

Project 6 – Plant Pathology : 2020

Team1

배병규 조재효 박주환 전용우

1. DataSet : Kaggle “Plant Pathology : 2020”

2. Data Analysis

- Correct Train Set Errors

3. Model Selection

IN Papers

- Image Size : 224 * 224
- Random Weight
- Categorical Cross Entropy
- Adam Optimizer : Learning rate 0.05
- Global Average Pooling
- Relu
- Inception

In First ranked team's Kaggle

- Augmentation : Brightness, Contrast, Blur, Flip, Shift, Scale, Rotate

Backbone recommendations : GoogLeNet, seresnexnet50

Image Size recommendations : 320 * 512

Inferences : 5-fold average and 5 times TTA

- Advice

The amount of data in this competition is very small and the data distribution is imbalanced. A wrong classification of a sample with multiple disease categories will have significant impact on the final results. So, don't trust the public leaderboard, trust your CV.

4. Model Training

Step 1 Preprocessing

- There are some images which is augmented form same image but in different lables
 - : We changed labels of that images as same as first one's label
 - * wrong-labeled sample file names
 - : [Train_1.jpg, Train_171], [Train_379, Train_1173]
- Use only 'healthy' label for the training and test for the first time

Step 2 Set Model

- pick 3 Models from keras.io : InceptionV3, ResNet152V2, Xception

Step 3 Learning

(Trial 1)

```
optimizer=Adam(0.001)
loss = "binary_crossentropy"
metrics = ["accuracy"])
epochs = 40
validation_data = valid_generator_iterator
callbacks = [checkpoint_cb, early_stopping_cb, reduce_lr_cb]
```

results :

```
InceptionV3 - val_loss: 0.0619 - val_accuracy: 0.9808
ResNet152V2 - val_loss: 0.0848 - val_accuracy: 0.9781
Xception - val_loss: 0.0919 - val_accuracy: 0.9781
```

As we all know this competition must have use "All Categories" on preproseccing and evaluate results as AUC so we can changed our stretagy using all features. For this, Reactivate all features and make Dense to 4 and use MSE for loss.

(Trial 2)

```
optimizer = SGD  
loss = "mse"  
metrics = [tf.keras.metrics.AUC()])  
epochs = 40  
callbacks = [checkpoint_cb, early_stopping_cb, reduce_lr_cb]
```

results :

```
InceptionV3 - val_loss: 0.0227 - val_auc: 0.9889  
ResNet152V2 - val_loss: 0.0231 - val_auc: 0.9874  
Xception - loss: 0.1398 - auc_2: 0.8534
```

(Trial 3)

Same conditions as Trial 2

results :

```
InceptionV3  
    epochs 64 - val_loss: 0.0295 - val_auc_2: 0.9899  
  
ResNet152V2  
    epochs 64 - val_loss: 0.0293 - val_auc_2: 0.9883  
  
Xception  
    epochs 64 - val_loss: 0.0161 - val_auc: 0.9903
```

Exceptionally high scores, but when is submitted to Kaggle competition scores were dropped as 0.4xxs. So we changed test_generator's parameter shuffle from True to False. And add some param

(Trial 4)

```
ResNet152V2, "loss = categorical_crossentropy", "Adam=0.001
```

```
val_loss : 0.1759, val_auc : 0.9911
```

```
(Changed : in test_generator_iterator, "shuffle=False")
```

All trials are tested in condition of image_siz : 299, 299, 3

5. Evaluation (by Kaggle)

ResNet152V2 (64 epochs, Adam 0.001, Categorical_CrossEntropy)

- SCORE : 0.96501
- RANK : 393th

6. Conclusion

“More preprocessing : Better Score”

Augmentations are very important in this project because so many pairs of train images are mostly augmented from “Same Image”. And visual features of plant pathology can be easily detected when its brightness and contrast are more versatile(variable).

Further more, there are some pairs of images which are augmented from same image but labeled differently, so we have to adjust this labels as same.

After project, we tried additional version ; image_size 320*512
(Other conditions and parameters are same as Trial ver.4)

Opinions & Comments

If we can find out specific numerical values or figures of brightness and contrast which makes features of plant pathology more “POP”, we can improve this model better. But there must be more studies for that.

But it was meaningful time that we can realize the importance of “Data Preprocessing” in Deep Learning.

Cheers.