DSCI 552 Machine Learning for Data Science
Professor Ke-Thia Yao
Group1: Rui Xue, Zijun Tan, Siqi Qiao, Siyi Chen
Final Project Report: Image CNN Classification: Two-Level Hierarchical Classification for Category and Landmark

Abstract
The project aims to utilize a convolutional neural network (CNN) pre-trained model that can effectively analyze colored images of objects and classify them into one of the landmark and category classes of the given dataset. We incorporated transfer learning and data augmentation in our experiment to increase the model's proficiency and robustness for the given small dataset. The base model is trained on VGG16 architecture with weights on ImageNet. The model is trained for 20 epochs and batch size 32, resulting in an accuracy of 75%.

Introduction
The deep CNN has shown remarkable classification results in the ImageNet challenge, as proposed by Krizhevsky et al. in 2012. The deep CNN model, consisting of convolutional and fully connected layers, is known for its strong feature learning ability and better classification performance. In this project, our goal is to classify images with their corresponding category and landmark class using pre-trained CNN models. The dataset consists of 6 categories and 30 landmarks, with 420 images. This task is particularly challenging due to the two-level hierarchical classification on a small dataset. In normal practice, deep learning models tend to achieve better performance when they are provided with a larger amount of input data. However, our given data is limited. It is necessary to develop methods for constructing deep learning models using limited amounts of supervised data.

Methods
In order to overcome the problem of a small dataset, transfer learning allows us to leverage the knowledge learned by the pre-trained network and apply it to a new task, resulting in better performance with less training data and time. There are studies on two-hierarchical classification using transfer learning under AlexNet, CaffeNet, and VGG16 net-based architectures. The research proposed by Song et al. (2016) adopted a deep CNN that is pre-trained on ImageNet as the base network and then fine-tuned the base network to the general dataset through transfer learning. The results showed that VGG16Net (fc6, k=4) achieves the highest accuracy of 82.45 among all models on the Caltech-256 dataset (Song, et al. 2016).

We flatten the input data label into a single label (eg. Modern/Eiffel to Modern_Eiffel) to train images on landmark classification and then infer the category class. Since landmark information can provide model knowledge about its corresponding category class. We tried three different pre-trained convolutional neural network architectures: EfficientNet, VGG16, and Inception3 on our training dataset. After comparing its accuracy, we employed a single model to predict both category and landmark class, with VGG16 as our base model for higher accuracy. Below is our model flow chart.

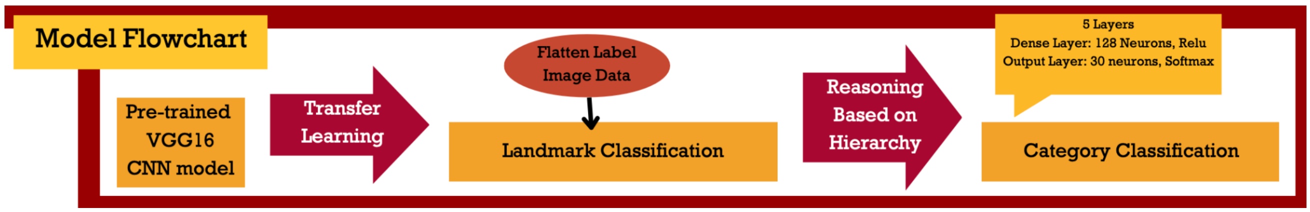

Fig1: Proposed model flow chart

There are five layers in our proposed CNN model under VGG16 architecture. The GlobalAveragePooling2D layer reduces the number of parameters in the model by taking the average of the feature maps. The Dense layer has 128 neurons with the activation function' relu', and the output layer has 30 neurons with the activation function' softmax'. Next, the layers in the pre-trained model are frozen using a for loop to prevent their weights from being updated during training. Finally, the model is compiled using the Adam optimizer with a learning rate of 0.01.

To achieve better performance on our train model, one approach is to increase the number of available samples through data augmentation. This method has been widely used in research papers in different fields, such as medical image classification (An et al., 2021). An's work proposed utilizes a hierarchical classification method to effectively classify diseases and their sub-types. The method involves using healthy/disease information from the first model to build subsequent models for subtype classification through data augmentation (An et al., 2021). Our results also show that the model achieves higher accuracy with data augmentation. After manipulating the origidal dataset using ImageDataGenerator, we ended up with 8,230 dataset to pass into our model. The accuracy of VGG16 model using data augmentation has increased.

Results

The accuracy of the validation dataset using EfficientNetB0 is extremely low. The validation accuracy is 0.0147. The loss and accuracy graph is shown below, with an F-1 score of 0.001 on landmark classification and 0.03 on category classification. We decided to move on with this approach due to the bad performance.

