Category Classification and Landmark Classification Based on EfficientNetB0 Network

Kayvan Shah, Zuqi Li, and Qianyou Wang
University of Southern California, Group 23 for DSCI-552

USC Viterbi School of Engineering

Introduction

- Image classification has always been an important topic in machine learning that involves assigning a label to an image based on its content.
- In this project, we aim to address the difficulty of landmark image classification, where the task is to predict both the category names and landmark names of given images.
- The dataset provided is organized into a two-level hierarchy structure. It is categorized into six categories: Gothic, Modern, Mughal, Neoclassical, Pagodas, and Pyramids, and for each category there are 5 landmarks for a total of 30 landmarks, with each landmark having 14 images.
- Based on existing research, we discussed and stated these issues.
  - Implement data augmentation to overcome the problem of small datasets.
  - For classification problems, a single multi-classifier is proposed.
  - The landmark classification benefits from knowing the output of category classification and multi-task learning is be considered to incorporate two tasks into one model for training and optimization.
  - Choose EfficientNetB0 as our pre-trained model for transfer learning..
- We implement EfficientNetB0 to deal with the category classification and landmark classification simultaneously and obtain a result of f1-score 0.9848 for category validation, and a result of f1-score 0.8030 for landmark validation.

Methods

Data Preparation

1. Loading data from given folders and generating metadata
2. Resizing the images into 224x224
3. Encoding categorical labels
4. Splitting into train, validation and test sets
5. Converting into TensorFlow dataset for better recalling.

Model Structure

- Use the preprocessing dataset as **input**.
- Since the given dataset is small, there possibly exists overfitting in our model. We implement **image augmentation** to improve the robustness and generalization of our models.
- They are connected to the pre-trained **EfficientNetB0** layer with only the feature vectors and frozen parameters.
- A flattened **fine-tuning** layer is connected to the pre-trained model layer in order to make it adapted to a new task or dataset by adjusting the parameters of some of its layers.
- After that, we have **two branches**, one for category classification and another for landmark classification. The final layers of these two branches have the same number as which of categories and landmarks.

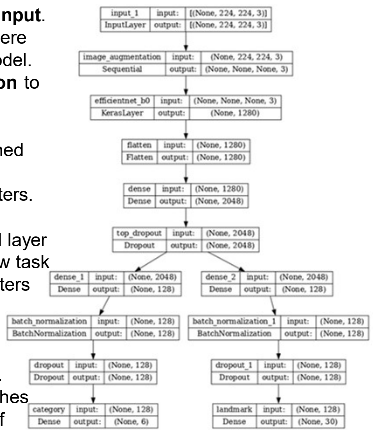
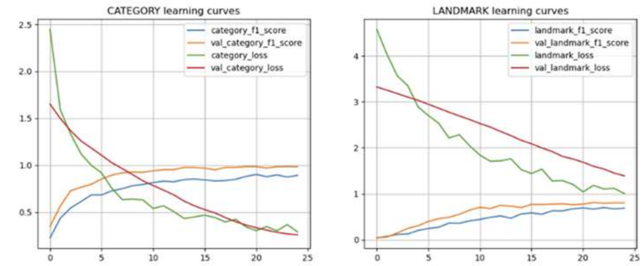


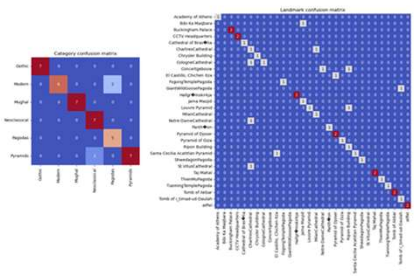
Fig1. Model Structure

Results

- Landmark dataset: 420 images in total, belonging to 30 landmarks and 6 categories
- **Image Augmentation** implemented:
  - rotates the image by an angle within 20% of the maximum rotation angle
  - adjusts the brightness up to 45% lighter or darker
  - flips the image horizontally with a probability of 50%
  - zooms up to 20% in both the x and y directions
  - adjusts the contrast of the image by up to 30% more or less contrast
  - resizes the image by up to 20% along the vertical and the horizontal axis.
- **Batch size**: 32 **Epoch**: 25 **Learning rate**:1e-4
- We obtain a result of f1-score 0.9848 for category validation, and a result of f1-score 0.8030 for landmark validation.
- After training, we obtain **learning curves** for loss function and f1-score.



