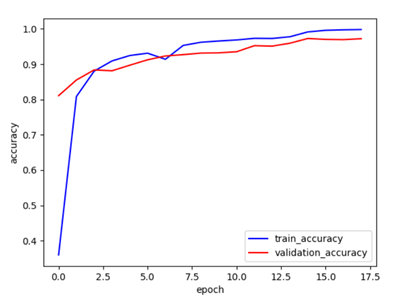
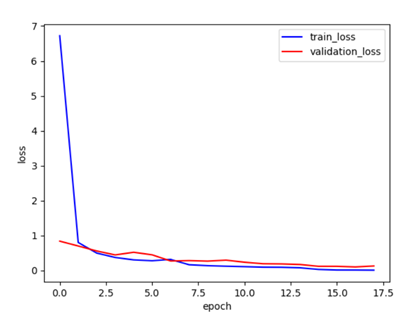
1.Our work

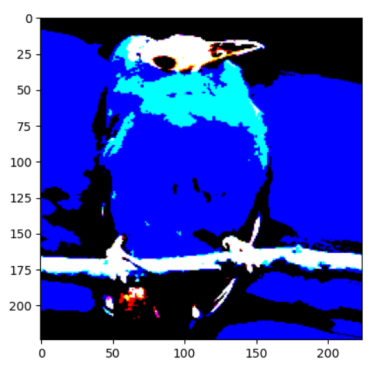
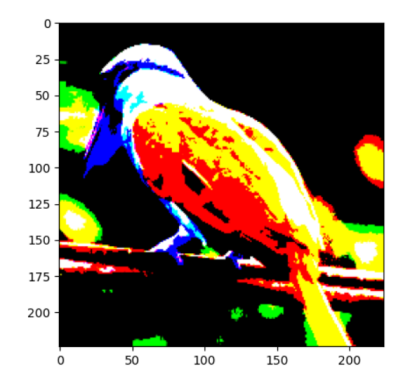
A standard VGG16 model is applied to train the model, with referring to a previous kaggle submission at [https://www.kaggle.com/blakewinters/vgg16-bird-species]. The detailed structure and implementation of VGG16 can be found in the paper published by Karen and Andrew at [https://arxiv.org/abs/1409.1556].

The original framework repeatedly fed the data three times to train the model to adjust the learning rate, our work involves improving the code to make the structure neater, and also fine-tuned the model to reduce the training time while preserving the accuracy. The accuracy on the test set is 97.21%, training/validation loss and accuracy are listed below:

