Plant Seedling Classification

Heqian Lu
ECE department
University of California, San Diego
La Jolla, CA
h7lu@ucsd.edu

Zhenkai Lai
ECE department
University of California, San Diego
La Jolla, CA
zhlai@ucsd.edu

Jiawei Duan
ECE department
University of California, San Diego
La Jolla, CA
jduan@ucsd.edu

***Abstract*—The classification of seedling is important in the agriculture and botany filed.Therefore in this project, we propose several classification methods to predict the plant sort result from the images of test data which include CNN(Convolutional Neural Network),ResNet(Residual Networks),MobileNet and VGG models.Through comparing the results and performances of different models with the same train set, we could conclude one of the best suitable for this model and data set. And in this project, how we process the data, tune the parameters of the model, construct the layers in CNN models is more important for us to learn more about computer vision field.**

***Index Terms*—Deeping learning;Plant seedling classification; weed control;RGB extraction;convolutional neural network;transfer learning.**

I. Introduction

Precision agriculture has the potential to revolutionize modern farming by tailoring treatments to match the variations and local properties of the field. For example, mapping local properties like crops to weeds ratio can help eliminate weeds at early stage, which brings benefits both economically and environmentally. To measure and quantify local properties across the farmland, data acquisition and reliable analysis plays a crucial role.

Due to the importance and the performance requirement of plant classification, computer aided methods were introduced both to improve classification accuracy and save human resources. Studies on plant classification using computer vision methods date back in 1995, Brendel and Schwanke [1] implement knowledge-based object recognition to classify leaves. By transforming features used by botanists to code manually, it can get some good results but is very limited in number of classes, Which greatly impacted its application scenario. Moreover, feature conversion needed to be done manually also fails the algorithm in expandability. To minimize human interference in feature extraction and make the whole process automatically, dimension reduction methods was proposed and it significantly improve the procedure of feature selection, as well as the expandability, but still very limited in both versatility and robustness. Golzaria and Frick [2] did some research in distinguishing three grass species using PCA method, but still cannot achieve acceptable results in large size of species. Development of machine learning gives possibilities to increase the performance of image classification, as well as plant seedling classification. Statistical and neural network [3] are used to classify five different cereal grains and achieve over 96% accuracy, which shows the potential of such methods. Hinton's [4] great success in ImageNet 2012 greatly stimulates the development of using deep learning methods in image classification and provides foundation for using deep learning methods in plant seedling classification. Later researches proved the performance of deep learning, Giselsson [5] can achieve 94.8% accuracy in 8 species classification using pure CNN, and till now Ashcar [6] in 2019 can achieve a shockingly 99.48% accuracy in 12 species classification using VGG.

To better understand different methods of deep learning and validate previous work, self-built CNN, VGG, ResNet, DenseNet and MobileNet are tested. Results shows that all methods can give great results on a 12 species plant seedling classification with minor modification, and MobileNet achieves a great balance in accuracy and time consumption when deployed.

II. Related Work

Significant works [7] have been done for plant image classification tasks, including classification between maize plants and weeds with 44580 segmented images at early growth with obtained training accuracy of 97.23%.

For plant disease classification in wild conditions, some researchers focus on automatic disease detection system like Johannes. The system [8] achieved excellent results for early recognition of three wheat diseases. An analysis was carried out using seven handheld devices of 3500 images across Spain and Germany.

For plant classification, a research [9] using CNN to identify plant by extracting vein morphological patterns, can get an accuracy of 95%.

For plant seedling classification, many researchers are using dataset [10] provided by Aarhus University, Department of Engineering Signal Processing Group. Benefited from Kaggle which host the dataset for competition, plenty of jobs have achieved great results. As we have mentioned earlier, Ashcar got the best results using VGG with an accuracy of 99.48%.

Deep learning's success in different kinds of plant classification testifies its versatility and expandability, not only the more recent methods like VGG can achieve good results, pure CNN can also performs good in some kind of task with finely tuned parameters.

