



Group 16 Plant Seedling Classification

Qinyan Li, Yunzhe Hu, Qian Wang, Zhanghan Liu
 {q3li, yuhu, qi018, zh1031}@ucsd.edu

Predicting

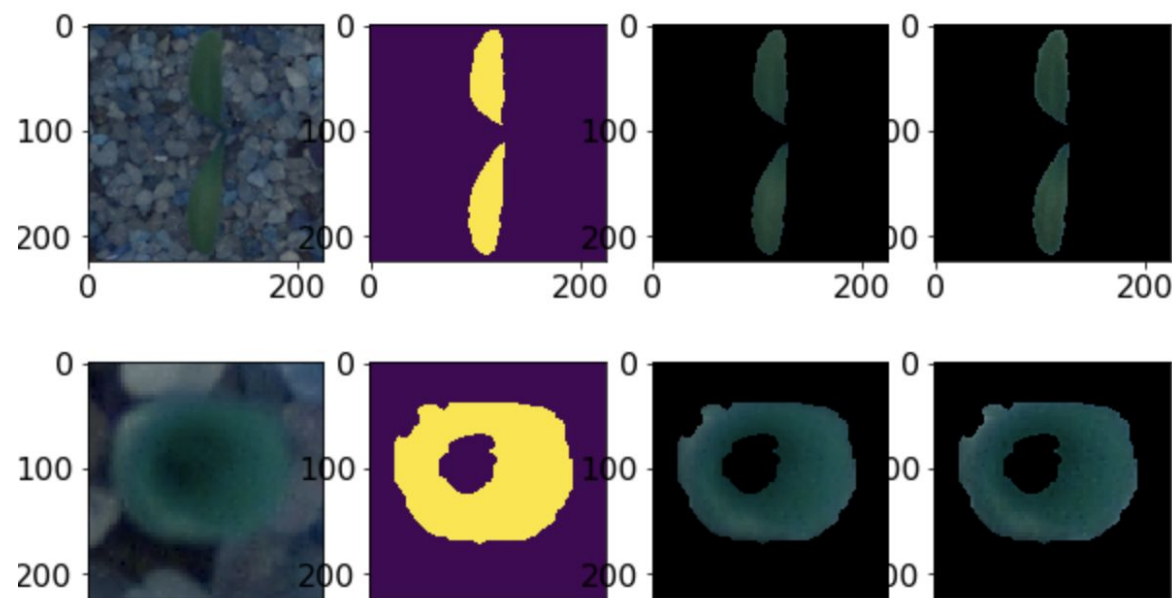
In this project, we will participate in a Kaggle competition and aim to define a classifier that can identify the species of a plant given a seed image. The problem here is the weed seedling is much like crop seedling and our goal is to be able to differentiate between them. The output of classification should be in the format as follows:

File	Species
0021e9014.png	Sugar Beet
026719f9b.png	Charlock
...	...

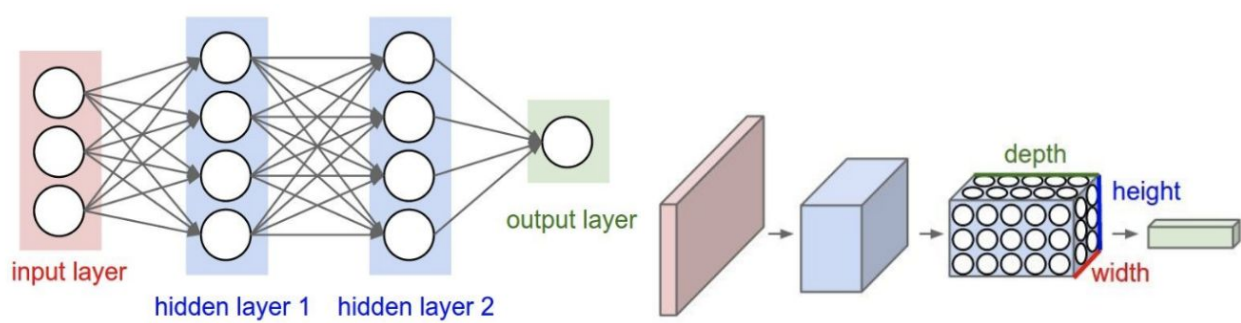
Features

First, we do data preprocessing based on raw images:

- Data augmentation
- Image segmentation



Then, we use deep neural networks to extract features automatically.



Data

The Dataset have been recorded at Aarhus University Flakkebjerg Research station in a collaboration between University of Southern Denmark and Aarhus University. The dataset:

- Contains 960 unique plants belongs to 12 species at different growth stages
- Comprises annotated RGB images with physical resolution of about 10 pixels per mm.
- Has totally 4750 labeled examples.

