Journal of Plant Diseases and Protection (2024) 131:227–236
https://doi.org/10.1007/s41348-023-00814-9

SHORT COMMUNICATION

A comparative analysis of deep learning methods for weed classification of high-resolution UAV images

Pendar Alirezazadeh[1] · Michael Schirrmann[1] · Frieder Stolzenburg[2]

Received: 30 March 2023 / Accepted: 28 September 2023 / Published online: 16 October 2023

Abstract

Because weeds compete directly with crops for moisture, nutrients, space, and sunlight, their monitoring and control is an essential necessity in agriculture. The most important step in choosing an effective and time-saving weed control method is the detection of weed species. Deep learning approaches have been proven to be effective in smart agricultural tasks such as plant classification and disease detection. The performance of Deep Learning-based classification models is often influenced by the complexity of the feature extraction backbone. The limited availability of data in weed classification problems poses a challenge when increasing the number of parameters in the backbone of a model. While a substantial increase in backbone parameters may only result in marginal performance improvements, it can also lead to overfitting and increased training difficulty. In this study, we aim to explore the impact of adjusting the architecture depth and width on the performance of deep neural networks for weed classification using Unmanned Aerial Vehicles (UAV) imagery. Specifically, we focus on comparing the performance of well-known convolutional neural networks with varying levels of complexity, including heavy and light architectures. By investigating the impact of scaling deep layers, we seek to understand how it influences attention mechanisms, enhances the learning of meaningful representations, and ultimately improves the performance of deep networks in weed classification tasks with UAV images. Data were collected using a high-resolution camera on a UAV flying at low altitudes over a winter wheat field. Using the transfer learning strategy, we trained deep learning models and performed species-level classification tasks with the weed species: *Lithospermum arvense, Spergula arvensis, Stellaria media, Chenopodium album,* and *Lamium purpureum* observed in that field. The results obtained from this study reveal that networks with deeper layers do not effectively learn meaningful representations, thereby hindering the expected performance gain in the context of the specific weed classification task addressed in this study.

Keywords Deep learning · Transfer learning · Transformers · Unmanned aerial vehicles (UAVs) · Weed classification

Introduction

Weeds can be found scattered throughout the field, posing a challenge as they compete with crops for essential resources like water, nutrients, space, and sunlight. If not effectively managed, this competition can negatively impact crop yield and quality. To tackle the weed problem, various methods have been employed, including manual, mechanical, and chemical weeding. Manual weeding, performed by hand or with basic tools, has been practiced for centuries, especially in small fields (Bakhshipour et al. 2017). However, this approach is labor-intensive, inefficient, and incurs high labor costs, making it unsuitable for modern weed control practices.

In contrast to manual weeding, mechanical methods offer higher efficiency and labor-saving benefits. However, without a targeting module, they may struggle to remove weeds within crop rows and risk causing crop damage (Hamuda et al. 2016; Wang et al. 2019).

In modern agriculture, chemical herbicides play a vital role in effectively managing weeds, safeguarding crop yields, and enhancing overall farming efficiency. These herbicides act as crucial tools in reducing competition

✉ Pendar Alirezazadeh
PAlirezazadeh@atb-potsdam.de

[1] Department of Agromechatronics, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee 100, 14469 Potsdam, Germany

[2] Automation and Computer Sciences Department, Harz University of Applied Sciences, Friedrichstr. 57-59, 38855 Wernigerode, Germany

for essential resources like water, nutrients, and sunlight, thereby fostering optimal crop growth and yielding higher agricultural productivity. Furthermore, herbicides contribute to the containment of diseases and pests that could otherwise pose significant threats to crops, consequently elevating food quality and ensuring greater safety for consumers (Pakdaman Sardrood and Mohammadi Goltapeh 2018).

Traditional weed control methods that rely on chemical herbicides involve uniformly spraying the herbicides across the entire field, irrespective of weed presence. This approach leads to increased herbicide expenses (Rodrigo et al. 2014). Additionally, the excessive utilization of herbicides in agriculture has led to significant environmental consequences. Recognizing the need to protect the environment and preserve the safety of drinking water, there is a collective agreement on the importance of reducing herbicide usage in agricultural practices (Wato et al. 2020). Given these challenges, it is imperative to adopt more environmentally friendly weed management practices. Utilizing precision agricultural technology becomes crucial in order to mitigate the adverse impacts of herbicides on the environment and optimize their usage (Osorio et al. 2020; Hasan et al. 2021).

Site-Specific Weed Management (SSWM) is a modern agricultural approach that utilizes advanced technologies, data analytics, and precision agriculture techniques to optimize weed control while reducing herbicide usage (Wang et al. 2019; de Camargo et al. 2021; Pflanz et al. 2018). SSWM relies upon a good understanding of the distribution of weeds in the field and involves tailoring weed management strategies to specific areas within a field based on weed distribution and severity. Two examples of effective methods used in SSWM are spot spraying (Hafeez et al. 2022; Villette et al. 2022) and band treatment (Loddo et al. 2019). Spot spraying involves the targeted application of herbicides to individual weed plants or small weed-infested areas. Spot spraying is particularly effective when dealing with scattered weed distributions or when weed populations are relatively low, resulting in reduced herbicide expenses and more efficient weed management. In band treatment, herbicides are applied in narrow strips along crop rows or specific zones where weeds are more likely to be present. The main goal is to target only the areas where weeds are expected to grow, while leaving weed-free zones untouched. This approach is commonly used in row crops or intercropped systems. By using herbicides in bands, farmers can minimize herbicide usage and reduce crop damage, leading to more effective and efficient weed control. SSWM involves the use of cutting-edge technology like GPS, sensors, and data analytics to identify and map weed infestations across the entire field. By considering the overall distribution and density of weeds, SSWM develops a customized weed management strategy.