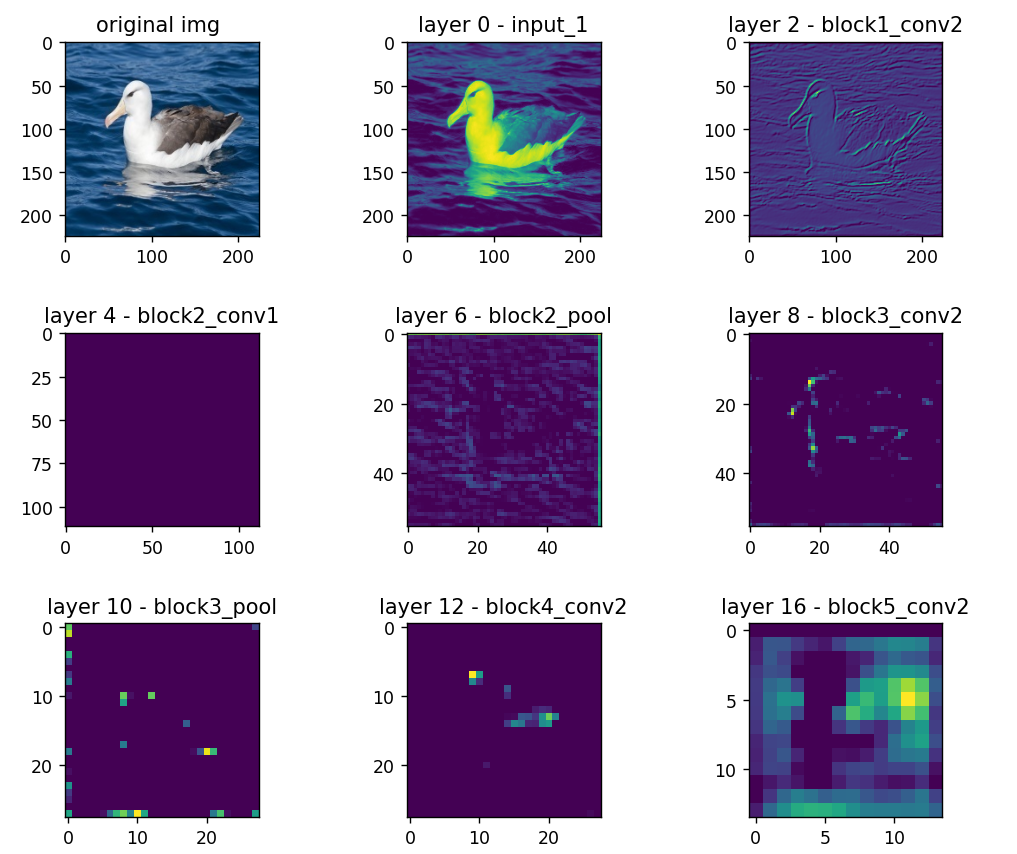


1. Visualize the model

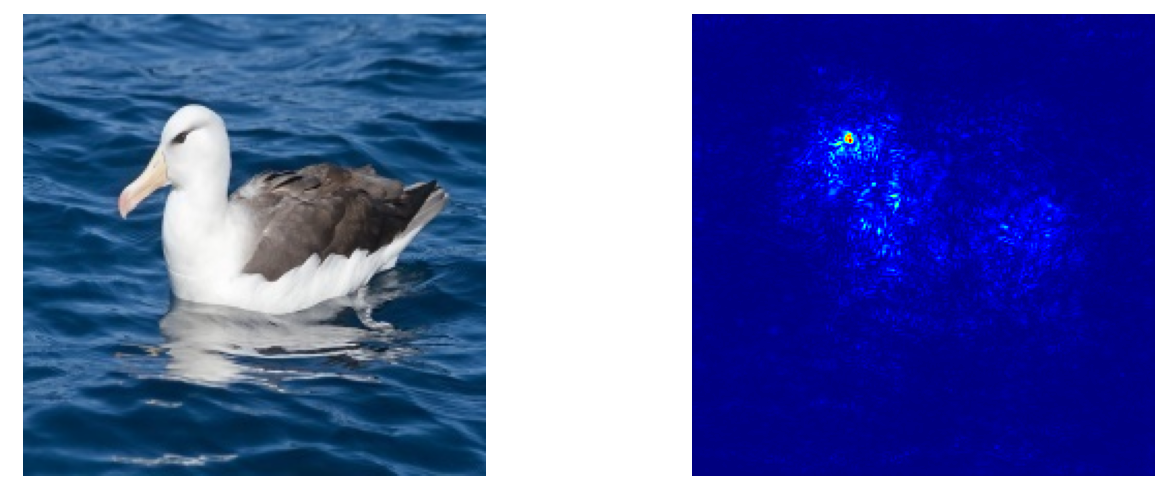
The image preprocessing step used function in keras library, yielding following effects(the original pictures are from the training data available at kaggle[https://www.kaggle.com/gpiosenka/100-bird-species]):

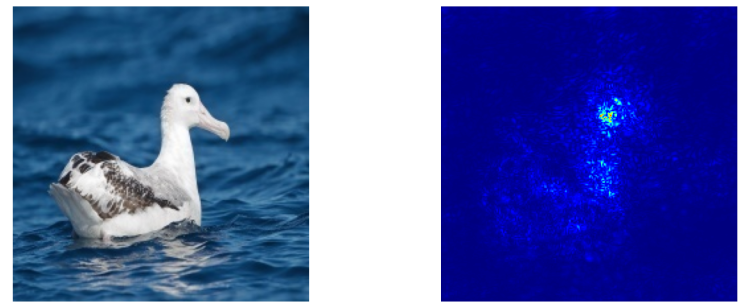


We checked the intermediate results in the model and tried to achieve a better understanding of the model. Since VGG has a deep structure of convolution layers(22 layers for vgg16), some early layers can reveal the effect of convolution on input image(e.g. layer2), but most layers can’t give us intuitive lower-level informations.



The saliency map provided a more intuitive visualization showing which features/parts of the image contributed more to the prediction, for example(picture from kaggle dataset, albatross 002/003):





The saliency map shows the eye regions significantly contributed to the final prediction result, also the highlighted region of the heatmap matched the bird region on left hand side and excluded the background. This helped to show the robustness of the model.

We also artificially transformed the original picture to see whether prediction result would be affected, the results are shown below:

|  |  |  |
| --- | --- | --- |
| figure | transform | Prediction result(Probability %) |
| original | Original image | 99.9999999987 |
| affine | Affine transformation | 99.99999994 |
| break_main_feature | Gaussian noise at eye region | 99.999984 |
| grayscale | Color transformation(Grayscale) | 99.999973 |
| break_side_feature | Gaussian noise at background | 99.9975 |
| distort | Distortion | 98.81 |
| gaussian_noise | Global gaussian noise | 92.1 |

None of those transformation would change the final prediction label, but they more or less brought some influence on the prediction probability as shown in the table. The result also proved the robustness of the model.