III. Dataset

- **data exploration**

The dataset of this project can be downloaded from Kaggle Competition which contains images of about 960 unique plants belonging to 12 species at several growth stages.And also the database have been recorded Aarhuas University Flakkebjerg Research station in a collaboration between Southern Denmark and Aarhus University.This can be found here:https://www.kaggle.com/c/plant-seedlings-classification/data. This dataset contains 5,539 images of crop

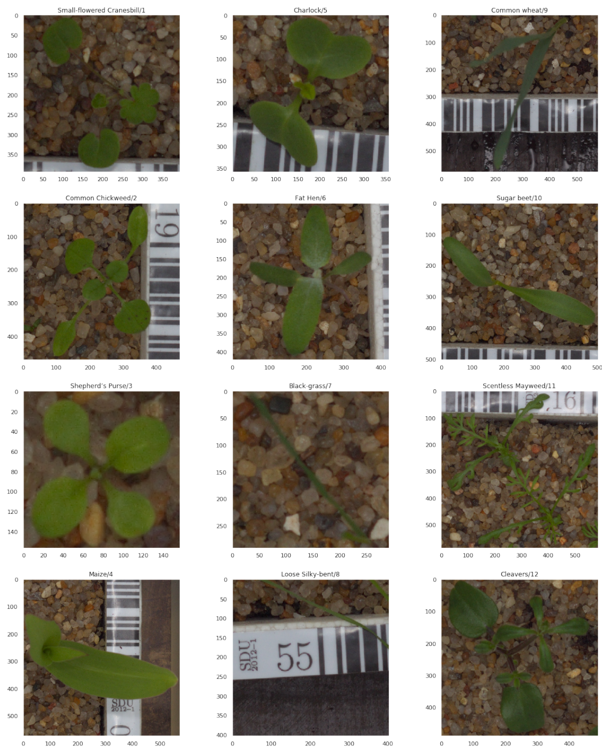


Fig. 1: data samples

and weed seedlings. The images are grouped into 12 classes as shown in the above pictures. Each class contains rgb images that show plants at different growth stages. The images are in various sizes and are in png format.So in our project data preprocess, we will change the size of picture and normalize to fit the model and convert the pixels to matrices.

- **Data preprocessing**

1.Resizing images
Images in the data set are not always the same, so we have to resize the size of images to 224*224 to feed it to the neural network.
2. Create mask for the images
Create masks means that we return the matrix with shape height and width of the original images,and in this matrix there are only 0 and 1 values. The 1 values define the interesting part of the original image. We create this mask using HSV of the image. The HSV color-space is suitable for color detection because with the Hue we can define the color and the saturation and value will define ”different kinds” of the color. (For example it will detect the red, darker red, lighter red too). We cannot do this with the original RGB color space.

RGB

- The RGB is an additive color space.
- Creating any color by mixing the three primitive additives Red, Green and Blue.
- A pixel with 3 channels contains information of (red, green, blue) light with values between 0 and 255.
- Given all values equal to 0 maps to black, all values 255 maps to white.

HSV

- Colors of each hue follow the radius on the circle. It's values have an angular dimension starting with the red primary at 0°, passing the green primary at 120° and the blue primary at 240° and finally ending up with red again at 360°.
- The vertical axis describes the gray scale ranging from black (0) to white (1).
- The horizontal axis covers the saturation of the color that can be reduced by tinting with white from 1 to 0. [11]

3.**Morphological operations**
one of most common morphological operation is closing : closing used to close the small halls in the images. This figure below illustrate the image before and after applying closing:
4.**Sharpening**

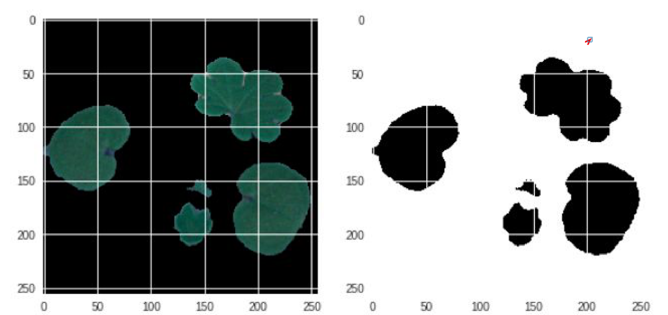


Fig. 2: after morphological operations

Sharpening an image increases the contrast between bright and dark regions to bring out features. It just looks like make the texture of the image more visible. Image before and after sharpening:
5.**Label encoding**
Using **LabelBinarizer** Binarize labels in a one-vs-all fashion
Input: the label of image and the **output** is vector represent the class in binary form
6.**Split data into training and testing set**
Splitting 70% for training and 30% for testing.Meanwhile, splitting testing data half for testing and half for validation.We use cross-validation to prevent the overfitting and check if it

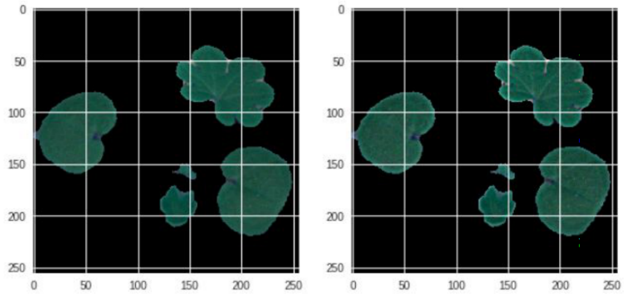


Fig. 3: after Sharpening

will happen.