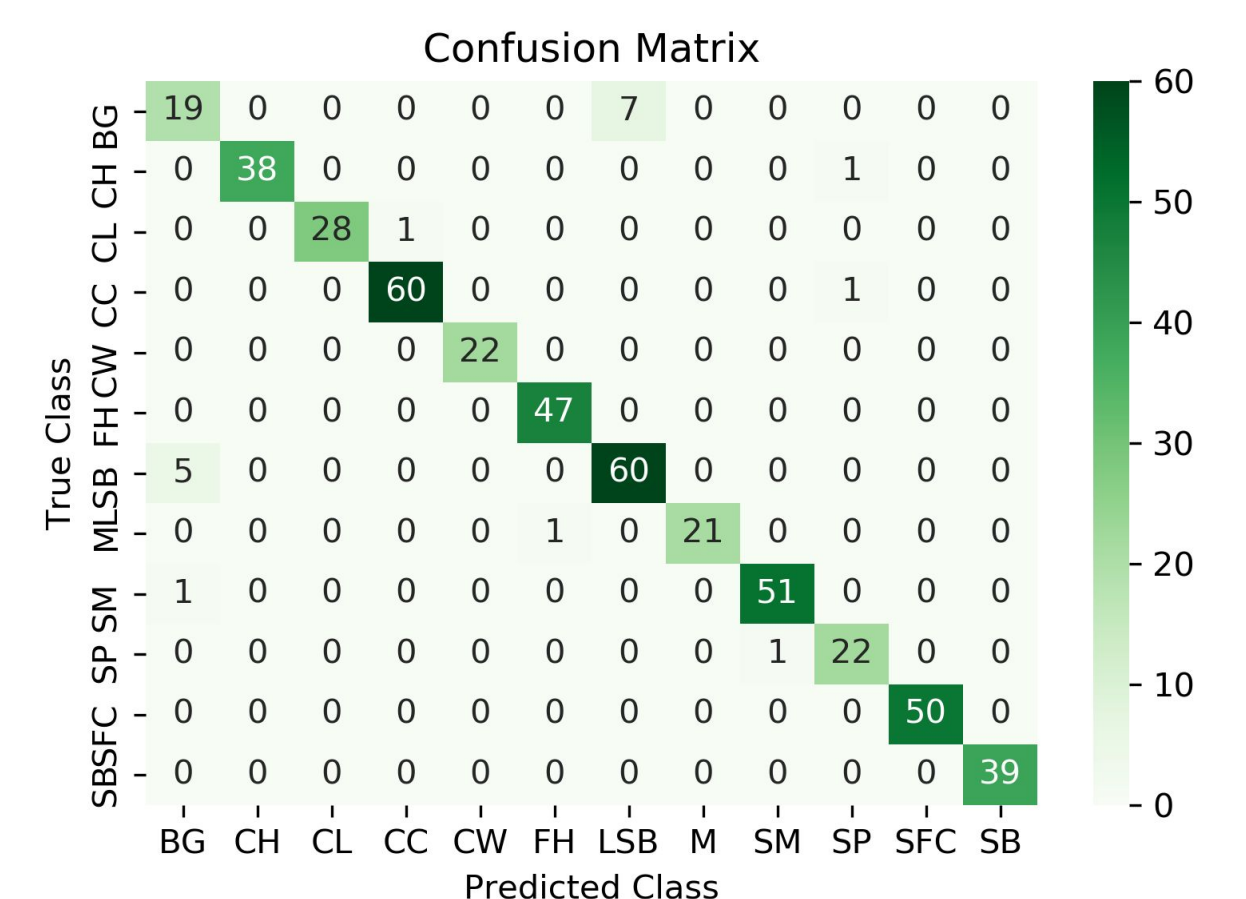
Models

- Resnet50 (pretrained network)
 - Logistic regression
- VGG19 (pretrained network)
 - Logistic regression
- Xception (pretrained network)
 - + Logistic regression
 - + 2-layer fully-connected network
- Customized CNN
 - 6 Convolution Layers
 - 8 Batch Normalization Layers
 - 3 Fully-Connected Layers
 - 5 Dropout Layers

Results

Model	Train accuracy	Test accuracy
Resnet50+LR	0.5522	0.5
VGG19+LR	0.8678	0.8510
Xception+LR	0.9989	0.8813
Xception+FCN	1.0	0.8906
6 Conv layers+3 FC layers	0.9995	0.9621

The confusion matrix for the customized CNN is:

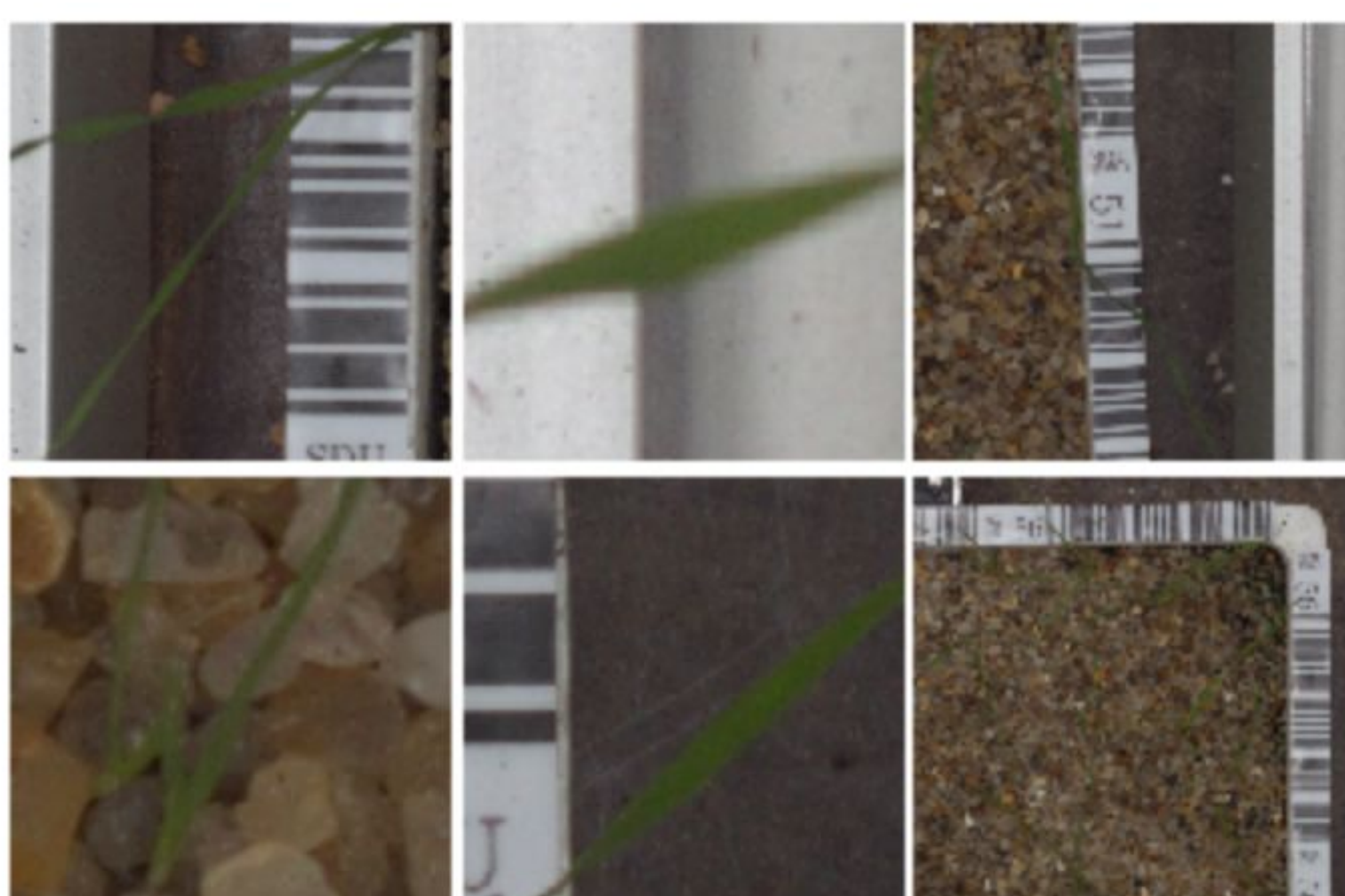


Future

- The convolutional neural network we build now is not deep, but the training time for this model is already 4 hours. If we could have time to train a much deeper neural network, the test results should be much better.
- We could try more combinations of the data preprocessing methods.

Discussion

- Using pretrained model can produce a pretty well result without spending much time on training the network. Training a shallow CNN by ourselves improves the accuracy by 7 percent, but it took 4 hours.
- When using the same pretrained model, classification algorithms will not affect the result much.
- From the confusion matrix, we can observe that the model has difficulty in differentiating Black-grass from Loss Silky-bent. As we visualize their seedlings, we can see that they are very similar, therefore, this result is plausible.



Black-grass

Loose Silky-bent

References

- "Plant Seedlings Classification," Kaggle, 12-Mar-2018. [Online]. Available: <https://www.kaggle.com/c/plant-seedlings-classification>. [Accessed: 11-May-2019].
- Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., Dyrmann, M., & Midtby, H. S. (2017). A public image database for benchmark of plant seedling classification algorithms. arXiv preprint arXiv:1711.05458.