- We also obtain the **confusion matrix** for each label.



Classification report for Category

	precision	recall	f1-score	support
0	1.00	1.00	1.00	7
1	1.00	0.75	0.86	8
2	1.00	1.00	1.00	7
3	0.88	1.00	0.93	7
4	0.71	1.00	0.83	5
5	1.00	0.88	0.93	8
accuracy			0.93	42
macro avg	0.93	0.94	0.93	42
weighted avg	0.95	0.93	0.93	42

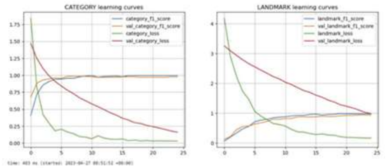
Classification report for Landmarks

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1
1	0.00	0.00	0.00	1
2	1.00	1.00	1.00	2
3	1.00	1.00	1.00	2
4	1.00	1.00	1.00	1
5	0.25	0.50	0.33	2
6	1.00	1.00	1.00	1
7	1.00	0.50	0.67	2
8	0.00	0.00	0.00	2
9	0.00	0.00	0.00	1
10	1.00	1.00	1.00	1
11	0.00	0.00	0.00	1
12	1.00	1.00	1.00	2
13	0.50	1.00	0.67	1
14	1.00	0.50	0.67	2
15	0.50	1.00	0.67	1
16	0.00	0.00	0.00	1
17	1.00	1.00	1.00	1
18	0.67	1.00	0.80	2
19	1.00	1.00	1.00	1
20	0.33	1.00	0.50	1
21	1.00	0.50	0.67	2
22	1.00	1.00	1.00	1
23	0.00	0.00	0.00	1
24	1.00	1.00	1.00	2
25	1.00	1.00	1.00	1
26	1.00	1.00	1.00	1
27	1.00	1.00	1.00	2
28	0.50	1.00	0.67	1
29	1.00	1.00	1.00	2
accuracy			0.74	42
macro avg	0.69	0.73	0.69	42
weighted avg	0.73	0.74	0.71	42

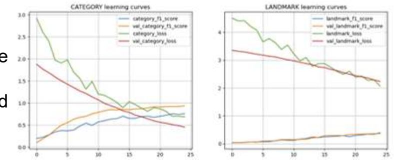
- we also have our classification reports that contain precision, recall, f1-score, and support for categories and landmarks. As shown on the right, we have reached an average accuracy of 0.93 for category classification, and an average accuracy of 0.74 for landmark classification.

- Ablation studies: Research on a model without image augmentation or without fine-tuning part.

From the learning curves without image augmentation, it's obvious that classification using EfficientNetB0 model on small datasets may cause great overfitting problems.



From the learning curves without fine-tuning part for pre-trained model, we can see that overfitting problems exist for landmark classification and the prediction accuracy is rather low.



Model	f1-score for Category	f1-score for Landmark
Optimized	0.93	0.74
Without image augmentation	0.98	0.88
Without fine-tuning layer	0.88	0.45

Conclusion

- We explored the transfer learning approach to deal with the category classification and landmark classification for the given small dataset.
- We implemented the EfficientNetB0 pre-trained model and tuned hyperparameters as well as added some layers to the top of the model to optimize the models' performance.
- Transfer learning with pre-trained models can achieve high accuracy but smaller datasets are more prone to overfitting.
- Based on ablation studies by deleting the image augmentation part and fine-tuning part separately, we also show that appropriate image augmentation and fine-tuning for the pre-trained model help to improve the robustness and generalization of our model and avoid overfitting problems.
- Our approach is beneficial for practical applications such as image classification with small learning sets and limited resource usage.
- Our results provide insights into the models' behavior and how we can improve the performance when implementing such models, which can inform future research directions.

References

[1] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946.
[2] Saba, T., Khan, A. M., Khurshid, K., & Lee, J. Y. (2020). Multi-class classification of diabetic retinopathy images using VGG16 and EfficientNet. Applied Sciences, 10(4), 1324.
[3] Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621
[4] Chollet, F. (2019). Introduction to Image Classification with Keras. TensorFlow. https://www.tensorflow.org/tutorials/keras/classification