frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 10 February 2023
DOI 10.3389/feart.2023.1146558

OPEN ACCESS

EDITED BY
Isa Ebtehaj,
Université Laval, Canada

REVIEWED BY
Hamidreza Bolhasani,
Islamic Azad University, Iran
Omid Memarian Sorkhabi,
University College Dublin, Ireland

*CORRESPONDENCE
Jingjing Yang,
r78z@foxmail.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 17 January 2023
ACCEPTED 02 February 2023
PUBLISHED 10 February 2023

CITATION
Guo X, Ge Y, Liu F and Yang J (2023), Identification of maize and wheat seedlings and weeds based on deep learning.
Front. Earth Sci. 11:1146558.
doi: 10.3389/feart.2023.1146558

Identification of maize and wheat seedlings and weeds based on deep learning

Xiaoqin Guo, Yujuan Ge, Feiqi Liu and Jingjing Yang*

School of Information Science and Engineering, Hebei North University, Zhangjiakou, China

Introduction: It is well-known that maize and wheat are main food crops in the world. Thus, promoting high quality and abundant maize and wheat crops guarantees the development of the grain industry, which is needed to support world hunger. Weeds seriously affect the growing environment of maize, wheat, and their seedlings, resulting in low crop yields and poor seedling quality. This paper focuses on the identification of maize and wheat seedlings and field weeds using deep learning.

Methods: Maize and wheat seedlings and field weeds are the research objects. A weed identification model based on the UNet network model and ViT classification algorithm is proposed. The model uses UNet to segment images. A Python Imaging Library algorithm is used to segment green plant leaves from binary images, to enhance the feature extraction of green plant leaves. The segmented image is used to construct a ViT classification model, which improves the recognition accuracy of maize and wheat seedlings and weeds in the field.

Results: This paper uses average accuracy, average recall, and F1 score to evaluate the performance of the model. The accuracy rate (for accurately identifying maize and wheat seedlings and weeds in the field) reaches 99.3%. Compared with Alexnet, VGG16, and MobileNet V3 models, the results show that the recognition effect of the model trained using the method presented in this paper is better than other existing models.

Discussion: Thus, this method, which accurately disambiguates maize and wheat seedlings from field weeds can provide effective information support for subsequent field pesticide spraying and mechanical weeding.

KEYWORDS

deep learning, image classification, weed identification, maize seedlings identification, wheat seedlings identification

1 Introduction

In crop production, weeds are considered to be the key negative factor affecting agricultural production, i.e., the most harmful factor for agricultural production (Wang et al., 2019). Weeds compete with maize and wheat seedlings for nutrients, fertilizer, sunlight, and growth space, which impedes ventilation and lighting in the field. This competition seriously affects the growth environment of maize and wheat, reduces the yield of maize and wheat, and also affects farming operations (Gharde et al., 2018). A large amount of research data shows that the competition with weeds for resources in the field has a strong correlation with crop yield loss, resulting in an increasing proportion of crop yield loss (Wang et al.,

2020) As a main food crop, maize and wheat have the largest crop sizes, the largest consumption, and the most economic value (Basit et al., 2019; Venkataraju et al., 2022). Clearing weeds in the field in time is helpful in maintaining crop yield and realizing precision agriculture. To efficiently remove weeds in the field, it is necessary to accurately determine information about maize and wheat seedlings and the types of weeds in the field (Ahmad et al., 2021). Traditionally, management methods focus directly on weeds, and manual weeding, chemical weeding, and mechanical weeding are used to control weeds in the field (Abbas et al., 2018). There are advantages and disadvantages to using traditional methods to control weeds. Manual weeding has a clear wedding goal, which is to not harm maize, wheat, and seedlings; it can also root out and improve the control effect of weeding, but the operation efficiency is low, time-consuming, and labor-intensive, and the labor cost is high. Chemical weeding methods have high weeding efficiency and low cost, but depending on the professional knowledge of researchers, improper use of herbicides will cause crop damage and environmental pollution (Manisankar et al., 2022). Mechanical weeding methods have high weeding efficiency, low labor intensity, and low labor cost. But may cause damage to maize and wheat seedlings (e.g., remove weeds in the field under the condition of close planting).

