-Sabaawi A, Salam MSH (2022) Classification of weed using machine learning techniques: a review—challenges, current and future potential techniques. J Plant Dis Prot 129:745–768

Alengebawy A, Abdelkhalek ST, Qureshi SR, Wang MQ (2021) Heavy metals and pesticides toxicity in agricultural soil and plants: ecological risks and human health implications. Toxics 9:42

Angarano S, Martini M, Navone A, Chiaberge M (2023) Domain generalization for crop segmentation with knowledge distillation. arXiv:2304.01029

Ba J, Caruana R (2014) Do deep nets really need to be deep? Pages 2654–2662 *in* Proceedings of the 28th Annual Conference on Neural Information Processing Systems. New York: Curran Associates

Bakhshipour A, Jafari A, Nassiri SM, Zare D (2017) Weed segmentation using texture features extracted from wavelet sub-images. Biosyst Eng 157:1–12

Brosnan JT, Elmore MT, Bagavathiannan MV (2020) Herbicide-resistant weeds in turfgrass: current status and emerging threats. Weed Technol 34:424–430

Buciluǎ C, Caruana R, Niculescu-Mizil A (2006) Model compression. Pages 535–541 *in* Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery

Cai X, Zhu Y, Liu S, Yu Z, Xu Y (2024) FastSegFormer: a knowledge distillation-based method for real-time semantic segmentation of surface defects in navel oranges. Comput Electron Agric 217:108604

Chen J, Ran X (2019) Deep learning with edge computing: a review. Proc IEEE 107:1655–1674

Cho JH, Hariharan B (2019) On the efficacy of knowledge distillation. Pages 4794–4802 *in* Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Washington, DC: IEEE Computer Society

Dai X, Xu Y, Zheng J, Song H (2019) Analysis of the variability of pesticide concentration downstream of inline mixers for direct nozzle injection systems. Biosyst Eng 180:59–69

Deng J, Dong W, Socher R, Li LJ, Li K, Li FF (2009) Imagenet: a large-scale hierarchical image database. Pages 248–255 *in* Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: Institute of Electrical and Electronics Engineers

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

El-Haggar S, Samaha A (2019) Sustainable urban community development guidelines. Pages 75–102 *in* Amer M, ed. Roadmap for Global Sustainability—Rise of the Green Communities. Cham, Switzerland: Springer

El-Rashidy N, El-Sappagh S, Islam SR, El-Bakry HM, Abdelrazek S (2020) End-to-end deep learning framework for coronavirus (COVID-19) detection and monitoring. Electronics 9:1439

Ghofrani A, Toroghi RM (2022) Knowledge distillation in plant disease recognition. Neural Comput Appl 34:14287–14296

Hamuda E, Glavin M, Jones E (2016) A survey of image processing techniques for plant extraction and segmentation in the field. Comput Electron Agric 125:184–199

Hasan AM, Sohel F, Diepeveen D, Laga H, Jones MG (2021) A survey of deep learning techniques for weed detection from images. Comput Electron Agric 184:106067

Hasanuzzaman M, Mohsin SM, Bhuyan MB, Bhuiyan TF, Anee TI, Masud AAC, Nahar K (2020) Phytotoxicity, environmental and health hazards of herbicides: challenges and ways forward. Pages 55–99 *in* Prasad MNV, ed. Agrochemicals: Detection, Treatment and Remediation. Oxford, UK: Elsevier

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Pages 770–778 *in* Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: Institute of Electrical and Electronics Engineers

Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Proc Mag 29:82–97

Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. arXiv:1503.02531 [stat.ML]

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. Pages 4700–4708 *in* Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: Institute of Electrical and Electronics Engineers

Jiang H, Jiang X, Ru Y, Wang J, Xu L, Zhou H (2020) Application of hyperspectral imaging for detecting and visualizing leaf lard adulteration in minced pork. Infrared Phys Technol 110:103467

Jin X, Bagavathiannan M, McCullough PE, Chen Y, Yu J (2022a) A deep learning-based method for classification, detection, and localization of weeds in turfgrass. Pest Manag Sci 78:4809–4821

Jin X, Han K, Zhao H, Wang Y, Chen Y, Yu J (2024) Detection and coverage estimation of purple nutsedge in turf with image classification neural networks. Pest Manag Sci 80:3504–3515