In the realm of weed management, SSWM adopts a holistic approach, encompassing the entire field and devising a tailored weed management plan (Monteiro and Santos 2022; Gerhards et al. 2022). This all-encompassing technique enhances efficiency, curtails herbicide consumption, and mitigates environmental repercussions by selectively applying herbicides where they are needed. Furthermore, SSWM employs a diverse array of weed management strategies to avert the emergence of herbicide-resistant weed populations. In addition, SSWM optimizes crop protection and yields by adroitly managing weeds in a site-specific manner, offering comprehensive safeguarding across the entire field rather than merely specific zones.

One of the SSWM systems is proposed by airborne remote sensing. Airborne remote sensing uses sensors on balloons, aircraft, unmanned aerial vehicles (UAVs), and satellites equipped with advanced sensors to collect high-resolution images and data of agricultural fields. The data obtained through airborne remote sensing helps identify and map weed infestations, assess weed density, and monitor crop health in real-time or near-real-time. By analyzing the collected data, farmers and agronomists can implement targeted and precise weed control measures to optimize herbicide usage, reduce environmental impact, and enhance overall weed management efficiency.

Weed classification plays a vital role in effectively identifying and classifying weed species and weed mapping using UAV aerial imagery. Through the application of machine learning and image processing techniques, the UAV system can distinguish between various weed types and differentiate them from crops. This weed classification process offers significant advantages to farmers, enabling them to precisely target and implement weed management strategies tailored to specific weed species. Such focused approaches optimize herbicide usage, minimize environmental impact, and enhance overall efficiency and effectiveness in agricultural weed management.

The recent progress in deep learning and computer vision has brought exciting opportunities for precision weed management through weed classification. Utilizing these technologies, particularly Convolutional Neural Networks (CNNs), researchers have achieved impressive accuracy in differentiating between various weed species and crops. However, weed data presents several unique challenges, such as class imbalance, variations within and between species, occlusion and overlap with crops, and limited dataset size (Rai et al. 2023). The effectiveness of Deep Learning-based classification models is frequently affected by the complexity of the feature extraction backbone. In weed classification tasks, the scarcity of available data becomes a challenge when trying to increase the parameters in the backbone of a model. Although a significant increase in backbone parameters might only yield marginal performance enhancements, it can also result in overfitting and make the training process more difficult. These complexities make weed classification

more intricate compared to other types of classification tasks.

In this study, our main objective is to explore how adjusting the architecture depth and width impacts the performance of deep neural networks for weed classification using UAV imagery. We focus on comparing the performance of well-known CNNs with varying complexities, ranging from heavy to light architectures. By investigating the effects of scaling deep layers, we aim to understand how attention mechanisms are influenced, leading to improved learning of meaningful representations and overall better performance in weed classification tasks using UAV images. We selected two pre-trained models, ResNet, known for its efficiency and high classification accuracy, and MobileNet, a lightweight architecture, to examine different transfer learning techniques. Additionally, we explored the importance of the attention mechanism using the MobileVit architecture, and it demonstrated competitive results when trained from scratch without any pre-trained parameters. The data for our study was collected using a high-resolution camera mounted on a UAV flying at low altitudes over a winter wheat field. By employing transfer learning, we trained deep learning models and conducted species-level classification tasks with five weed species observed in that field: *Lithospermum arvense, Spergula arvensis, Stellaria media, Chenopodium album,* and *Lamium purpureum*. Our findings indicate that applying lightweight convolutional neural network architectures to limited data scenarios yields promising results for weed classification.

Materials and methods

Data collection

The study's field was located at the ATB Potsdam Marquardt experimental station (52.467913° N, 12.956533° E, 42 m above sea level) with a soil type of diluvium and a size of 7700 m^2. The mean annual air temperature and mean annual precipitation in this area are 8.3 °C and 527 mm, respectively. The crop under investigation was winter wheat (variety 'Matrix').

For data acquisition, the flying platform used was an octocopter called OktopusXL (CiS GmbH, Bentwisch, Germany), featuring an 8-propeller design and utilizing Pixhawk autopilot hardware. The platform was equipped with an HD video transmission and control system, capable of streaming control, video, and drone telemetry data via 2.4 GHz to a remote control unit within a line-of-sight of 30 km (see Fig. 1). The images were captured with an RGB-camera, namely the Sony α-6000 with a resolution of 24.7 MP. The flying height was maintained at 1.5–3 m above the ground, with a flying speed of 1 m/s, and images were taken every second. The camera was positioned in a nadir view of the canopy. The ISO and aperture were set to auto, while the exposure time was fixed at 1/1250 s. On November 8, 2021, at the beginning of the growing season, a total of 277 images were captured.