With the progress of artificial intelligence, it has been applied in earth science, such as water level prediction (Deng et al., 2022), precipitation prediction (Luo et al., 2022), rock classification (Chen et al., 2023) and flood prediction (Ebtehaj and Bonakdari, 2022), and so on. It is very important for agricultural production. Deep learning is an important direction of artificial intelligence. We use deep learning to identify maize, wheat and field weeds. Currently, image feature extraction based on machine vision and the recognition network modeling and deep learning are widely used in agricultural image recognition. Wu et al. (2021) use machine vision recognition methods to identify crops or weeds by extracting the texture, shape, color, or size features of images. Experiments show that weeds can be accurately identified in specific crops and specific environments, but the method is not suitable for large-scale identification or classification of crops. Osorio et al. (2020) proposed a recognition method based on a gradient histogram (HOG) and support vector machine (SVM). The HOG algorithm is used to extract image features, which mainly includes gradient calculations of the image, statistics of the gradient direction and magnitude of the image, grasping the characteristics of local shapes using the extracted edge and gradient features, using an NDVI index as the preprocessing stage of background estimation, then adding SVM for crop detection, and finally segmenting only weed class objects, and then determining the coverage of the weeds by calculating the pixel ratio. Although HOG can capture local shape information well, has good generalization ability, and shows good performance in identifying maize and wheat seedlings and weeds in the field, the process of feature acquisition is complicated and the dimensionality is high, which leads to poor real-time performance. Additionally, the blocked crops are not easily detected. Bah et al. (2018) collected data on field weeds *via* unmanned aerial vehicle (UAV), and then trained data sets from the images collected by using CNNs and unsupervised training data markers, and established a recognition model of maize and wheat seedlings and field weeds. Compared with traditional recognition methods, experimental results show that CNN-based training is simple and fast, and the experimental results are close to the performance of monitoring training data markers. However, the image resolution acquired by unmanned aerial vehicles is low and is affected by light conditions. Because maize and wheat are similar to weeds, this complication makes it difficult to distinguish between crops and weeds. Jiang et al. (2020) put forward a GCNAlexnet-101 network model to extract features through CNN. This model can improve the recognition accuracy of crops and weeds under the condition of limited label data, and its accuracy rate is better than that of Alexnet-101. However, the method doesn't consider the influence of soil on crops and leaves with abnormal leaf morphology due to poor crop growth. Louargant et al. (2018) introduced an unsupervised classification algorithm that uses multispectral images to collect data and identifies maize and weeds in the field by combining spatial and spectral information. Without considering the low resolution or distortion of images, this method does not need to manually label data, thus reducing the cost of data labeling. Du et al. (2022) trained the ALWeeds data set using MobileNet V2. The model can still maintain high accuracy on low-memory equipment, but its application environment is limited and is only suitable for weeding in enclosed environments with light control. Fu et al. (2020) introduced a VGGI model, which is improved on the initial VGG model. It reduces the number of convolution layers to reduce the network parameters while ensuring good classification accuracy, but the VGGI model can't maintain high accuracy in complex backgrounds or with images having different shooting angles. Wang et al. (2018) built a multi-scale convolution neural network model and introduced a method of multi-scale layered features to identify maize and weeds. The accuracy rate of weed identification by this method can reach 98.92% and has a good effect on the overlapping part of maize and weeds. However, the training process of this model is complicated and computationally expensive. Tannouche et al. (2022) evaluated and tested six models, i.e., VGG16, VGG19, Inception v3, Inception v4, MobileNet V1, and MobileNet V2, to detect weeds. The experimental results show that Inception v4 has the highest accuracy on the mixed image set, and MobileNet V2 has a fast processing speed and a small model. However, since there are many kinds of weeds, and its model recognition ability is limited, accurately identifying different weed species is nearly impossible. Pei et al. (2022) used a lightweight model such as YOLOv4-Tiny to detect image data sets. YOLOv4-tiny uses a structure that is a simplified version of YOLOv4. To improve its performance, an ACON activation function and CBAM model are used. In this way, the marking cost is reduced, the efficiency is greatly improved, and the accuracy of crop detection is improved. A large number of experiments prove that the ACON activation function (Ma et al., 2021) is superior to Relu and Swish activation functions in classification and detection. CBAM (Wang et al., 2022) is an Attention mechanism module that combines spatial features and channels, which can enhance the ability to extract network features. Compared with attention mechanism modules, which only pay attention to one aspect, CBAM can give attention to both aspects and achieve better results. Although the YOLOv4-tiny model has achieved good recognition results in all single aspects, it is difficult to detect when maize seedlings, wheat seedlings, and weeds overlap and block each other.

Therefore, disambiguating maize and wheat seedlings from field weeds is an important but difficult task. Although the reviewed

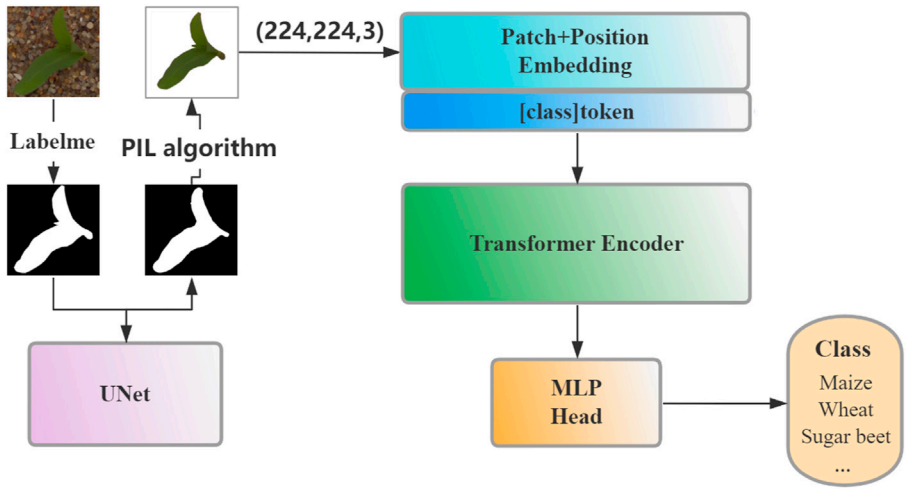

FIGURE 1
Flow Chart of maize and Wheat Seedlings and Field Weeds.

training models have achieved very good performance and can achieve good identification effects, their application scope is often limited. Most reviewed methods can't accurately distinguish between maize seedlings, wheat seedlings, and weeds under complicated background conditions or the mutual shielding of maize leaves, wheat leaves, and weeds. This study explores the identification of maize and wheat seedlings and weeds in the field and proposes an identification method based on the UNet network model and ViT classification algorithm. This method pays more attention to the extraction of local features and edge features of images, and can accurately disambiguate maize and wheat seedlings from weeds in the field. This will in turn effectively promote the healthy growth of maize and wheat seedlings, and improve the yield and quality of maize and wheat. The main contributions of this paper are summarized as follows:

1 Combining Unet and ViT, the important characteristics of target crops and weeds are highlighted, and this further improves the ability to identify different weeds and crops in complex environments. This is also the first time that a UNet network segmentation model and ViT classification model are combined to identify maize and wheat seedlings and weeds.
2 Unet is used to remove the background, and the PIL algorithm is utilized to segment the leaves of green plants and get a clear weed image.
3 Compared with Alex, VGG16 and MobileNet V3, the proposed method is superior to MobileNet V3 in accuracy, recall and F1 value.

