Landmark Detection

Jayanth Dasari
2019BCS-016
ABV-IIITM, Gwalior

Mahesh Bommisetty
2019BCS-013
ABV-IIITM Gwalior

Satya Pavan Kalyan Vemula
2019BCS-069
ABV-IIITM Gwalior

Anunay Nalam
2019BCS-034
ABV-IIITM Gwalior

I. Problem Statement

The Landmark Detection project aims for Instance Level Recognition, to detect the Landmarks inside given images with different perspectives, from various camera angles and lighting conditions. This technology can predict landmark labels directly from image pixels to help people better understand and organize their photo collections. In this project we are focusing on implementing landmark detection for our own college premises with top performing models.

II. Current Research Area

In this domain, the most popular dataset, the Google Landmarks Datasetv2, consists of over 5M images and over 200k distinct instance labels, making it the largest instance recognition dataset to date[4]. Apart from GLDv2, the Oxford and the Paris datasets[3] are other famous datasets. In image detection/recognition problems, there are mainly three methods, Basic Recognition, Fine-Grained Recognition involving distinction of species/models/styles, and Instance Level Recognition. Currently, the main focus is on Instance Level Recognition and Google Landmark Recognition Dataset due to the vast number of included images and classes.

III. Motivation and Justification

Landmark Recognition has many applications like Visual Search for recognizing famous buildings, Personal Photo Search for organizing one's image library, grouping photos taken at a landmark, and gaming through augmented reality. This project has vast datasets, including many images that are distractors, not landmarks, and have much larger number of classes. Solving these problems is a huge challenge that motivated us to take up this project as our course project. We are implementing this project for our own college Dataset.

IV. Literature Survey

Cheng Xu et al.[5] developed their model on the Google Landmark detection dataset using features and classification logits further optimized with an ArcFace Loss. Included an efficient pipeline for re-ranking predictions by adjusting retrieval scores and got a 0.489 score with the ensembled model.

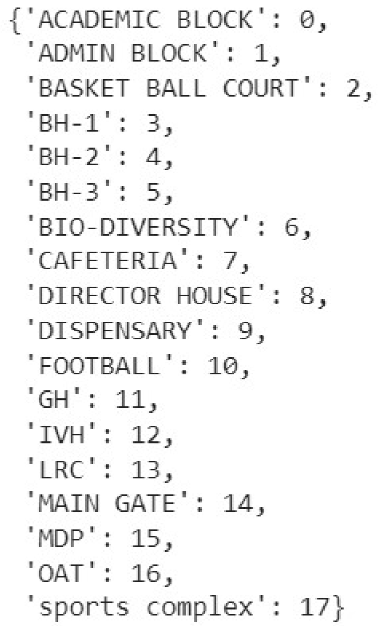
Nilwong et al.[2] presented a method alterantive to conventional ways for outdoor localization relying on Faster RCNN and the feedforward neural network (FFNN) trained on their custom dataset images with geotags and labeled bounding boxes of the koganei campus of Hosei University.

Christof Henkel et al.[1] developed an end-to-end instance level recognition method for labeling and ranking landmark images. In their method, they embed images in a high-dimensional feature space, classify them based on visual similarity, re-ranking predictions, and filter noise based on their similarity to out-of-domain images.

V. Methodology

A. Datasets

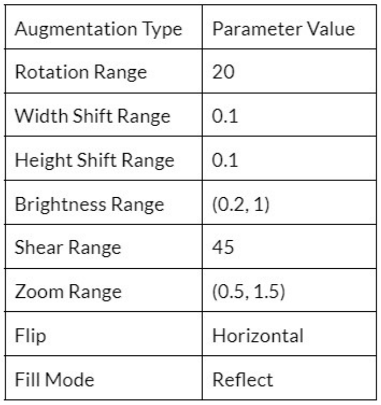
1) Data Collection: We have carefully considered 18 landmarks from all the available college premises. We collected around 35 different images for each landmark under various conditions, like sunset, sunrise, morning, and night with and without proper lights, from multiple angles. We also tried to diversify the dataset as much as possible.



```
{'ACADEMIC BLOCK': 0,
 'ADMIN BLOCK': 1,
 'BASKET BALL COURT': 2,
 'BH-1': 3,
 'BH-2': 4,
 'BH-3': 5,
 'BIO-DIVERSITY': 6,
 'CAFETERIA': 7,
 'DIRECTOR HOUSE': 8,
 'DISPENSARY': 9,
 'FOOTBALL': 10,
 'GH': 11,
 'IVH': 12,
 'LRC': 13,
 'MAIN GATE': 14,
 'MDP': 15,
 'OAT': 16,
 'sports complex': 17}
```

Fig. 1. Landmarks

2) Data Augmentation: Initially, the dataset is around 700 images with 18 classes. To further increase the diversity and amount of data, we performed data augmentation techniques to the existing data. Following table represents the data augmentations we performed and parameters we took.



