Edison Wang

Exploration and Deployment of Migrating Ensemble's Model for Fruit Classification Performance

Oxford Brookes University in collaboration with Chengdu University of Technology

Supervised by Dr Happy Nkanta Monday

Abstract

There are many types of fruits, and human identification may have more errors, in addition, human identification in the orchard also brings high labor cost. The current deep learning techniques used in this field mainly rely on a single CNN model, and if the classification problem becomes more complex, the performance of a single CNN model will be degraded. In order to solve these problems, this project will use an ensemble model to provide a more accurate and powerful network for fruit recognition. The model uses a weighted ensemble algorithm to merge three pre-trained CNN models obtained through transfer learning: AlexNet, ResNet-50, and EfficientNet_b7. The process also involves the proper introduction of the attention mechanism, which proves to have a positive effect on the model classification.

Dataset

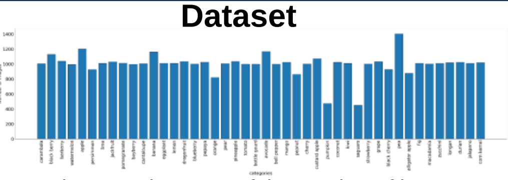


Figure 1 Histogram of the number of images in each category

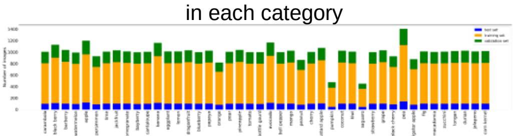


Figure 2 Histogram of proportions of training set and test set for each category

The training set, validation set and test set are divided in the ratio of 7:2:1.

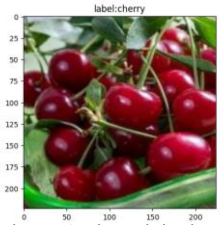


Figure 3 The original one

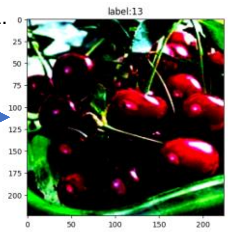


Figure 4 After the data augmentation

In image preprocessing, the training set is first cropped with random scaling and then image enhancement is performed. Since the input to a neural network is usually a tensor, the image is converted to a pytorch tensor and finally normalized.

Ensemble models

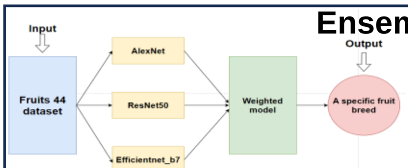
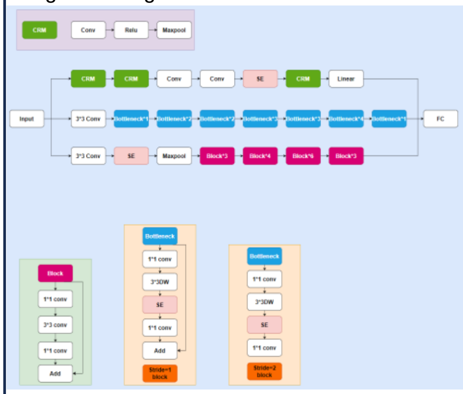


Figure 5 weighted ensemble model

The weights of AlexNet ,ResNet50 ,EfficientNet_b7 models are set to 0.1,0.2,0.7 respectively. The final outputs of the three networks are summed according to the corresponding weights of the three networks.



Techniques such as batch normalisation and average pooling are employed for each of the three independent models to speed up training and improve the generalisation ability of the models. In addition, the SE attention mechanism is introduced to enhance the model's attention to the important features in the fruit images to further improve the model performance. The SE module is mainly used to enhance the channel attention by reducing the spatial dimensions to 1x1 through global average pooling and generating weights to compress and excite the input features. The SE module is placed at different stages of the three network architectures in order to enhance the ability to enhance the model capability of the important feature learning.

Figure 6 Ensemble Modelling with Attention Mechanisms.