2 Methodology

Data sets containing images of maize, wheat, and weeds at various growth stages were used, where D represents the data set, x_1, y_1 represents the first sample, x_n represents n samples, X represents the sample features, Y represents the sample label, x_i represents the latitude of the sample feature vector is D, and y_i represents that there are 12 categories in the data set. The model training component of the data set uses the label category of the input image that can be accurately identified.

In this study, a method to identify maize and wheat seedlings and weeds in the field by combining the UNet network model and ViT classification algorithm is developed. The proposed method can effectively improve the identification accuracy of maize and wheat seedlings and weeds in the field. As shown in Figure 1, the UNet network model for image segmentation can generate a network segmentation model. Next, the PIL algorithm is used to segment green plant leaf features, i.e., the input of the classification model, and the ViT classification model is used to output the image classification results.

When using UNet to segment images, the image background of the dataset of maize and wheat seedlings and weeds in the field used in this study has noise and different resolutions. To better improve the segmentation effect of the model, a preprocessing step extracts the relevant information of the image by eliminating irrelevant noise and interference in the image; this step effectively improves the accuracy of the algorithm (Ranganathan, 2021). First, we use PIL to batch adjust the resolution of images to 256×256 with a bit depth of 8, and store them in PNG format for use. Second, a random rotation, flipping, scaling, translation, and clipping is used to expand the data set. By performing these data enhancement operations on the image, the training data is effectively increased and the robustness of the model is improved. These operations also improve the generalization ability of the model and helps to avoid over-fitting. After the image data sets are preprocessed uniformly, a closed polygonal polyline of Labelme is used to label the data sets. The labeled JSON file is converted into an 8-bit binary label map, as shown in Figure 3. Labelme pays more attention to the edge and outline details of the image, which can focus the model on the leaf details and obtain better segmentation results.

10.3389/feart.2023.1146558

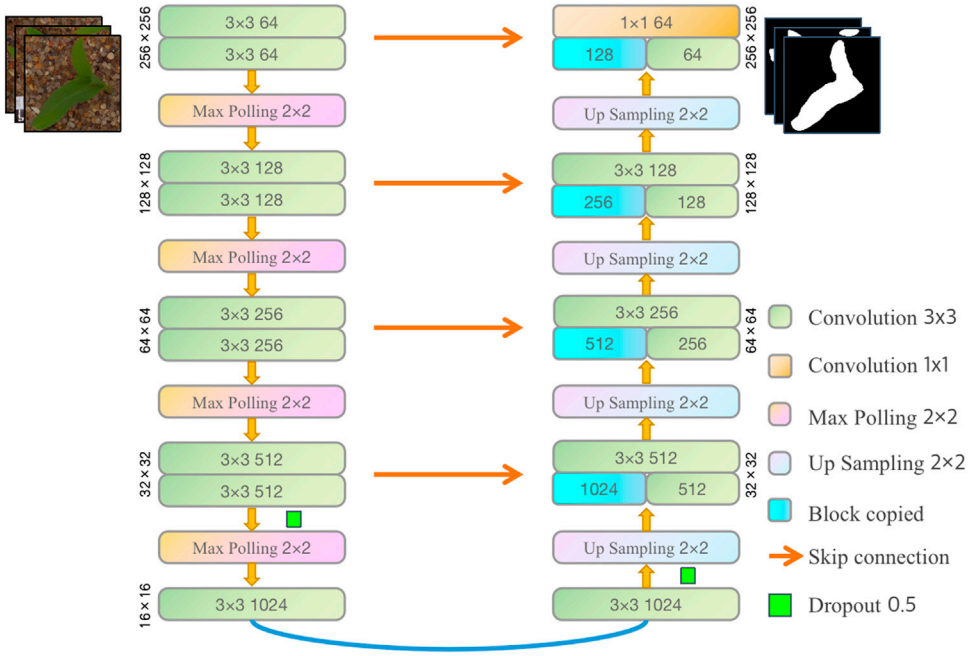

FIGURE 2
Model architecture diagram.

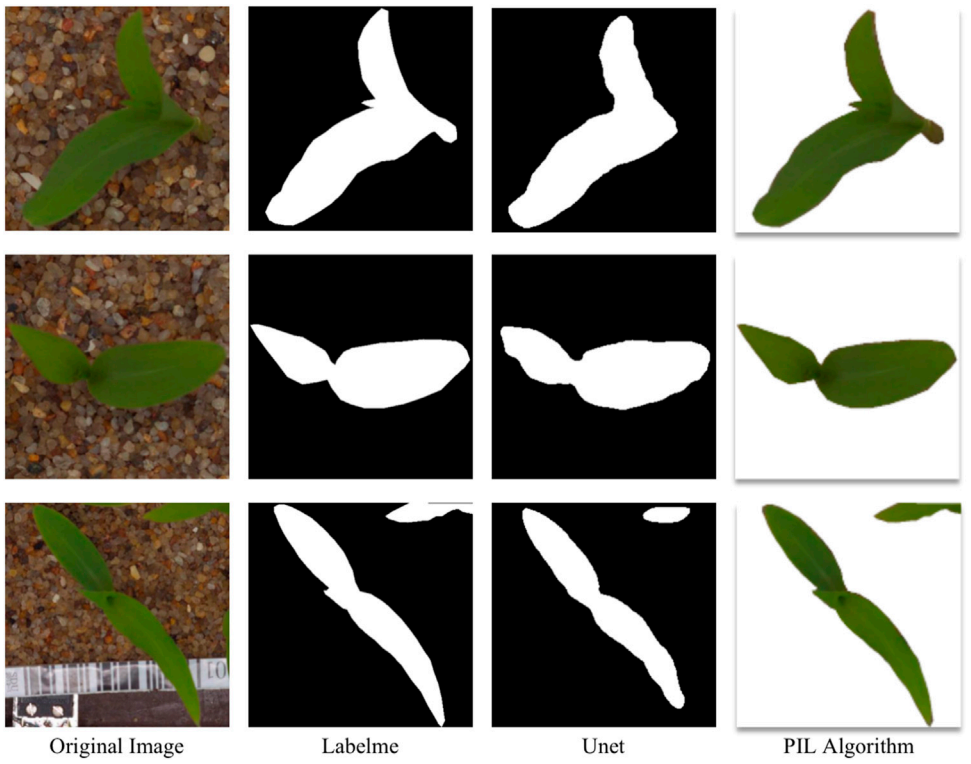

FIGURE 3
Sample segmentation diagram of partial data sets.