Jin X, Liu T, Chen Y, Yu J (2022b) Deep learning-based weed detection in turf: a review. Agronomy 12:3051

Jin X, Liu T, Yang Z, Xie J, Bagavathiannan M, Hong X, Xu Z, Chen X, Yu J, Chen Y (2023) Precision weed control using a smart sprayer in dormant bermudagrass turf. Crop Prot 172:106302

Jung JY, Lee SH, Kim JO (2022) Plant leaf segmentation using knowledge distillation. Pages 1–3 *in* Proceedings of the 2022 IEEE International Conference on Consumer Electronics—Asia. Piscataway, NJ: Institute of Electrical and Electronics Engineers

Kakarla SC, Costa L, Ampatzidis Y, Zhang Z (2022) Applications of UAVs and machine learning in agriculture. Pages 1–19 *in* Zhang Z, Liu H, Yang C, Ampatzidis Y, Zhou J, Jiang Y, eds. Unmanned Aerial Systems in Precision Agriculture. Singapore: Springer

Khanam R, Hussain M, Hill R, Allen P (2024) A comprehensive review of convolutional neural networks for defect detection in industrial applications. IEEE Access 12:94250–94295

Krichen M (2023) Convolutional neural networks: a survey. Computers 12:151

Li Z, Li X, Yang L, Zhao B, Song R, Luo L, Li J, Yang J (2023) Curriculum temperature for knowledge distillation. Pages 1504–1512 *in* Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, DC: Association for the Advancement of Artificial Intelligence Press

Liu B, Bruch R (2020) Weed detection for selective spraying: a review. Curr Robot Rep 1:19–26

McCullough PE, Yu J, Raymer PL, Chen Z (2016) First report of ACCase-resistant goosegrass (*Eleusine indica*) in the United States. Weed Sci 64: 399–408

McCullough PE, Yu J, Shilling DG, Czarnota MA, Johnston CR (2015) Biochemical effects of imazapic on bermudagrass growth regulation, broomsedge (*Andropogon virginicus*) control, and MSMA antagonism. Weed Sci 63:596–603

McElroy J, Martins D (2013) Use of herbicides on turfgrass. Planta Daninha 31:455–467

Mennan H, Jabran K, Zandstra BH, Pala F (2020) Non-chemical weed management in vegetables by using cover crops: a review. Agronomy 10:257

Monteiro JA (2017) Ecosystem services from turfgrass landscapes. Urban For Urban Green 26:151–157

Nithya R, Santhi B, Manikandan R, Rahimi M, Gandomi AH (2022) Computer vision system for mango fruit defect detection using deep convolutional neural network. Foods 11:3483

Pantazi XE, Moshou D, Bravo C (2016) Active learning system for weed species recognition based on hyperspectral sensing. Biosyst Eng 146:193–202

Partel V, Kakarla SC, Ampatzidis Y (2019) Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. Comput Electron Agric 157:339–350

Perez A, Lopez F, Benlloch J, Christensen S (2000) Colour and shape analysis techniques for weed detection in cereal fields. Comput Electron Agric 25:197–212

Pincetl S, Gillespie TW, Pataki DE, Porse E, Jia S, Kidera E, Nobles N, Rodriguez J, Choi DA (2019) Evaluating the effects of turf-replacement programs in Los Angeles. Landscape Urban Plan 185:210–221

Reed TV, Yu J, McCullough PE (2013) Aminocyclopyrachlor efficacy for controlling Virginia buttonweed (*Diodia virginiana*) and smooth crabgrass (*Digitaria ischaemum*) in tall fescue. Weed Technol 27:488–491

Shakarami A, Shahidinejad A, Ghobaei-Arani M (2021) An autonomous computation offloading strategy in mobile edge computing: a deep learning-based hybrid approach. J Netw Comput Appl 178:102974

Shuping F, Yu R, Chenming H, Fengbo Y (2023) Planning of takeoff/landing site location, dispatch route, and spraying route for a pesticide application helicopter. Eur J Agron 146:126814

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inform Process Manag 45:427–437

Stewart R, Nowlan A, Bacchus P, Ducasse Q, Komendantskaya E (2021) Optimising hardware accelerated neural networks with quantisation and a knowledge distillation evolutionary algorithm. Electronics 10:396