Each image has dimensions of 6000 × 4000 pixels. Each image in the dataset was annotated with weed species using bounding boxes by two annotators. The annotations were then verified by a subject matter expert to ensure accuracy

Table 1 Weed dataset details

Class label	Latin name	No. of images
CHEAL	*Chenopodium album*	3161
LAMPU	*Lamium purpureum*	1051
LITAR	*Lithospermum arvense*	2846
SPRAR	*Spergula arvensis*	725
STEME	*Stellaria media*	5982

Fig. 1 The OktopusXL flying platform utilized for data acquisition in this study. It features eight propellers and employs the Pixhawk autopilot hardware. The octocopter is equipped with an HD video transmission and control system, enabling it to stream control, video, and drone telemetry data via a 2.4 GHz connection to a remote control unit

and consistency. Following the annotation process, the coordinates (x_{min}, y_{min}, x_{max}, y_{max}) of each bounding box are utilized to crop the corresponding region from the original image, resulting in a new separate image containing only the annotated weed species. These cropped images are then saved to create a dataset specifically focused on weed species (see Fig. 2). A total of 13,765 square images from 5 weed species were obtained from the annotated images, as shown in Table 1. Figure 3 represents a sample image for each class. As you can see, due to capturing the images in an uncontrolled condition, the images have challenges like lighting, occlusion, and low resolution.

Convolutional neural network backbones

In this study, we compared three different CNNs architectures which varied in depth and number of trainable parameters. We remove the fully-connected layers of these networks and pre-train the networks with ImageNet dataset. ImageNet dataset is a very large-scale image dataset containing 1.2 million images with 1000 categories. After the pre-training, we fine-tune the convolutional and fully-connected layers with our weed image dataset. In MobileViT, we trained the model from scratch without pre-trained weights to show the importance of transformers. The following architectures were evaluated:

- *ResNets* To avoid network complexity by increasing network depth and thus reducing system performance, the Residual Network architecture (ResNet) has been proposed (Alirezazadeh et al. 2021). ResNet is composed of several residual blocks with shortcut connections between layers. ResNet has been developed with a number of different layers (i.e., 50, 101, and 152.). ResNet-50 which is composed of 26 million parameters, ResNet-101 with 44 million parameters and ResNet-152 which is deeper with 152 layers. ResNet-50 is used in some object detection frameworks such as BlitzNet and RetinaNet. ResNet-101 is used in Faster R-CNN, R-FCN, and CoupleNet, etc.
- *MobileNet* By replacing the standard convolution layers with depthwise separable convolution blocks, MobileNet is proposed as an efficient and lightweight deep neural network to be used in mobile applications (Howard et al. 2019). Unlike standard convolution layers which include 3×3 convolution layer followed by batch normalization and ReLU, in MobileNet each convolution layer split into a 3×3 depthwise convolution layer and a 1×1 pointwise convolution layer. Depthwise convolution layer filters the inputs and pointwise convolution layer combines the filtered values into a new set of outputs. We used MobileNetV2 and MobileNetV3 in our experiments. MobileNetV2 has 3.5 million parameters. MobileNetV3 is proposed in two versions small and large with 1.5 and

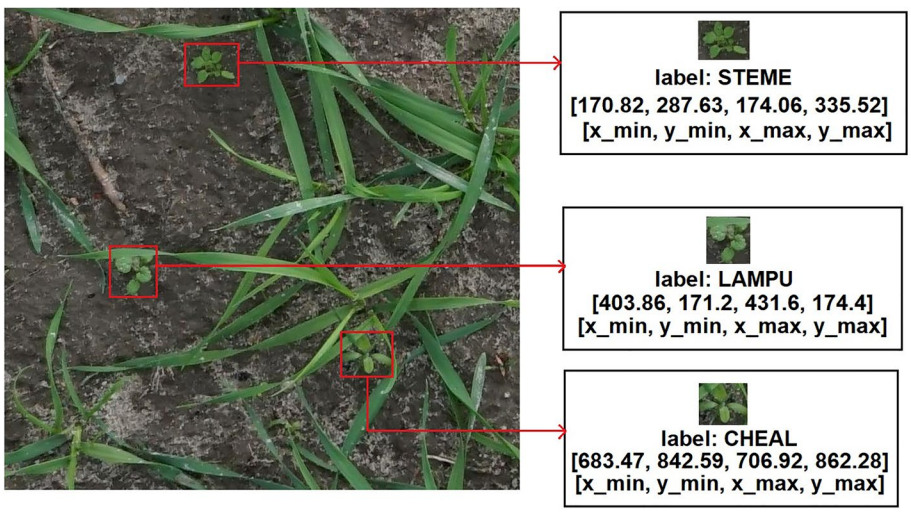

Fig. 2 The coordinates (x_{min}, y_{min}, x_{max}, y_{max}) of each bounding box are employed to crop the corresponding region from the original image to create separate images for each annotated weed species

Fig. 3 Sample images of the dataset for each class

4.2 million parameters, respectively. MobileNetV2 is used as the backbone of YoloV4-tiny detection model.