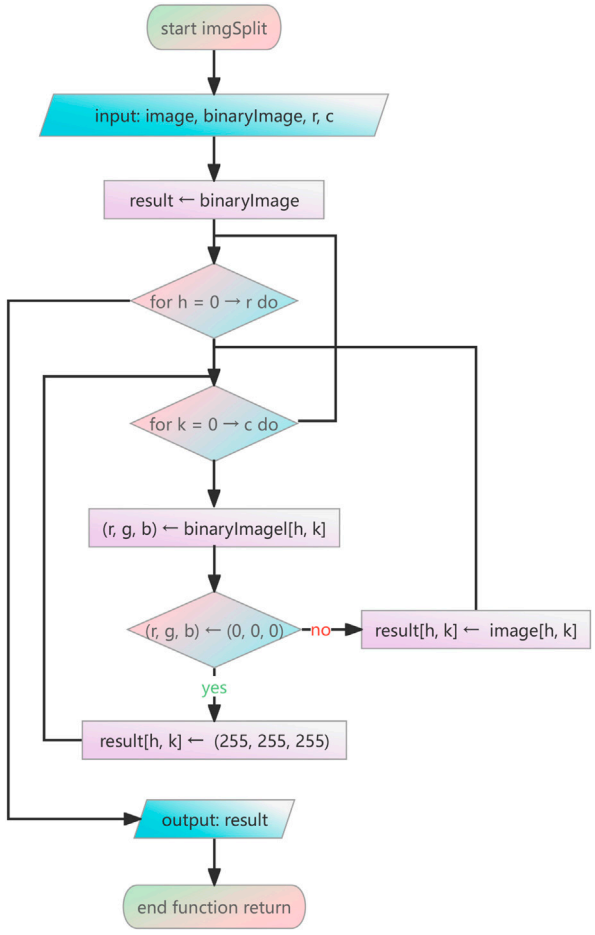

FIGURE 4
Flow chart of the proposed method.

UNet is used to train and label the data set. Through this network, the input image is segmented at the pixel level; specifically, the leaves of maize and wheat seedlings and field weeds are separated from the complex background. Thus, the leaf characteristics of maize and wheat and weeds are extracted to improve the recognition ability of maize and wheat seedlings and field weeds.

In this paper, a convolved layer (conv) with a convolution kernel of 3 × 3 is used to extract the characteristic information of maize, wheat, and weeds. When performing the convolution, uniform padding is used to ensure that the output layer size is equal to the input layer size. Relu is the active layer, MaxPooling is the maximum pool layer of 2 × 2, and UpConv is the transposed convolution of 2 × 2. The model architecture is shown in Figure 2. After each convolution, the height and width of the characteristic layer of maize, wheat, and weeds remain unchanged. For each downsampling, the height and width of the characteristic layer of maize, wheat, and weeds are reduced by half, and the number of channels is doubled. This layer is performed to extract the texture and some detailed features of maize, wheat, and weeds. In the third and fourth layers, 50% of neurons will be discarded, reducing over-fitting. By transposing the convolution, the height and width of the characteristic layer of maize, wheat, and weeds are doubled, the number of channels is halved, and the size of the picture is gradually enlarged. This step is made to extract the deep information, contouring, and shape features of maize, wheat, and weeds. In the process of sampling at each level, the feature map of maize, wheat, and weeds at the corresponding position of the encoder is combined with deep features and the shallow features *via* Skip Connection, so that more feature information is retained and the segmentation accuracy is improved. Finally, a Conv+Sigmoid operation with a convolution kernel of 1×1 is performed to generate a segmented image, and the segmented result is output, as shown in Figure 3.

UNet is effectively used to improve the recognition speed and accuracy of maize and wheat seedlings and field weeds detection. After maize and wheat seedlings and field weeds were separated from the complex background, and to better extract the leaf characteristics of maize and wheat, the PIL algorithm was used to traverse every pixel point of the whole binary image. The algorithm evaluated and segmented the green plant leaves pixel-wise (pixel-by-pixel) and strengthened the extraction of the main features of maize and wheat and field weeds, as shown in Figure 3. This improved the performance of the classification model and realized an accurate classification of images. The flow chart and pseudo code of the algorithm are shown in Figures 4, 5.

The segmented data sets of maize and wheat seedlings and weeds in the field were used to train the ViT classification model. ViT abandoned the convolution network structure, completely adopted the pure transformer structure to complete the classification task, and achieved better results on large-scale data sets than CNN (Dosovitskiy et al., 2020). The transformer's standard is the input sequence, so it was necessary to convert the input to two-dimensional images of maize and wheat seedlings and weeds in the field. An image-blocking strategy was used to divide an H×W×C image into a non-overlapping patch of size $N = HW/P^2$ using a linear transformation to map each image block to a D-dimensional feature vector. By embedding position codes into each patch, the spatial position information between input image blocks was retained, as shown in Formula 1. A learnable category vector $x_0^0 = X_{class}$ for the embedded sequence of length N was used to learn the category information in the process of training the model,

$$x = [X_{class}]; X_p^1 E, X_p^2 E, \ldots; X_p^N + E_{pos}, E \in \Re^{(P^2 \cdot C) \times D}, E_{pos} \in \Re^{(N+1) \times D} \quad (1)$$

where W is the width of the input image, H is the height of the input image, C is the number of channels, P is the size of patches, and N is the number of patches. The vector dimension of each image block is $x_p \in \Re^{N \times (P^2 \cdot C)}$.