7.**Data augmentation**

The data augmentation will be applied when the distribution of our data is not quite uniform or the number of samples is not very lagre to feed the model. Through crop,spinning,color or something else to generate more images based on the original data,the sample images rotation,size,width,height, hrozontal flip,vertical flip are randomly transformed and then the size of the training data could be larger than before.

IV. Methods

In this project, we use several model(classifiers): self-built CNN model, ResNet model,MobileNet model.The CNN model is a state-of-the-art alogorithm of most image processing tasks, which is classification for different kinds of plant seeding in this task.CNN's are concurrently employed in agriculture mainly, for recognition and classification tasks [3], and have been proven to provide superior results.It needs a large amount of training data compared to other approaches; and this dataset we have been provided are already sufficient enough to fit this criteria.

A. *Convolutional Neural Network*

The Convolutional Neural Network(CNN) is a deep learning algorithm and consists of an input layer, hidden layers,and an output layer.

The original seedling images are all equally resized to 224x224pixels(The size of images can be resized empirically so as to get more satisfactory performance and fit to the input layer).And the hidden layers consist of a lot of layers which will be illustrated in the following table: Following are the layers which are used to construct the CNN model:

- **CONV layer**

The CONV layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. In this project, we choose 32,64,128 filters for the Conv layers.

- **POOL layer**

Pool layer performs downsampling operation along the spatial dimension(width,height), outputting a reduced volume than the

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 224, 224, 32)	2432
conv2d_2 (Conv2D)	(None, 224, 224, 32)	25632
max_pooling2d_1 (MaxPooling2	(None, 112, 112, 32)	0
dropout_1 (Dropout)	(None, 112, 112, 32)	0
conv2d_3 (Conv2D)	(None, 112, 112, 64)	18496
conv2d_4 (Conv2D)	(None, 112, 112, 64)	36928
max_pooling2d_2 (MaxPooling2	(None, 56, 56, 64)	0
dropout_2 (Dropout)	(None, 56, 56, 64)	0
conv2d_5 (Conv2D)	(None, 56, 56, 128)	73856
conv2d_6 (Conv2D)	(None, 56, 56, 128)	147584
max_pooling2d_3 (MaxPooling2	(None, 28, 28, 128)	0
dropout_3 (Dropout)	(None, 28, 28, 128)	0
global_max_pooling2d_1 (Glob	(None, 128)	0
dense_1 (Dense)	(None, 256)	33024
dropout_4 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 12)	3084

Fig. 4: summary of layers in CNN.

previous layer.These are used to reduce computational cost and to some extent also reduce overfitting.

- **Dense layer**

Dense layer is also called fully-connected layer, each neuron will be connected to all the neurons of the previous layer. Each of the nodes of dense layer outputs a score corresponding to a class score.

- **Dropout layer**

Dropout layer is used as a method of regularization of combat over-fitting of the training set. It 'drops' randomly neurons(setting their weights to zero), resulting in a simpler version of the CNN for each iteration and hence giving the model a hard time to overfit.

B. *Transfer Learning*

Transfer learning is a machine learning technique where a model trained on one task is repurposed on a second related task.In transfer learning, we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. This process will tend to work if the features are general, meaning suitable to both base and target tasks, instead of specific to the base task. There are two common approaches as belows.

- Develope Model Approach
- Pre-trained Model Approach

In our project, we choose the second approach and use four pre-trained models: ResNet, DenseNet, VGG, MobileNet. Below we will briefly introduce these four models.

- **VGG** [12]

VGG mainly contains 16 or 19 layers, including fully-connected layers, convolutional layers, max-pool layers. Its kernel size is 3 and max-pool size is 2.

- **ResNet** [13]

ResNet is composed of a series of residual blocks, which can be divided into two parts, direct mapping part and residual part. Its general expression is :

$$x_{l+1} = x_l + F(x_l)$$

- **DenseNet** [14]

DenseNet is composed of a series of dense blocks.For each layer, the feature maps of all previous layers are directly used as input for this layer.ResNet adds features directly through the "Summation" operation, impeding the information flow in the network to a certain extent. DenseNet combines feature maps through concatenate operations, and each layer is related to other layers, which maximizes the flow of information.Its general expression is :

$$x_{l+1} = F(x_0, x_1 ... x_l)$$

- **MobileNet** [15]

The MobileNet model is based on depthwise separable convolutions which factorize a standard convolution into a depthwise convolution and a pointwise convolution.