- *MobileViT* Because vision transformers (ViTs) are hefty, they are not appropriate for mobile vision activities. MobileViT is recommended for taking advantage of ViTs (Mehta and Rastegari 2021). MobileViT builds a lightweight, low-latency network by combining the strengths of CNNs and ViTs. To combine the benefits of transformers and convolutions, MobileViT offers a new layer that replaces local processing in convolutions with global processing using transformers. Long-distance dependencies are captured by transformers, resulting in global representations. Convolutions, on the other hand, can model locality by capturing spatial interactions. In addition, MobileNetV2 blocks have been used in MobileViT to speed up processing time and increase accuracy. Considering three expansion factor values for the MobileNetV2 blocks, we developed and compared three types of MobileViT: MobileViT-2, MobileViT-8, and MobileViT-8. MobileViT-2, MobileViT-8, and MobileViT-16 have 1.3, 1.5, and 1.7 million parameters, respectively. MobileViT is used as the backbone of SSD detection model.

Experimental setting

The convolutional neural networks were evaluated toward their classification accuracies. Since pre-trained models have trained in the ImageNet datasets (Krizhevsky et al. 2017), they have an output layer of 1000 neurons. Therefore, we changed them according to the number of classes (i.e., 5). We also tried to fine-tune all the layers of the networks. For this purpose, we set the trainable configuration of the layers as True.

In the dataset we collected, there is a notable discrepancy in the number of images among different species, resulting in an imbalanced dataset. This imbalance can lead to biased model training and subpar generalization performance. To address this issue, we employed data augmentation techniques (cropping, resizing, and rotating), which involve artificially increasing the size and diversity of the minority class. This is achieved by generating new synthetic samples through various transformations and modifications of the existing data. Additionally, data augmentation aids neural networks in distinguishing images of each class from noise or background interference. By creating multiple variations of each image while keeping the species consistent and altering the noise or background, data augmentation enhances the robustness and accuracy of the training process.

The input size resolution is 224×224 for all networks except for MobileViT where the input size is 256×256. The experimental framework was written in Python 3.9 and the Keras deep learning 2.10 library based on TensorFlow 2.10 backend. Models were executed on a single NVIDIA GeForce RTX 2080Ti GPU with 11 GB of video memory. We divided the dataset into training, validation, and test sets with a ratio of 70%, 20%, and 10%, respectively. To train the models, we set the number of epochs and the batch size to 100 and 16, respectively. The input images of the CNNs underwent one-to-one augmentation without duplication. To avoid the plateau phenomenon, the model's validation loss was monitored during the reduction in the learning rate to stop it when it does not improve. A learning-rate of 0.002 was set. We used the Adam algorithm for optimization. The performance of the models was evaluated by macro-averaged and micro-averaged versions of precision, recall, F-score, and overall classification accuracy extracted by the confusion matrices as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$F\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{3}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations, whereas Recall (Sensitivity) is the ratio of correctly predicted positive observations to all observations in the actual class. Thus, precision focuses on the prediction, whereas recall focuses on the measurements. F-Score is the harmonic average of Precision and Recall. The measure selected by the authors for ranking the systems was the overall classification accuracy score.

Results

The results obtained by the six pre-trained CNNs and three different types of MobileViT for the classification of the weed images are presented in Table 2. We conducted a comprehensive performance comparison of various deep learning models, including ResNet, MobileNet, and MobileViT, as they are widely recognized as state-of-the-art models for image classification tasks. These models have demonstrated remarkable capabilities in learning complex features and achieving high accuracy on diverse datasets. We evaluated the performance of each model on our weed dataset and analyzed their strengths and weaknesses in weed classification. By comparing their classification accuracy, training efficiency, and feature discrimination, we aimed to identify the most suitable model for our application in SSWM.

Table 2 Micro and macro measurements of the deep learning models on the test set for classification of the "CHEAL," "LAMPU," "LITAR," "SPRAR" and "STEME" classes

Network	Micro			Macro			
	Pr	Re	F1	Pr	Re	F1	Acc
MobileNetV2	87.78	87.13	87.78	84.83	86.12	85.42	87.83
MobileNetV3Small	88.08	87.60	88.04	86.05	85.20	85.32	88.26
MobileNetV3Large	**88.92**	**88.40**	**88.95**	**86.38**	**86.29**	**86.24**	**89.06**
ResNet50	88.70	**88.20**	**88.38**	87.57	**84.82**	85.41	**88.77**
ResNet101	88.70	87.72	87.76	89.55	83.26	85.20	88.26
ResNet152	**88.75**	87.84	88.03	**90.58**	83.12	**85.66**	88.48
MobileViT-2	88.55	84.70	86.52	84.88	83.73	84.26	86.38
MobileViT-8	89.58	86.85	88.16	86.82	84.76	85.67	87.61
MobileViT-16	**89.83**	**88.07**	**88.91**	**87.23**	**85.00**	**85.82**	**88.84**

The best results for each network group are indicated in bold for each network

We compared the performances of the three different types of ResNet i.e., ResNet50, ResNet101, and ResNet152. The ResNet50, ResNet101, and ResNet152 achieved accuracies of 88.77, 88.26 and 88.48%, respectively. These networks are very huge and heavy models that have many parameters to learn. On the other hand, our weed dataset is an imbalanced small dataset with 13,765 images. As we can see, using a huge deep learning model with many parameters on a small dataset can reduce system performance in the test set. In the following, we compared the performance of lightweight networks on the weed dataset (i.e., MobileNet and MobileViT). MobileNetV2, MobileNetV3Small, and MobileNetV3Large achieved accuracies of 87.83, 88.26, and 89.06%, respectively. As shown in Table 2, our observations show consistent and substantial enhancements from MobileNetV2 to MobileNetV3Large networks, outperforming ResNet networks in all performance metrics. This highlights the superiority of adopting lightweight backbone networks for small datasets. In contrast, the performance improvement from ResNet50 to ResNet152 is messy and irregular, emphasizing that heavy-weight deep learning models are more prone to overfitting when used with small datasets. We trained MobileViT from scratch to investigate the significance of ViT. In MobileViT, the attention mechanism is adapted and optimized to be computationally efficient and suitable for deployment on resource-limited devices like smartphones or edge devices. It achieves this by reducing the number of parameters and employing depthwise separable convolutions, which helps to maintain performance while reducing computational complexity. MobileViT achieved acceptable outcomes when compared to pre-trained MobileNet networks.