The combined sequence of category vector, image block embedding, and position coding are imported into the transformer encoder for feature extraction. This was done to calculate the weight coefficient of each feature vector *via* a self-attention module, and pay attention to the importance of each feature vector and the relationship between each feature vector. By alternately stacking the transformer encoder several times, the

Algorithm 1 PIL image segmentation

Input: *image*: original image, *binaryImage*: binary image, r: number of rows in binary graph, c: number of columns in binary graph

Output: split graph

```
function IMAGESPLIT(image, binaryImage, r, c)
    result ← binaryImage
    for h = 0 → r do
        for k = 0 → c do
            (r, g, b) ← binaryImage[h, k]
            if (r, g, b) = (0, 0, 0) then
                result[h,k] ← (255, 255, 255)
            else
                result[h,k] ← image[h, k]
            end if
        end for
    end for
    return result
end function
```

FIGURE 5
The pseudo-code of the proposed method.

FIGURE 6
Part of the data set samples is segmented by the proposed method.

TABLE 1 The total number of training, validation, and testing images of crop and weed species for building the proposed model.

Weeds/Crops	Training	Validation	Testing	Total
Black-grass	179	75	75	329
Charlock	256	100	100	465
Cleaver	194	79	79	353
Common Chickweed	389	143	143	675
Common wheat	155	66	66	287
Fat Hen	307	117	117	541
Loose Silky-bent	413	153	153	719
Maize	155	66	66	287
Scentless Mayweed	332	125	125	582
Shepherds Purse	161	68	68	297
Small-flowered Cranesbill	320	121	121	562
Sugar beet	252	98	98	448

features corresponding to the learnable class embedding vector Class Token are extracted from MLP and used for image classification, as shown in Formula 2,3,4,5. Finally, the probability distribution of the current image belonging to each category is the output, which is usually also called a vector. The value of each vector dimension is 0–1, and the highest probability distribution is the category to which the predicted image belongs.

$$Attention(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{soft\,max}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_k}}}\right) \tag{2}$$

$$\mathrm{x}_{\ell}^{'} = \mathrm{MSA}\left(\mathrm{LN}\left(\mathrm{x}_{\ell-1}\right)\right) + \mathrm{x}_{\ell-1}, \ell = 1...L \tag{3}$$

$$\mathrm{x}_{\ell} = MLP\left(LN\left(\mathrm{x}_{\ell}^{'}\right)\right) + \mathrm{x}_{\ell}^{'}, \ell = 1...L \tag{4}$$

$$y = LN\left(\mathrm{x}_{L}^{0}\right) \tag{5}$$

3 Experiment

3.1 Experimental data set

The data set used to train and evaluate the images in this experiment was published by Giselsson et al. (2017). Giselsson dataset includes maize and wheat at different growth stages, and many kinds of weeds: Black-grass, Charlock, Cleaver, Common Chickweed, Common wheat, Fat Hen, Loose Silky-bent, Maize, Scentless Mayweed, Shepherds Purse, Small-flowered Cranesbill, and Sugar beet. The data set includes 5545 images with different resolutions. To reduce over-fitting, the Keras ImageDataGenerator function library is used to preprocess the data set, and the data set is expanded *via* random rotation, flipping, scaling, translation and cropping. The images are then labeled *via* a closed polygonal polyline of Labelme and the corresponding class label is generated. Then, the PIL algorithm is used to segment green plant leaves, as shown in Figure 6. Finally, the data set is randomly divided into the training set, testing set, and verification set according to a ratio of 60:20:20. Table 1 shows the number of images in each crop and weed species category used for training, verification, and testing data sets. The training set is used to construct the recognition model of maize and wheat seedlings and weeds in the field, the verification set is used to preliminarily evaluate the recognition ability of the model, and the test set is used to evaluate the generalization ability of the final model.

3.2 Experiment settings

To verify the accuracy and performance of the UNet network model and ViT classification algorithm in identifying maize and wheat seedlings and weeds in a complex environment, under the same operating environment and super-parameters, an experiment was designed that compared the proposed algorithm to three models, Alexnet, VGG16, and MobileNet V3. The hardware environment used consists of an Intel (R) Xeon (R) CPU E5-2690v3 @ 2.60 GHz processor, 32G memory, NVIDIA GeForce RTX 3060 graphics card, and Win10 operating system. The programming language used is *Python*3.9 and a Tensorflow 2.5.0 framework is used to build the model. To ensure comparative effectiveness, all models adopt the Adam optimizer, the weight attenuation was set to 1e-4, the learning rate was set to be consistent, and the number of images trained in each iteration was set to 8, with a total of 160 iterations. The hyperparameter settings of Alexnet, VGG16, Mobilenet V3, and the proposed model are shown in Table 2.

Experimental Accuracy, Precision, and Recall were used as model evaluation indicators to evaluate algorithmic performance (Grandini et al., 2020). Because accuracy and recall affect each other, an F1-Score is used as a comprehensive index to balance the influence of accuracy and recall and comprehensively evaluate the classification model. The calculation formulas of accuracy, precision, recall, and F1 score are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FF + FN} \tag{6}$$

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{7}$$

$$\mathrm{Recall} = \frac{TP}{TP + FP} \tag{8}$$

$$F1 = \frac{2\mathrm{Pr}ecision \times \mathrm{Recall}}{\mathrm{Pr}ecision + \mathrm{Recall}} \tag{9}$$

where TP refers to predicting positive samples as positive samples, TN refers to predicting negative samples as negative samples, FP refers to predicting negative samples as positive samples, and FN refers to predicting positive samples as negative samples.

