Classification on Fine-Grained Plant Pathology Image Dataset

Zhengyang Wei [1] Ranchao Yang [1]

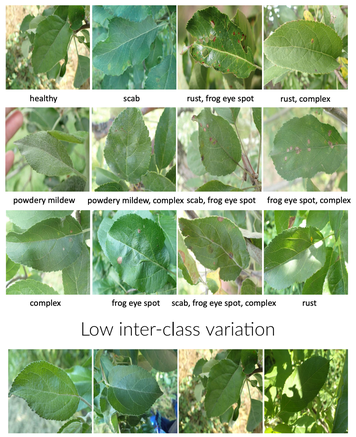
[1]ICME

Introduction

- With high-accuracy plant pathology classification, farmers can avoid the loss of yield of agricultural production. However, current human scouting for plant pathology is expensive and inefficient.
- Image-based plant pathology classification is usually regarded as a **fine-grained visual classification (FGVC)** problem.

FGVC task is very challenging:

- **High intra-class variation** caused by different ages of infected tissues, different light conditions, and genetic variations.
- **Low inter-class variation** due to the fact that the images are all similar leaves of plants.



Low inter-class variation

High intra-class variation

Figure 1. Examples from Dataset

A variety of methods have been proposed in the past years on fine-grained visual classification:

- **Feature-encoding methods** aim to extract fine-grained features by encoding higher order information on features.
- **Region localization methods** have also been used widely with weakly supervised learning. They are used to detect discriminative regions in images.

Several powerful attention blocks are proposed to enhance classification performance:

- Squeeze-and-Excitation (SE) block
- Self-attention mechanism (SAM)
- Convolutional Block Attention Module (CBAM)

Dataset

We use Plant Pathology Challenge 2020 dataset as our dataset. It consists of 18632 images of plant leaves classified into 12 categories. Some examples are shown in Fig. 1. The resolution of the images is 448×448. We split the dataset into training, validation and test sets. Our data augmentation methods include :

- randomly cropping
- randomly horizontal flipping (with the default probability parameter 0.5)
- randomly changing brightness (with brightness parameter 0.126)
- randomly changing saturation (with saturation parameter 0.5)

Methods

Our framework is illustrated in Fig. 3.

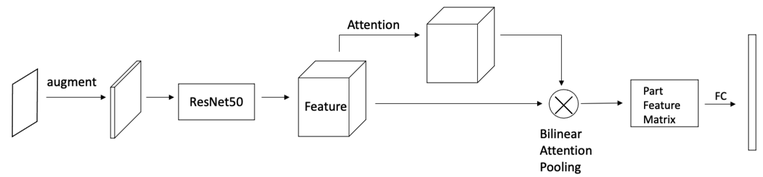


Figure 2. Model architecture.

Feature and Attention Extraction

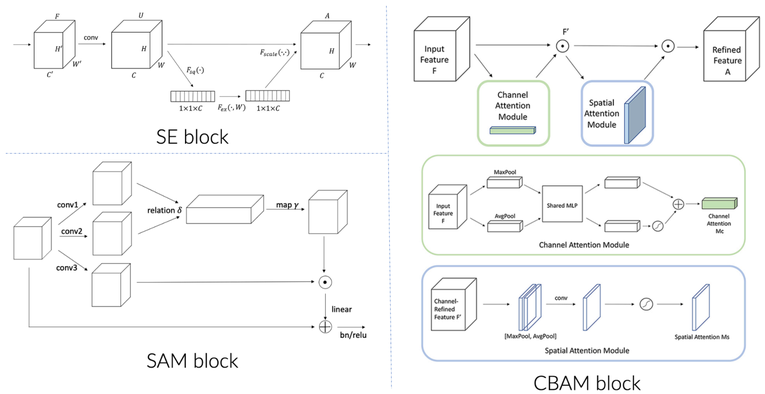


Figure 3. Attention Block.