Confusion matrices related to the performance of all networks are shown in Fig. 4. Confusion matrices offer a potent and intuitive means to compare models and gain valuable insights into their performance across various classes and aspects of the classification problem. These matrices are extensively utilized in machine learning and data analysis to assess class-specific performance and to comprehend misclassifications. In a confusion matrix, the main diagonal represents the true positive (TP) counts for each class, and the values on the diagonal are the number of correctly classified instances for each class, and the off-diagonal elements represent the misclassifications. The results clearly demonstrate that ResNet networks show remarkable performance for the STEME class, achieving TP counts of 555, 569, and 559 for ResNet50, ResNet101, and ResNet152, respectively. This finding suggests that high-weight networks tend to excel with more data and exhibit more effective training when provided with a larger dataset. Additionally, the higher number of samples in the STEME class, compared to the other five classes, contributes to the enhanced performance of ResNet in this specific category.

To better illustrate the performance of the networks for each class, we drew the Receiver Operating Characteristic (ROC) curves. Figure 5 shows the ROC curves per class for all networks. The ROC curve is a graphical representation of the relationship between the true positive rate and the false positive rate at different classification thresholds. The diagonal line on the ROC curve represents a random classifier with no discrimination power, while a classifier close to this line performs no better than random guessing. The Area Under the ROC Curve (AUC) summarizes the overall performance of the model, where a perfect classifier has an AUC of 1 (100%) and a random classifier has an AUC of 0.5 (50%). A higher AUC value indicates better discriminative power and overall performance, and the closer the ROC curve is to the top-left corner, the better the model's discrimination power and performance. Upon analyzing the ROC curves, it becomes evident that MobileNetV3Large exhibits the highest AUC and provides the most visually optimal roc curves for all classes. Moreover, the roc curve of each class within MobileNetV3Large consistently approaches the top-left corner, indicating excellent discriminative power and performance when compared to other methods.