V. Results

A. *Model evaluation and confusion matrix*

As shown in the dataset section, we split data into training data, validation data and testing data to train and evaluate the model.

- **accuracy and loss**

Loss function is categorical crossentropy.Suppose there are N samples and K labels. y is the true labels, p is the probability of the prediction and L is the crossentropy.

$$L(y,p) = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{K-1} y_{i,j} log(p_{i,j})$$

Accuracy is a common metric for classifiers; it takes into account all true prediction with equal weight.

$$accuracy = \frac{true\ prediction}{dataset\ size}$$

- **prediction time**

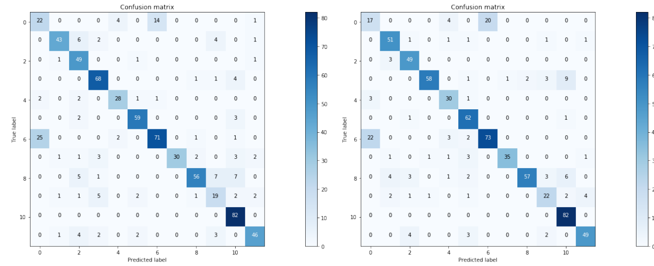
we use the prediction time of the test data to evaluate the prediction speed of our models.
Here is the summary of all models we used.

TABLE I: Results summary

Model	*CNN*	*Data augmentation*	*ResNet*	*MobileNet*	*DenseNet*	*VGG*
Accuracy	0.7921	0.8230	0.9059	0.8947	0.8610	0.8722
Loss	0.4416	0.5209	0.2601	0.3214	0.4115	0.8290
Time	0.5594	0.8207	4.5817	1.0843	3.7533	2.5685

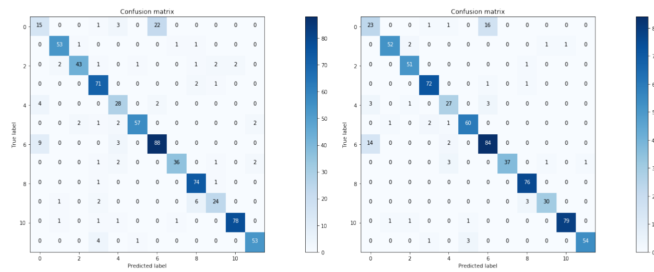
- **confusion matrix**

A confusion matrix C is such that Ci,j is equal to the number of observations known to be in group i but predicted to be in group j.The confusion matrix could be very helpful to see the model drawbacks on classes of plant. Fig.5 - Fig.6 is the confusion matrixs of our models.

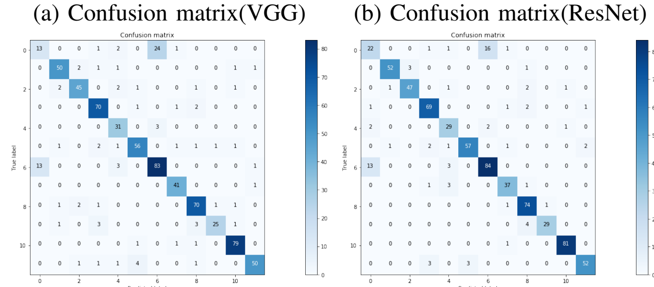


(a) Without augmentation (b) With augmentation

Fig. 5: self-built CNN model confusion matrix



(a) Confusion matrix(VGG) (b) Confusion matrix(ResNet)



(c) Confusion matrix(DenseNet) (d) Confusion matrix(MobileNet)

Fig. 6: Confusion matrices of different models using transfer learning

VI. Conclusion

First we try the self-built CNN model illustrated above, while we found that the accuracy is not very ideal. Then we try the ResNet model to improve the accuracy and get the expected result. Further, we want to decrease computation time and apply MobileNet model to the dataset. Computation time reduce a lot as we expect and the accuracy still remains high. To have a better understanding of the performance, we try the other two models listed in the above sections.In general, we prefer MobileNet model with high accuracy and low latency. According to the results shown in the above section, the prediction of the model is acceptable for the application, and from the confusion matrix, we could know more about the precision of the prediction. This model could be applied to help farmers to automatically classify the seedling plants and weed plants.