Model evaluation

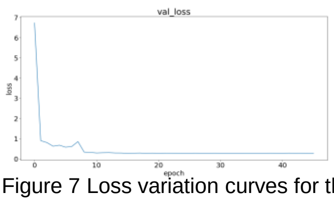


Figure 7 Loss variation curves for the Ensemble model on the validation set

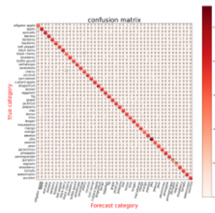


Figure 9 The best model's confusion matrix

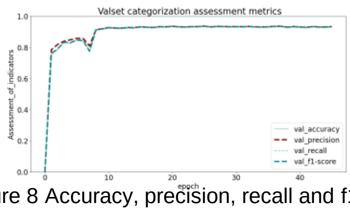


Figure 8 Accuracy, precision, recall and f1-score variation curves of Ensemble model

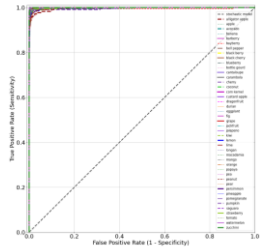


Figure 10 The best model's ROC graph

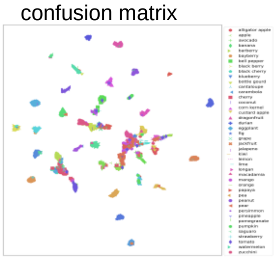


Figure 12 UMAP downscaling to 2D visualization

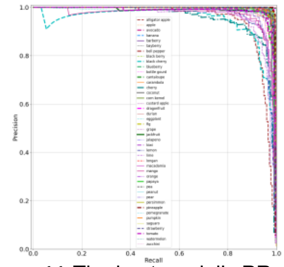


Figure 11 The best model's PR graph

The figure above shows the feedback obtained from the weighted integration of the three models using transfer learning, which was used to classify the fruit dataset for this experiment.

Interpretability and significance analysis of the model

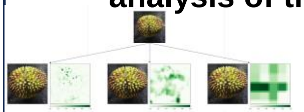


Figure 13 *Occlusion Interpretability Analysis*

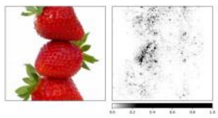


Figure 14 *Integrated Gradients Interpretable Analyses*

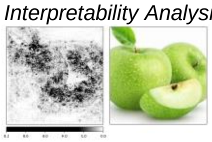


Figure 15 *GradientShap Interpretability Analysis*

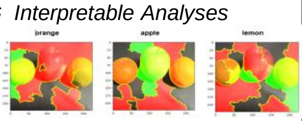


Figure 16 *LIME Interpretability Analysis*

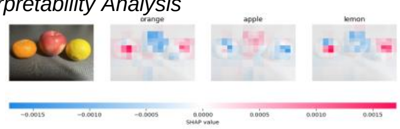


Figure 17 *The shap interpretability analysis algorithm*

Decomposability analysis aims to explain the model's decision-making process and help understand why the model makes particular predictions. Saliency analysis, on the other hand, focuses on identifying the regions or features in the image that have the most influence on the model's predictions, helping to understand the model's focus and decision-making rationale.

WEB APPLICATION DEVELOPMENT

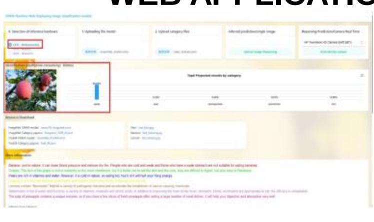


Figure18 Web interface to the model

Figure19 real-time prediction in web front end

The deployment runs in the browser, but only provides a web front-end interface and a JavaScript runtime environment, while the inference prediction still runs in the local terminal

Future work

1. More advanced integration methods and attention mechanisms will be explored in the future to further improve model performance in complex image classification tasks.
2. For post-deployment model optimization, strategies such as model lightweighting, incremental learning and model monitoring can be used.