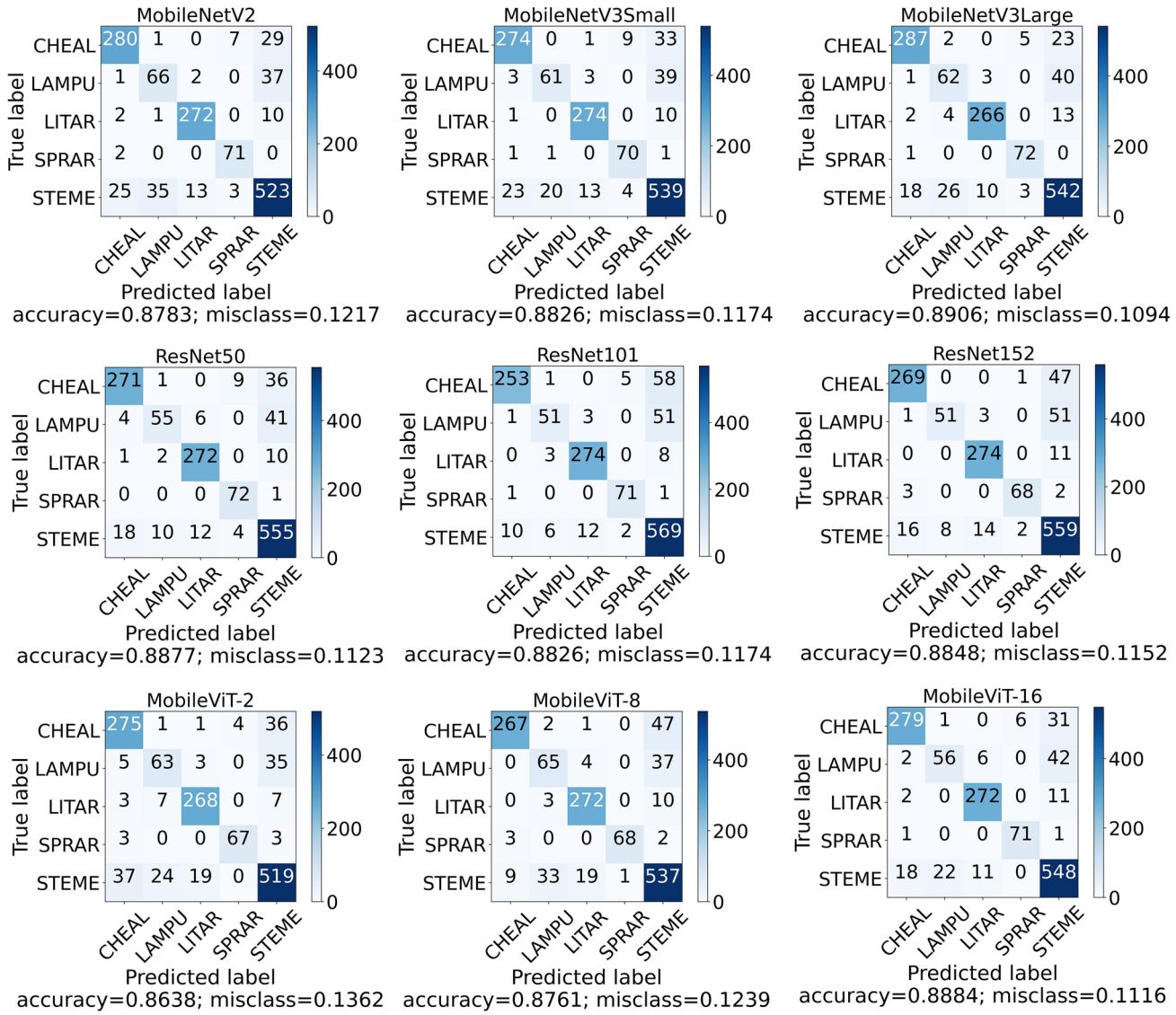

Fig. 4 Confusion matrices related to all networks for classifying weed species images of the weed dataset

To gain a deeper understanding of MobileNetV3Large's superiority, we conducted a comparison of bidimensional embeddings generated by MobileNetV3Large and other networks in the weed dataset (Fig. 6). The plot was generated by projecting a 128-dimensional embedding of the networks trained on the weed dataset, utilizing only CNN layers as a feature extractor, into two dimensions using the t-distributed stochastic neighbor embedding (t-SNE) algorithm (Alirezazadeh et al. 2022). Each class has been assigned a distinct color for visualization: CHEAL (purple), LAMPU (blue), LITAR (dark green), SPRAR (light green), and STEME (yellow). Employing t-SNE for feature embedding is an effective method to assess inter-class and intra-class relationships within the data. A distinct and well-separated cluster for each class in the t-SNE plot indicates successful discrimination by the model. It also reveals the variance and distribution of data points within each class. If data points of the same class form tight clusters and are closely grouped in the t-SNE plot, it indicates that the model has learned to recognize common patterns within that class. Conversely, if data points of the same class scatter across the t-SNE plot, it may indicate that the model struggles to capture intra-class variations effectively. The observations reveal that the MobileNetV-3Large network excels in effectively separating all classes, creating a highly discriminative space between them. The tight clusters and close grouping of points from the same class indicate that MobileNetV3Large demonstrates strong intra-class compactness, outperforming other methods in this aspect.