3.3 Comparison of experiment results

Figure 7 shows the identification accuracy and loss curve of the method in the training set and verification set of maize and wheat seedlings and weeds in the field. As can be inferred from the figure, when the Epoch is 40, the accuracy of this method in the training set

TABLE 2 Hyperparameter setting.

Hyperparameter	Alexnet	VGG16	MobilenetV3	Proposed method
Number of classes	12	12	12	12
Optimizer	Adam	Adam	Adam	Adam
Image patch size	8	8	8	8
Initial learning rate	0.0001	0.0001	0.0005	0.001
Dropout rate	0.5	0.5	0.5	0.5
Epochs	160	160	160	160

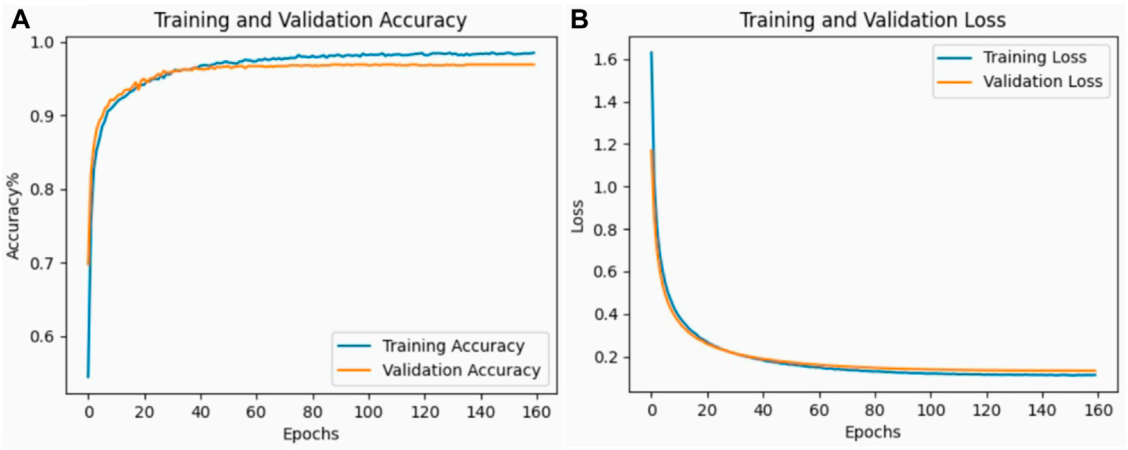

FIGURE 7
(A) Accuracy curve of the proposed method. **(B)** Loss curve of the proposed method.

and verification set gradually increases and tends to be stable after the 120th Epoch. When the Epoch is 120, the loss in the training set and verification set decreases and tends to be stable. When Epoch is set to 150, the accuracy of the proposed method is 99.3% in the training set and 98.1% in the verification set, which shows that the recognition accuracy in the training set and the verification set is stable. When Epoch is set to 160, the loss rate in the training set and verification set converges.

Alexnet, VGG16, and MobileNet V3 models are trained with the same parameters as the methods in this paper. All models achieved good recognition accuracy in the training set, and the loss converges to a stable value. When the number of iterations reaches 40, the accuracy tends to rise steadily, and when the number of iterations reaches 160, the accuracy of the training set approaches 99%. Thus, when the number of iterations reaches 150, the loss of each model in the training set gradually converges. But the effect on the verification set is average. The accuracy rate of Alex is 80.2%, that of VGG16 is 82.5%, and that of MobileNet V3 is 88.7%. Thus, the accuracy rate and loss value on the verification set fluctuate greatly between methods. With the increase of iterations, the accuracy of the validation set does not improve, the loss rate does not converge well, and the generalization ability of the model is poor.

The experimental results of the four different identification models, Alexnet, VGG16, MobileNet V3, and the proposed method are shown in Table 3. According to the data in the table, the accuracy of each model is similar for the training sets, however, the verification set shows that the training time of Alexnet is short, but the accuracy is the lowest, which is due to the number of layers that Alexnet has and its relatively simple network. The training time for VGG16 is 4 times longer than that of Alexnet, but the accuracy rate is only increased by 2%. Although MobileNet V3 uses less training time, its accuracy rate is only 88.7%, which does not reach the ideal accuracy rate. By contrast, although the training time of the model is not the shortest, the proposed method shows high recognition accuracy in the training set and the verification set, which are 99.3% and 98.1%, respectively. Therefore, the proposed method can effectively improve the identification accuracy of maize and wheat seedlings and weeds in the field.

The experimental results, including the total parameter number and single image prediction time of Alex, VGG16, MobileNet V3, and the four different recognition models in this paper are shown in Table 4. The experiment uses 1200 images in the test set for batch prediction and calculates the prediction time of a single image. The table shows that MobileNet V3 has the lowest number of parameters, meaning the model is the smallest, but the prediction

TABLE 3 Experimental results of different models.

Method	Training duration (hour)	The accuracy of the training set (%)	The accuracy of the verification set (%)
Alexnet	0.8	0.969	0.802
VGG16	3.3	0.987	0.825
MobileNet V3	2	0.988	0.887
Proposed Method	4.5	0.993	0.981

TABLE 4 Total parameter quantity and prediction time of different models.