VII. Contributions

TABLE II: Team member contribution

Member name	***Contribution***
Jiawei Duan	load data and process data,study background and related work
Heqian Lu	Build and evaluate CNN model
Zhenkai Lai	Build and evaluate other models

VIII. References

References

[1] T. Brendel, J. Schwanke, P. F. Jensch, and R. Megnet, "Knowledge-based object recognition for different morphological classes of plants," in *Optics in Agriculture, Forestry, and Biological Processing*, vol. 2345, pp. 277–284, International Society for Optics and Photonics, 1995.

[2] E. Neilson, A. Edwards, C. Blomstedt, B. Berger, B. Møller, and R. Gleadow, "Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a c4 cereal crop plant to nitrogen and water deficiency over time," 2015.

[3] X. Luo, D. Jayas, and S. Symons, "Comparison of statistical and neural network methods for classifying cereal grains using machine vision," *Transactions of the ASAE*, vol. 42, no. 2, p. 413, 1999.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[5] T. M. Giselsson, H. S. Midtiby, and R. N. Jørgensen, "Seedling discrimination with shape features derived from a distance transform," *Sensors*, vol. 13, no. 5, pp. 5585–5602, 2013.

[6] B. A. Ashqar, B. S. Abu-Nasser, and S. S. Abu-Naser, "Plant seedlings classification using deep learning," 2019.

[7] C.-C. Andrea, B. B. M. Daniel, and J. B. J. Misael, "Precise weed and maize classification through convolutional neuronal networks," in *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, pp. 1–6, IEEE, 2017.

[8] A. Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, and A. Ortiz-Barredo, "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case," *Computers and electronics in agriculture*, vol. 138, pp. 200–209, 2017.

[9] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, 2016.

[10] S. Skovsen, M. Dyrmann, A. K. Mortensen, K. A. Steen, O. Green, J. Eriksen, R. Gislum, R. N. Jørgensen, and H. Karstoft, "Estimation of the botanical composition of clover-grass leys from rgb images using data simulation and fully convolutional neural networks," *Sensors*, vol. 17, no. 12, p. 2930, 2017.

[11] *Computer vision with seedlings*, 2020. https://www.kaggle.com/allunia/computer-vision-with-seedlings.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

Critique Response

critique review from group 86

Q1: How did MobileNet come up as your final choice as different model have different advantages? In another words, how did you weight each factor up? (Accuracy, Loss, and Time)

A: The least prediction time among models is 0.82s, which is predicted by CNN model, while its accuracy is 0.82. The best accuracy among models is 0.91, which is predicted by ResNet model, while its prediction time is 4.6s. As for MobileNet model, we can see that its accuracy is 0.89 and its prediction time is 1.1s. Compared with the other two model, we can conclude that MobileNet model improve shortcomings of other models and doesn't affect their advantages. That's why we choose MobileNet model as our final choice.

critique review from group 65

Q2: Adding some traditional machine learning methods for comparison might be better to outstand the neural network model in computer vision. In the presentation, I can only see the one self-built CNN and two transfer learning models applied.

A: In images classification which is also shown in the related work, CNN has achieve the best result currently than other traditional work. we think that traditional machine learning method can't achieve as accurate results as CNN models. That's why we only choose the CNN model.

Q3: It might be better to have a detailed explanation of the bird view of data before starting the data preprocessing step, such as explosive data analysis (EDA), and finally, discuss how the dataset quality affects the model performance.

A:As we can see in the report, the data augmentation could be seen as improving the quality of dataset which make the whole dataset more uniformly and sufficient to robust the model.That is how the dataset quality afftects the model performance in our project.

Q4: plant seeding classification should not be the final goal of the project, it would be great if they can discuss more the application of this plant seeding classification, and how this can be improved.

A:Actually, the plant seeding classification is just the application which can be realized and used in the agriculture and botany which are shown in the introduction part work, and to improve the performance based on the current loss and accuracy, I think we could add more clean and suitable data or apply data augmentation to transfer learning models.

critique review from group 74

Q5: The presentation didn't illustrate how the dataset is augmented, specific augmentation way may be added at report. And is the ResNet, VGG and following models all use the augmented dataset?

A: We use the image generator to conduct data augmentation. Only the self-built Model use the augmented dataset.

Q6: At the literature review part, you can write the method name like 'VGG' instead of using the author's name. And the image resolution is also a very important attributes, I think you should add a column to specify what resolution is used in those articles if possible. The comparison between results using different resolution may be less persuasive.

A: Thanks for pointing out, we did not mention that in the presentation as time is limited, and some of the references did not cover this info.

Q7: Is it a self-designed CNN model? If so, a more detailed explanation about the model's structure is preferred.

A: Yes. We include more details about the model in the report. Because of the limited presentation time, we didn't talk much about the details.