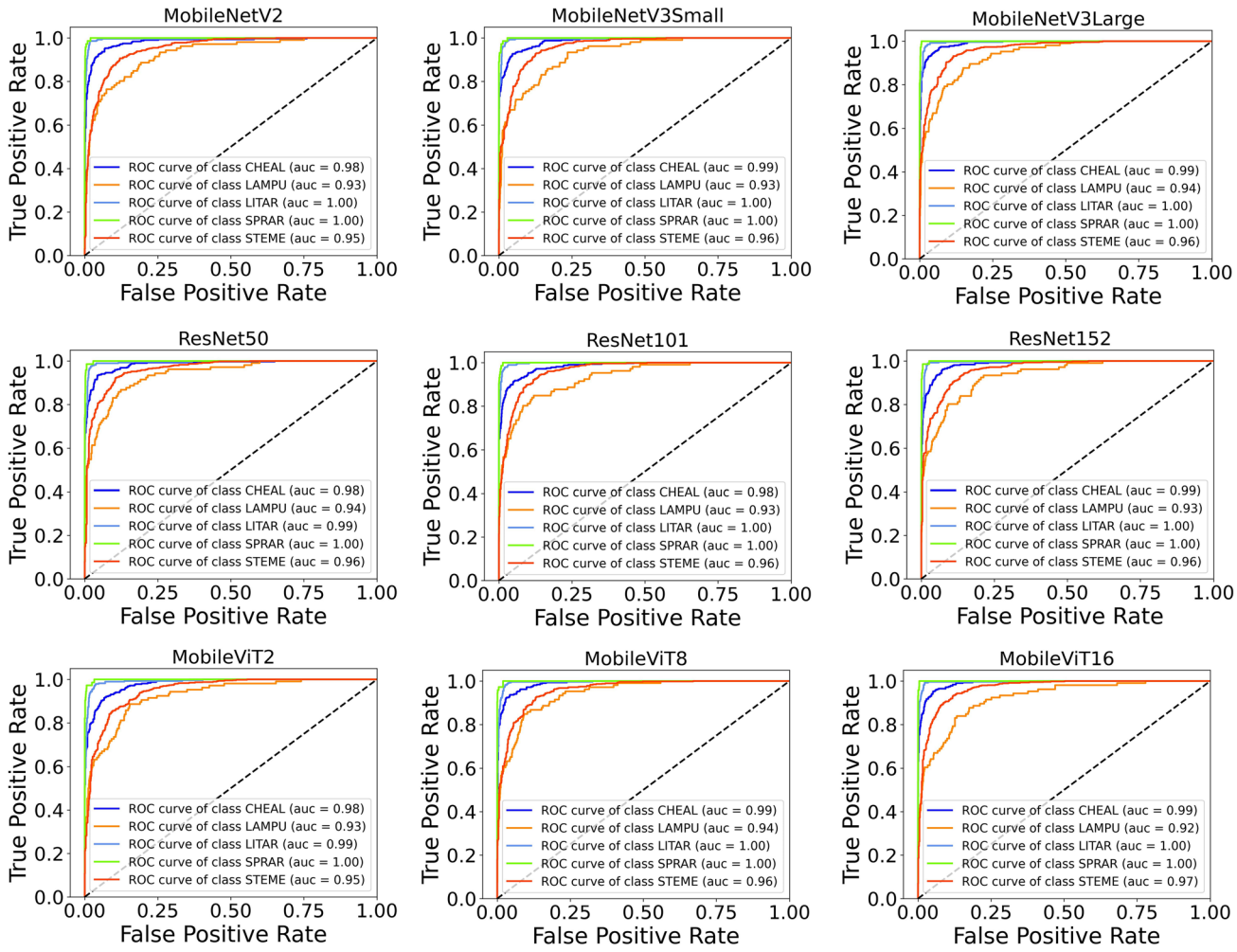

Fig. 5 ROC curves of the networks on the weed dataset for each class

Discussion

The proper classification of different weed species holds great importance in weed management, as it enables selective herbicide treatment. Selective herbicide treatment involves using herbicides that target specific weed species while sparing non-target plants and crops. By accurately categorizing weeds, farmers can choose herbicides that are most effective against the prevailing weed species in their fields. Moreover, understanding the distribution of weed species within the field empowers farmers to apply herbicides only where they are most needed, reducing waste and optimizing resource utilization. This targeted approach is especially valuable in SSWM, where precision agricultural techniques are employed to tailor weed control strategies to specific areas based on weed distribution and intensity. Through proper weed classification and selective herbicide treatment, farmers can enhance the efficiency and sustainability of weed management practices in agriculture.

This study explored the relationship between light-weight deep neural networks and small datasets in the context of weed classification for SSWM. The results revealed that light-weight models, such as MobileNet and MobileViT, showed a significant advantage over high-weight models, like ResNet, when dealing with small datasets. Despite achieving good overall accuracy, high-weight models suffered from overfitting issues on the limited data, leading to reduced generalization. In addition, the importance of the vision transformer was demonstrated in global representations combined with spatial relationships of CNNs.

Furthermore, the evaluation of inter-class and intra-class distances using t-SNE feature embeddings demonstrated the importance of not solely relying on accuracy as a performance metric. A good network should effectively compact the distance within each class, enhancing the model's ability to distinguish between different samples and improving class separation in the feature space.