Method	Total parameter quantity	Number of layers	Model size (MB)	Single image prediction time (ms)
Alexnet	14,606,028	8	55.72	87
VGG16	70,317,900	16	268.24	212
MobileNet V3	4,241,804	28	16.18	484
Proposed Method	85,802,501	12	327.31	352

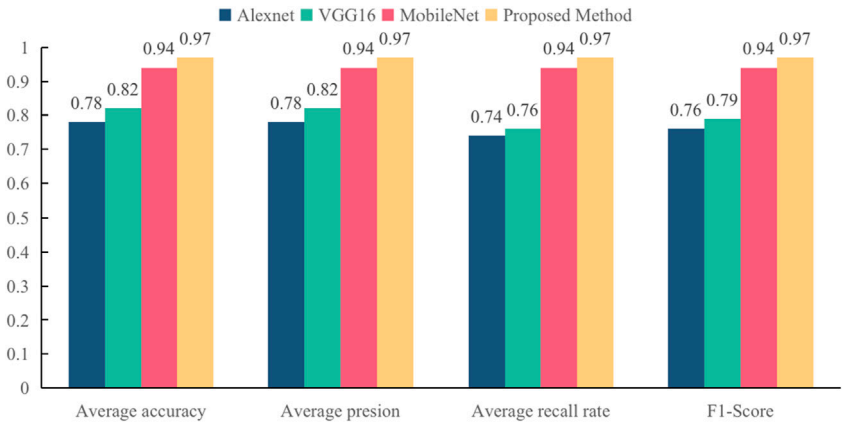

FIGURE 8
Accuracy rate, Precision rate, Recall rate, and F1 score of different models.

time of a single image is the longest, which may be due to the complex network and deep layers of this model. Although MobileNet V3 redesigns the time-consuming layer structure and improves the running time, it is still just as complex as other networks. The Alexnet model is small, with few layers and the fastest prediction time. However, it is not suitable for weed identification because of its low prediction accuracy. From the total number of parameters and model size, VGG16 and the proposed method has similar prediction time for a single image, and it is close to VGG16 when there are many parameters. However, the guarantee of recognition accuracy makes the proposed method more efficient when considering the test set.

Comparison results of average accuracy, average precision, average recall, and F1 score of four different recognition models, Alex, VGG16, MobileNet V3, and the proposed method, are shown in Figure 8. The accurate rates of each model are 78%, 82%, 94%, and 97% percent respectively. The data demonstrates that the accuracy rate and precision rate of Alexnet is not high. The accuracy of VGG16 is not as good as that of MobileNet V3, and the accuracy of VGG16 is 82%, while that of MobileNet V3 is 12% higher than that of VGG16. This is because the channel attention mechanism is added in the MobileNet V3 network, and the different weights of each feature layer are analyzed. Thus, the important features are given more weight, whereas the opposite is given less weight. Therefore, MobileNet V3 pays more attention to extracting the important features of the feature layer. Although the accuracy rate of MobileNet V3 is high, it is still lower than that of the proposed method. The average accuracy rate, average precision rate, average recall rate, and F1 score of the proposed algorithm are 97%, 97%, 97%, and 97%, respectively. Alexnet and VGG16 models have the

lowest recall rate, and the proportion of positive samples is the lowest. The difference between Mobilenet V3 and the proposed model is 3%. According to the F1 score, the F1 score of the proposed model is higher than that of Alexnet, VGG16, and MobileNet V3 models by 21%, 18%, and 3% respectively. Therefore, the proposed method has the best recognition effect in the test set of maize and wheat seedlings and weeds in the field. Because the proposed method introduces the self-attention mechanism, the self-attention mechanism analyzes the correlation between vectors, gives different weights to feature layers, and calculates the weighting of feature channels, to improve the representation ability of the model; this allows the model to pay more attention to the target features.

4 Conclusions and prospect

In this paper, a recognition model based on the UNet network model and ViT classification algorithm is proposed. First, the UNet network model is used to separate maize and wheat seedlings and weeds, in different growth stages, from complex backgrounds. Then, the PIL algorithm is used to extract the segmented green plant leaf features. Finally, the image features are input into the Vision Transformer model to identify and classify maize and wheat seedlings and weeds in the field. Experiments show that the average accuracy, average precision, average recall, and F1-score of the proposed model are 97%, 97%,97%, and 97%, respectively. This method not only effectively improves the recognition accuracy of maize and wheat seedlings and weeds in the field, but also improves the recognition speed of the model. Compared with other existing research results, this model shows better performance on the test set than other models, which provides effective information support for pesticide spraying and mechanical weeding.

In the current study, Although the proposed recognition model has high accuracy, it also has some insufficient. Compared with Alexnet, VGG16, and MobilenetV3 models, the parameters are 5.8 times that of Alexnet, 1.2 times that of VGG16, and 20 times that of Mobilenet V3. Compared with Alexnet and VGG16, the recognition speed of the proposed model is 4 times and 1.6 times slower, but it is 1.3 times faster than Mobilenet V3. The data show that the proposed model is larger and has more parameters, so the process of recognition, takes up a lot of computing resources and the recognition speed is relatively long. In future work, we will plan to study the following two aspects: (1) Building a lightweight model. To solve the problem that the existing model has many parameters and occupies high computational resources, model pruning, and fine-tuning technology are used to adjust the model weights, and all the weights close to 0 are set to 0 until the parameters reach the target sparsity, the calculation amount is reduced. Based on ensuring the recognition performance, a model with a similar size to the existing mobile terminal network is obtained. (2) Model deployment and application. The proposed model will be deployed on Android, and users can identify crops and weeds by taking photos, making weed control simple and efficient.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Giselsson T M, Jørgensen R N, Jensen P K, et al. A public image database for benchmark of plant seedling classification al-gorithms[J]. https://arxiv.org/abs/1711.05458.