Bilinear Attention Pooling

The features of the object are represented by a part feature matrix $P \in R^{M \times N}$ which is achieved by stacking these part features f_k. Let $\Gamma(A, F)$ indicate bilinear attention pooling between an attention map A and a feature map F. It can be represented as follows:

$$P = \Gamma(A, F) = \begin{pmatrix} g(a_1 \odot F) \\ g(a_2 \odot F) \\ \dots \\ g(a_M \odot F) \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_M \end{pmatrix} \quad (1)$$

where $g(\cdot)$ is the additional feature extraction function that extracting discriminative local features.

Training with Attention-guided Data Augmentation

In the training process, for each training image, we randomly choose one attention map A_k to guide the data augmentation, and normalize it as the k_{th} augmentation map $A_k^* \in R^{H \times W}$. Then, with augmentation map A_k^*, we can first get the crop mask C_k from A_k^*.

$$C_k(i,j) = \begin{cases} 1, & \text{if } A_k^*(i,j) > \theta_c \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

We then find a bounding box that can cover the whole selected positive region of C_k. We enlarge this region from the raw image and use it as our augmented input data. We can also obtain attention drop mask D_k from A_k^*.

Test with Object Localization and Refinement

In the testing process, after the model outputs the coarse-stage classification result and the corresponding attention maps for the raw image, we can predict the entire region.

Experiments Analysis

The accuracy of our baseline models and improved models

Model	Backbone	Accuracy
Resnet50	Resnet50	88.34
WSDAN	Resnet50	89.79
WSDAN + SE	Resnet50	89.84
WSDAN + SAM	Resnet50	**90.03**
WSDAN + CBAM	Resnet50	89.90

Table 1. Accuracy comparison.

Attention Map Visualization

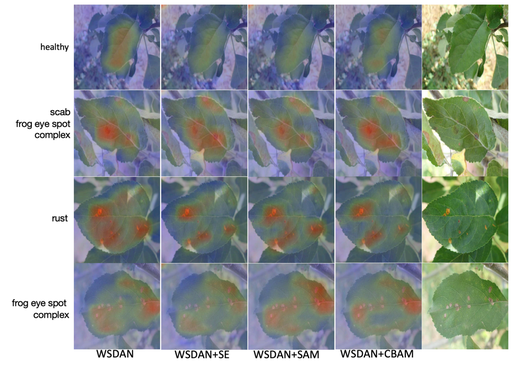


Figure 4. Examples of attention maps for each of the four models.The rightmost column is the original image.

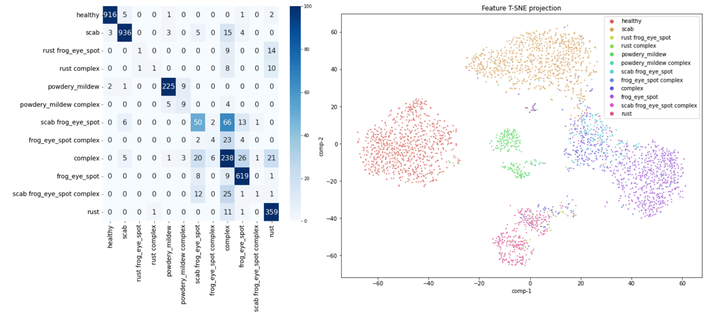


Figure 5. Confusion Matrix and t-SNE of WSDAN + SAM.

Conclusion

- Contribution:
  - Investigate models to predict plant diseases based on images of the leaves
  - Apply ResNet50 and WSDAN as baselines to our dataset and experiment with three different attention blocks
  - Find that WSDAN+SAM performs the best on our dataset and achieves the highest accuracy.
- Limitation:
  - The results in earlier sections might be sub-optimal
  - Our dataset is highly imbalanced: some of the classes have a large number of training data while others have only a few
- Future work: We would like to adjust for the data imbalance through resampling methods. We would also like to optimize the models we have experimented with and explore other models such as transformers which might improve model performance on fine-grained visual classification tasks.

zywei@stanford.edu audreyrc@stanford.edu