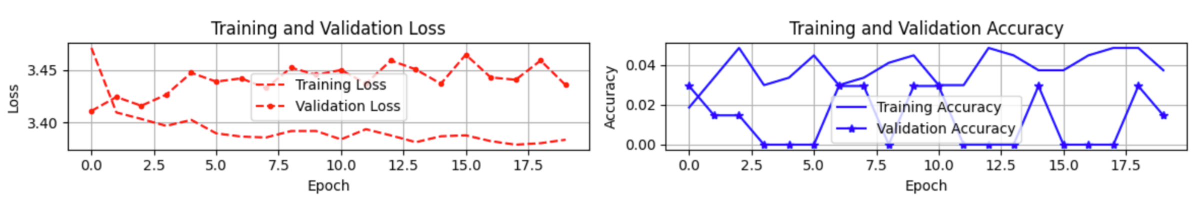

Fig2&3: loss and accuracy graph of EfficientNetB0 model

The accuracy of the InceptionV3 model after data augmentation is 0.6765, with an F-1 score of 0.6610 on landmark classification and 0.9012 on category classification.

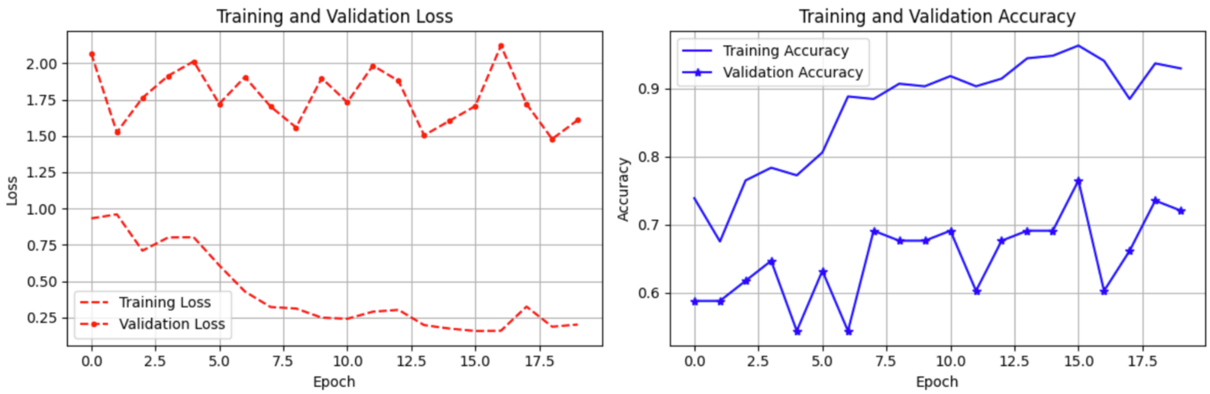

Fig4&5: loss and accuracy graph of InceptionV3 model after data augmentation

The accuracy of the validation dataset using VGG16 is 0.70. However the accuracy on training dataset using VGG model is approximately 1 which indicates that this model may overfit the training data.

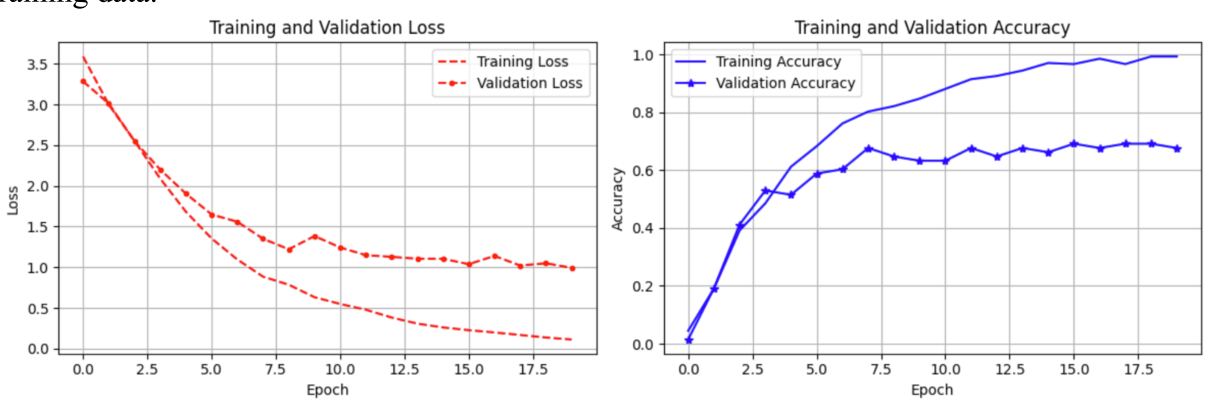

Fig6&7: loss and accuracy graph of VGG16 model

Then we tried with data augmentation, the accuracy achieved 0.75. With an F-1 score of 0.6670 on landmark classification and 0.9153 on category classificationIt is the highest accuracy overall.