Author contributions

JY: Conceptualization, supervision, writing—review, and editing. XG: Methodology, writing—original draft. YG: Software, resources, visualization. FL: Validation. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Hebei North College Provincial Universities Infrastructure Scientific Research Business Fee Project under Grant JYT2022021, the General Project of Hebei North University under Grant XJ2021005, and the Hebei Province Population Health Information Technology Innovation Center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbas, T., Zahir, Z. A., Naveed, M., and Kremer, R. J. (2018). Limitations of existing weed control practices necessitate development of alternative techniques based on biological approaches. *Adv. Agron.* 147, 239–280.

Ahmad, A., Saraswat, D., Aggarwal, V., Etienne, A., and Hancock, B. (2021). Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. *Comput. Electron. Agric.* 184, 106081. doi:10.1016/j.compag.2021.106081

Bah, M. D., Hafiane, A., and Canals, R. (2018). Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sens.* 10 (11), 1690. doi:10.3390/rs10111690

Basit, A., Irshad, M., Salman, M., Abbas, M., and Hanan, A. (2019). Population dynamics of weeds (canary grass, broad leave and wild oats), aphid and abiotic factors in association with wheat production in southern Punjab: Pakistan. *J. Appl. Microb. Res.* 2, 17–23.

Chen, W., Su, L., Chen, X., and Huang, Z. (2023). Rock image classification using deep residual neural network with transfer learning. *Front. Earth Sci.* 10, 1079447. doi:10.3389/feart.2022.1079447

Deng, B., Liu, P., Chin, R. J., Kumar, P., Jiang, C., Xiang, Y., et al. (2022). Hybrid metaheuristic machine learning approach for water level prediction: A case study in dongting lake. *Front. Earth Sci.* 1545. doi:10.3389/feart.2022.928052

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Du, Y., Zhang, G., Tsang, D., and Jawed, M. K. (2022). "Deep-cnn based robotic multi-class under-canopy weed control in precision farming," in 2022 International Conference on Robotics and Automation (ICRA) (IEEE), 2273–2279.

Ebtehaj, I., and Bonakdari, H. (2022). A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in Quebec, Canada. *J. Hydrol.* 614, 128592. doi:10.1016/j.jhydrol.2022.128592

Fu, L., Lv, X., Wu, Q., and Pei, C. (2020). Field weed recognition based on an improved VGG with inception module. *Int. J. Agric. Environ. Inf. Syst. (IJAEIS)* 11 (2), 1–13. doi:10.4018/ijaeis.2020040101

Gharde, Y., Singh, P. K., Dubey, R. P., and Gupta, P. K. (2018). Assessment of yield and economic losses in agriculture due to weeds in India. *Crop Prot.* 107, 12–18. doi:10.1016/j.cropro.2018.01.007

Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., Dyrmann, M., and Midtiby, H. S. (2017). A public image database for benchmark of plant seedling classification algorithms. arXiv preprint arXiv:1711.05458.

Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

Jiang, H., Zhang, C., Qiao, Y., Zhang, Z., Zhang, W., and Song, C. (2020). CNN feature based graph convolutional network for weed and crop recognition in smart farming. *Comput. Electron. Agric.* 174, 105450. doi:10.1016/j.compag.2020.105450

Louargant, M., Jones, G., Faroux, R., Paoli, J. N., Maillot, T., Gée, C., et al. (2018). Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information. *Remote Sens.* 10 (5), 761. doi:10.3390/rs10050761

Luo, S., Zhang, M., Nie, Y., Jia, X., Cao, R., Zhu, M., et al. (2022). Forecasting of monthly precipitation based on ensemble empirical mode decomposition and Bayesian model averaging. *Front. Earth Sci.* 10, 926067. doi:10.3389/feart.2022.926067

Ma, N., Zhang, X., Liu, M., and Sun, J. (2021). "Activate or not: Learning customized activation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8032–8042.

Manisankar, G., Ghosh, P., Malik, G. C., and Banerjee, M. (2022). Recent trends in chemical weed management: A review. *Pharma. Innov.* 11 (4), 745–753.

Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., and Rodríguez, L. (2020). A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering* 2 (3), 471–488. doi:10.3390/agriengineering2030032

Pei, H., Sun, Y., Huang, H., Zhang, W., Sheng, J., and Zhang, Z. (2022). Weed detection in maize fields by UAV images based on brop row preprocessing and improved YOLOv4. *Agriculture* 12 (7), 975.

Ranganathan, G. (2021). A study to find facts behind preprocessing on deep learning algorithms. *J. Innovative Image Process. (JIIP)* 3 (01), 66–74. doi:10.36548/jiip.2021.1.006

Tannouche, A., Gaga, A., Boutalline, M., and Belhouideg, S. (2022). Weeds detection efficiency through different convolutional neural networks technology. *Int. J. Electr. Comput. Eng.* 12 (1), 1048. doi:10.11591/ijece.v12i1.pp1048-1055

Venkataraju, A., Arumugam, D., Stepan, C., Kiran, R., and Peters, T. (2022). A review of machine learning techniques for identifying weeds in corn. *Smart Agric. Technol.* 3, 100102. doi:10.1016/j.atech.2022.100102

Wang, C., Wu, X., and Li, Z. (2018). Recognition of maize and weed based on multi-scale hierarchical features extracted by convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* 34 (5), 144–151.

Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 158, 226–240. doi:10.1016/j.compag.2019.02.005

Wang, A., Xu, Y., Wei, X., and Cui, B. (2020). Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access* 8, 81724–81734. doi:10.1109/access.2020.2991354

Wang, W., Tan, X., Zhang, P., and Wang, X. (2022). A CBAM based multiscale transformer fusion approach for remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 15, 6817–6825. doi:10.1109/jstars.2022.3198517

Wu, Z., Chen, Y., Zhao, B., Kang, X., and Ding, Y. (2021). Review of weed detection methods based on computer vision. *Sensors* 21 (11), 3647. doi:10.3390/s21113647