Stier JC, Steinke K, Ervin EH, Higginson FR, McMaugh PE (2013) Turfgrass benefits and issues. Pages 105–145 *in* Stier JC, Horgan BP, Bonos SA, eds. Turfgrass: Biology, Use, and Management. Wiley Online Books, https://doi.org/10.2134/agronmonogr56

Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. Pages 6105–6114 *in* Proceedings of the 36th International Conference on Machine Learning. Maastricht, Netherlands: ML Research Press

Tang JL, Chen XQ, Miao RH, Wang D (2016) Weed detection using image processing under different illumination for site-specific areas spraying. Comput Electron Agric 122:103–111

Tate TM, McCullough PE, Harrison ML, Chen Z, Raymer PL (2021) Characterization of mutations conferring inherent resistance to acetyl coenzyme A carboxylase-inhibiting herbicides in turfgrass and grassy weeds. Crop Sci 61:3164–3178

Tulbure AA, Tulbure AA, Dulf EH (2022) A review on modern defect detection models using DCNNs–deep convolutional neural networks. J Adv Res 35:33–48

Upadhyay A, Sunil GC, Zhang Y, Koparan C, Sun X (2024a) Development and evaluation of a machine vision and deep learning-based smart sprayer system for site-specific weed management in row crops: an edge computing approach. J Agric Food Res 18:101331

Upadhyay A, Zhang Y, Koparan C, Rai N, Howatt K, Bajwa S, Sun X (2024b) Advances in ground robotic technologies for site-specific weed management in precision agriculture: a review. Comput Electron Agric 225:109363

Urban G, Geras KJ, Kahou SE, Aslan O, Wang S, Caruana R, Mohamed A, Philipose M, Richardson M (2017) Do deep convolutional nets really need to be deep and convolutional? arXiv:1603.05691 [stat.ML]

[USEPA] U.S. Environmental Protection Agency (2023) Ingredients Used in Pesticide Products—Atrazine. http://www.epa.gov/ingredients-used-pesticide-products/atrazine. Accessed: March 10, 2024

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press

Wei X, Zhang H, Shi C, Yang X, Han H, Li B (2022) A lightweight flower classification model based on improved knowledge distillation. Pages 2236–2239 *in* Proceedings of the IEEE 10th Joint International Information Technology and Artificial Intelligence Conference. Piscataway, NJ: Institute of Electrical and Electronics Engineers

Xie S, Hu C, Bagavathiannan M, Song D (2021) Toward robotic weed control: detection of nutsedge weed in bermudagrass turf using inaccurate and insufficient training data. IEEE Robot Autom Lett 6:7365–7372

Yang G, Wang B, Qiao S, Qu L, Han N, Yuan G, Li H, Wu T, Peng Y (2022) Distilled and filtered deep neural networks for real-time object detection in edge computing. Neurocomputing 505:225–237

Yu J, McCullough PE (2016) Triclopyr reduces foliar bleaching from mesotrione and enhances efficacy for smooth crabgrass control by altering uptake and translocation. Weed Technol 30:516–523

Yu J, McCullough PE, Czarnota MA (2018) Annual bluegrass (*Poa annua*) biotypes exhibit differential levels of susceptibility and biochemical responses to protoporphyrinogen oxidase inhibitors. Weed Sci 66: 574–580

Yu J, Schumann AW, Cao Z, Sharpe SM, Boyd NS (2019a) Weed detection in perennial ryegrass with deep learning convolutional neural network. Front Plant Sci 10:1422

Yu J, Schumann AW, Sharpe SM, Li X, Boyd NS (2020) Detection of grassy weeds in bermudagrass with deep convolutional neural networks. Weed Sci 68:545–552

Yu J, Sharpe SM, Schumann AW, Boyd NS (2019b) Deep learning for image-based weed detection in turfgrass. Eur J Agron 104:78–84

Yu J, Sharpe SM, Schumann AW, Boyd NS (2019c) Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. Pest Manag Sci 75:2211–2218

Zhou W, Song C, Song K, Wen N, Sun X, Gao P (2023) Surface defect detection system for carrot combine harvest based on multi-stage knowledge distillation. Foods 12:793

https://doi.org/10.1017/wsc.2024.85 Published online by Cambridge University Press