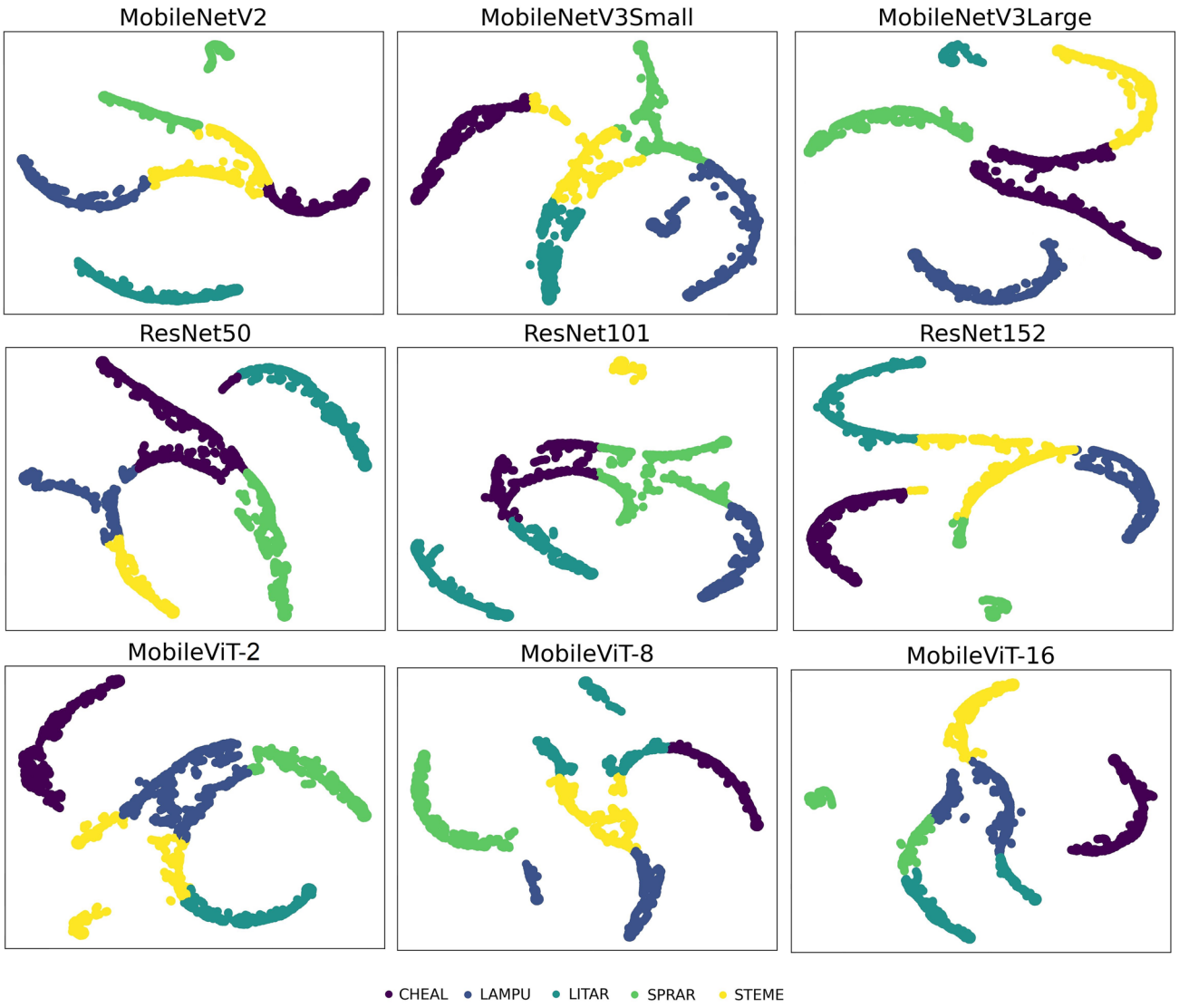

Fig. 6 The comparison of the t-SNE embedding features resulting from the networks trained on the weed dataset without fully connected layers (only CNN layers) as a feature extractor. Each class has been assigned a distinct color for visualization: CHEAL (purple), LAMPU (blue), LITAR (dark green), SPRAR (light green), and STEME (yellow)

In conclusion, a well-optimized deep learning model, particularly one with attention to lightweight architecture and compact intra-class distances, is capable of discriminating between different weed species effectively and generalizing well to new data. By leveraging lightweight architectures and paying attention to compact intra-class distances, the models can achieve excellent performance without overfitting to the training data.

Acknowledgements This project is supported by funds from the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany. The Federal Office for Agriculture and Food (BLE) provides coordinating support for artificial intelligence (AI) in agriculture as a funding organization, Grant Number 28DK105A20.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declaration

Conflict of interest The authors have no relevant financial or non-financial interests to disclose and have no conflict of interest to declare that are relevant to the content of this article.

References

Alirezazadeh P, Rahimi-Ajdadi F, Abbaspour-Gilandeh Y, Landwehr N, Tavakoli H (2021) Improved digital image-based assessment of soil aggregate size by applying convolutional neural networks. Comput Electron Agric 191:106499

Alirezazadeh P, Dornaika F, Moujahid A (2022) A deep learning loss based on additive cosine margin: application to fashion style and face recognition. Appl Soft Comput 131:109776

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

de Camargo T, Schirrmann M, Landwehr N, Dammer K-H, Pflanz M (2021) Optimized deep learning model as a basis for fast UAV mapping of weed species in winter wheat crops. Remote Sens 13(9):1704

Gerhards R, Andujar Sanchez D, Hamouz P, Peteinatos GG, Christensen S, Fernandez-Quintanilla C (2022) Advances in site-specific weed management in agriculture-a review. Weed Res 62(2):123–133

Hafeez A, Husain MA, Singh S, Chauhan A, Khan MT, Kumar N, Chauhan A, Soni S (2022) Implementation of drone technology for farm monitoring and pesticide spraying: a review. Inf Process Agric

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

Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V et al (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 1314–1324

Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90

Loddo D, Scarabel L, Sattin M, Pederzoli A, Morsiani C, Canestrale R, Tommasini MG (2019) Combination of herbicide band application and inter-row cultivation provides sustainable weed control in maize. Agronomy 10(1):20

Mehta S, Rastegari M (2021) Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv:2110.02178

Monteiro A, Santos S (2022) Sustainable approach to weed management: the role of precision weed management. Agronomy 12(1):118

Osorio K, Puerto A, Pedraza C, Jamaica D, Rodríguez L (2020) A deep learning approach for weed detection in lettuce crops using multispectral images. AgriEngineering 2(3):471–488

Pakdaman Sardrood B, Mohammadi Goltapeh E (2018) Weeds, herbicides and plant disease management. Sustain Agric Rev Biocontrol 31:41–178

Pflanz M, Nordmeyer H, Schirrmann M (2018) Weed mapping with UAS imagery and a bag of visual words based image classifier. Remote Sens 10(10):1530

Rai N, Zhang Y, Ram BG, Schumacher L, Yellavajjala RK, Bajwa S, Sun X (2023) Applications of deep learning in precision weed management: a review. Comput Electron Agric 206:107698

Rodrigo M, Oturan N, Oturan MA (2014) Electrochemically assisted remediation of pesticides in soils and water: a review. Chem Rev 114(17):8720–8745

Villette S, Maillot T, Guillemin J-P, Douzals J-P (2022) Assessment of nozzle control strategies in weed spot spraying to reduce herbicide use and avoid under-or over-application. Biosyst Eng 219:68–84

Wang A, Zhang W, Wei X (2019) A review on weed detection using ground-based machine vision and image processing techniques. Comput Electron Agric 158:226–240

Wato T, Amare M, Bonga E, Demand B, Coalition B (2020) The agricultural water pollution and its minimization strategies-a review. J Resour Dev Manag 64:10–22