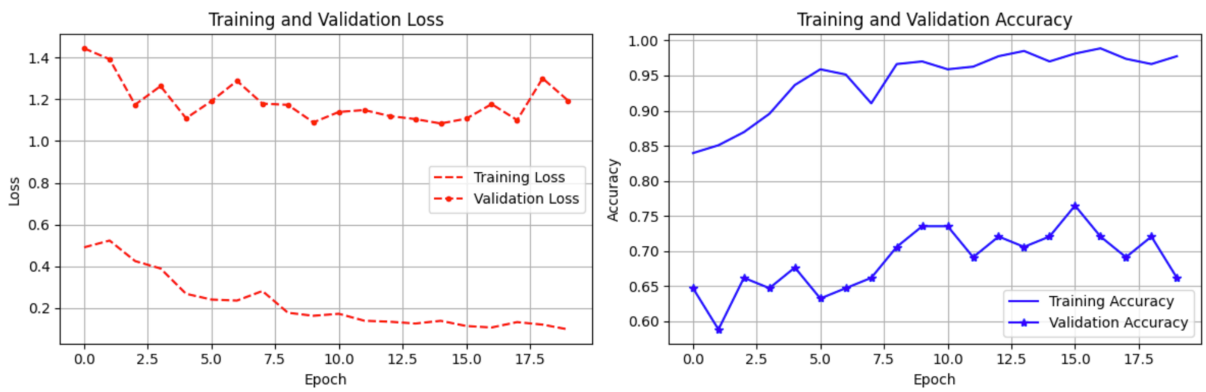

Fig8&9: loss and accuracy graph of VGG16 model after data augmentation

For better visualization, we create a side by side bar chart to evaluate the accuracy of three models in predicting landmarks and category.

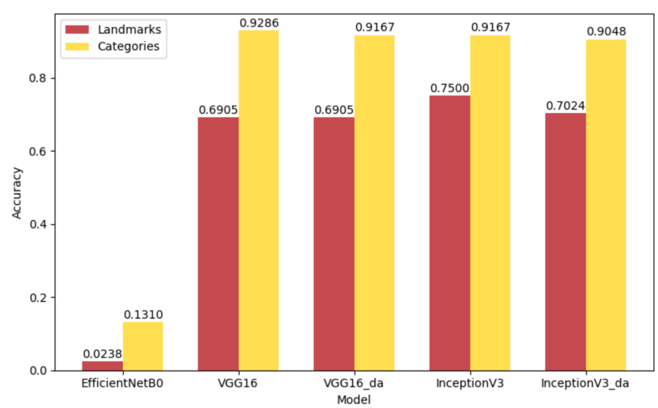

Fig10: Model comparison on landmark and category accuracy

Conclusions/Discussions

In this project, we proposed to use pre-trained model VGG16 to classify images into its landmark class and category class. The experimental results indicates that the choice of pre-trained model and utilization of data augmentation can improve the accuracy of the classification task. The category and landmark are learned together in our model, the higher category classification results showing that it benefits from knowing the output of landmark classification. In the future, we plan to test the versatility of the proposed approach by perform experiments on a larger dataset. We will keep exploring the different parameter under VGG16 framework, such as the number of layers to further improve the classification accuracy.

Bibliography

An G, Akiba M, Omodaka K, Nakazawa T, Yokota H. Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. Sci Rep. 2021 Mar 1;11(1):4250. doi: 10.1038/s41598-021-83503-7. PMID: 33649375; PMCID: PMC7921640.

Donahue, J., Jia, Y., Vinyals, O., et al., 2014. DeCAF: a deep convolutional activation feature for generic visual recognition. Proc. 31st Int. Conf. on Machine Learning, p.1-9.

Alex Krizhevsky et al., “ImageNet Classification with Deep Convolutional Neural Networks”, Proceedings of the 25th International Conference on Neural Information Processing Systems 1 (2012): 1097–1105.

Guang-hui SONG, Xiao-gang JIN, Gen-lang CHEN, Yan NIE. Two-level hierarchical feature learning for image classification. Front. Inform. Technol. Electron. Eng, 2016, 17(9): 897–906 https://doi.org/10.1631/FITEE.1500346