DSCI 552 Final Project Report

By Lian Lian
Aining Ding
Weixin Sheng
Yuqi Zhang

April 30, 2023

Abstract:

The main task of our project is to predict the category names of images and the landmark names of images by using transfer learning through applying a pretrained model. Transfer learning involves using pre-trained neural network weights for the lower layers, instead of starting the training process with randomly initialized weights. This means that only the upper layers of the network need to be trained. In the project, we used the EfficientNetB0 model to freeze the lower layers and only fine-tune the newly added layers on top of the pre-trained model. This approach allows us to train the model with fewer data points and computational resources. We split the dataset with 80% training size and 20% test size. To address the issue of limited datasets, we utilized data augmentation techniques such as rotation, translation, flipping, and contrast adjustment during the preprocessing stage. By decreasing the learning rate, we were able to achieve greater accuracy. Ultimately, we attained an accuracy of approximately 96% for the category and roughly 91.6% for the landmarks. Overall, our category and landmark classifications demonstrate a high level of accuracy. As a result, our image classification task is able to effectively recognize the majority of image categories and landmarks.

Introduction:

The project categorizes images of landmarks into different categories and identifies the specific landmark within each category. The dataset consists of images of five famous landmarks organized into six categories, including Gothic, Modern, Mughal, Neoclassical, Pagodas, and Pyramids. This dataset is relatively small to develop comprehensive CNNs. To address this issue, we used transfer learning with a pre-trained model and applied several data augmentation techniques to generate additional data for training and to help improve model performance. Two branches are needed to accomplish both learning tasks, so we invented separate layers for category prediction and landmark prediction. Additionally, the pre-trained model can easily overfit the data, so we decided to freeze the lower layers to tune the data on the top layers. Another difficulty we met is to decide which pre-trained model to use between EfficientNetB0 and VGG16. Each model has its strengths and weaknesses, we compared both models and decided to use EfficientNetB0 to improve model performance.

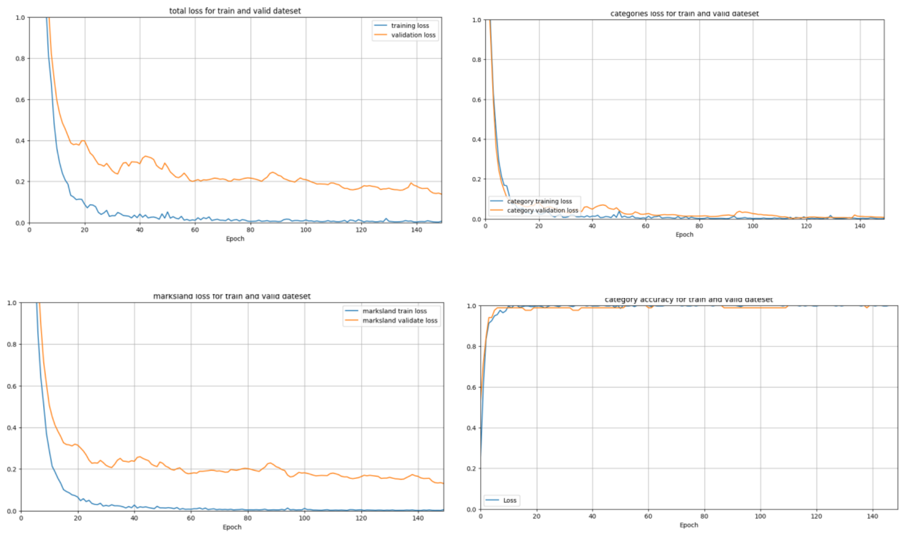
Methods:

We put the data into google drive and define a function to load image data from a specified path in Google Drive. It first creates a list of all the categories in the path and then loops through each category to create a list of all the landmarks within that category. For each landmark, it loops through all the images within that landmark and tries to load each image using Keras' load_img() function. If an error occurs, it prints an error message and continues with the next image. Three images cannot be loaded into the dataset, because they are not defined as images.

We split the dataset into 20% for validation and 80% for training for both categorized and landmark predictions. Defined a function by using image augmentation techniques that can be applied to images to generate new images with slightly modified versions of the original images. These techniques include RandomRotation, RandomTranslation, RandomFlip, RandomContrast. The purpose of applying these augmentation techniques to the images is to generate additional training data and to prevent overfitting by introducing variability into the dataset. We decided to use EfficientNetB0 as the pre-trained model since it's a relatively small model with fewer parameters compared to VGG16, the parameters can be easily adjusted to this project to enhance the performance. We freezed 217 layers when using EfficentNetB0. By using data augmentation and pre-trained weights, the goal of these operations is to improve the performance of the model by allowing it to learn from a wider variety of images while also benefiting from the feature extraction capabilities of the EfficientNetB0 model. "GlobalAveragePooling2D" applies average pooling on the spatial dimensions until each spatial dimension is one, and leaves other dimensions unchanged. After building the model, we create a label encode and normalize the label with "LabelEncoder", then fit the encoder to the target categories and transform the non-numerical labels to numerical labels. Finally we convert the numerical labels to one-hot encoded vectors. Then, we use train and validation datasets to train the model.

Results:

We evaluated the performance of the models using various metrics such as accuracy, precision, and recall. We found that data augmentation helped improve the performance of the models. The model achieved an accuracy of approximately 96% for category classification and 91.6% for landmark classification. The F1 score of category prediction is 0.9904 and the F1 score of landmark prediction is 0.9731. In the autograder test, the F1 score of category prediction is 0.9503 and the F1 score of landmark prediction is 0.7595.



Conclusions/Discussions:

In conclusion, our project successfully demonstrated the feasibility of using transfer learning with EfficientNetB0 and data augmentation techniques to classify landmark images into their respective categories and identify the specific landmark depicted.

Our project demonstrates that transfer learning can be a powerful technique for image classification tasks, particularly when working with limited datasets. By utilizing the pre-trained EfficientNetB0 model and freezing 217 layers, we were able to achieve high accuracy in both category and landmark classifications. We also acknowledge that data augmentation also enhances the model’s performance. In earlier steps, we use the data augmentation techniques to generate better data in the training set as well as reducing the chances of overfitting.

However, it is worth noting that there are limitations to our approach. The size of the dataset is limited which will increase the chance of overfitting. So having larger datasets with more landmarks examples and categories could enable our model to classify a wider range of data. Other than that, our future work involves fine-tuning the entire model to improve the accuracy of classification. We will also use more advanced regularization techniques to enhance the performance of our model.

Overall, our project demonstrates the effectiveness of transfer learning and data augmentation techniques for image classification tasks, and we believe that these approaches will continue to be important tools for researchers and practitioners in the field of machine learning.

Bibliography:

Chollet, F., & others. (2015). Keras. .

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X.. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B, Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Antoniadis, Panagiotis. “Machine Learning: What Is Ablation Study?” *Baeldung on Computer Science*, 4 Nov. 2022, https://www.baeldung.com/cs/ml-ablation-study.