Augmentation Type	Parameter Value
Rotation Range	20
Width Shift Range	0.1
Height Shift Range	0.1
Brightness Range	(0.2, 1)
Shear Range	45
Zoom Range	(0.5, 1.5)
Flip	Horizontal
Fill Mode	Reflect

Fig. 2. Data Augmentations

B. Models Used

1) EfficientNet: EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. EfficientNet uses a compound coefficient to uniformly scales network width, depth, and resolution in a principled way.

The compound scaling method is justified by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image.

2) ResNet101V2: Resnet uses identity mapping by adding 1layer without having no additional function, which means getting the output exactly the same with the input

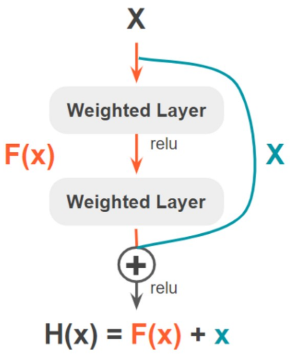


Fig. 3. Resnet

3) MobileNet: MobileNet, a single-shot multi-box detection network open-source by Google, uses the Caffe framework for object detection tasks. As its name suggests, it is designed for mobile applications usage. It uses depthwise separable convolutions, significantly reducing the number of parameters compared to others resulting in a lightweight deep neural network.

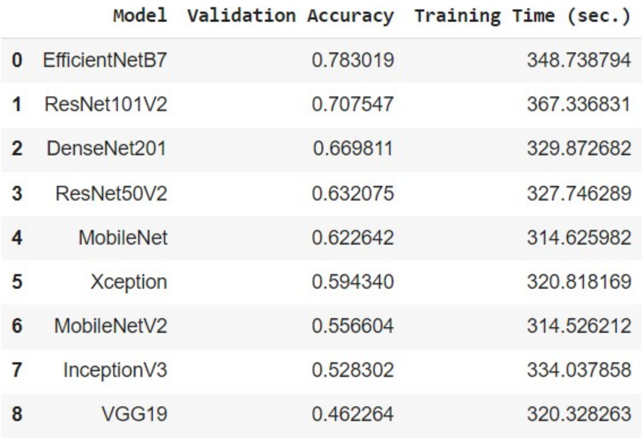
Depthwise separable convolution is a depthwise convolution followed by pointwise convolution. Mobile Nets mainly differ from CNN models in that they split the convolution into a 3x3 depth-wise convolution and a 1x1 pointwise convolution followed by batch normalization and ReLU. The model output is a typical vector of the tracked object data.

VI. Experiments

A. Experiment-1

Here, we explored various existing top SOTA models and trained them for some epochs to compare and find out which SOTA model is best suited for our Dataset.

Performed on Custom College Landmarks Dataset,
Optimization: ADAM,
Activation : ReLu
Regularization: Dropout,
Pooling: Max Pooling



	Model	Validation Accuracy	Training Time (sec.)
0	EfficientNetB7	0.783019	348.738794
1	ResNet101V2	0.707547	367.336831
2	DenseNet201	0.669811	329.872682
3	ResNet50V2	0.632075	327.746289
4	MobileNet	0.622642	314.625982
5	Xception	0.594340	320.818169
6	MobileNetV2	0.556604	314.526212
7	InceptionV3	0.528302	334.037858
8	VGG19	0.462264	320.328263

Fig. 4. Pre-Trained Models accuracy

B. Experiment-2

After Exploring various SOTA models, we further explored the top performing SOTA model EfficientNetB7.
We also explored MobileNet SOTA model due to its simplicity and reliability in real time as it is faster, Also considering the dataset size we have. We got the following results.

EfficientNetB7:
Training Accuracy : 0.9612
Validation Accuracy : 0.7879

MobileNet:
Training Accuracy : 0.9871
Validation Accuracy : 0.7424

VII. WorkDone

- IIITM Campus dataset collection, Collected the IIITM Campus images under different perspectives, lighting conditions, and from various angles.

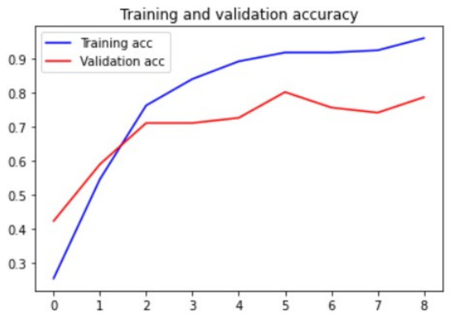


Fig. 5. EfficientNetB7 accuracy

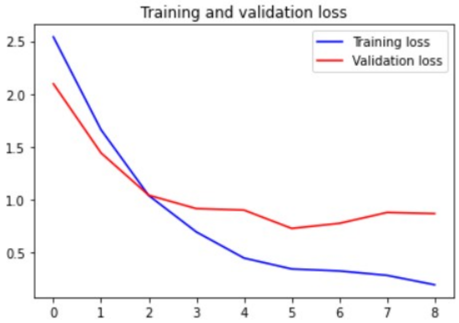


Fig. 6. EfficientNetB7 Loss

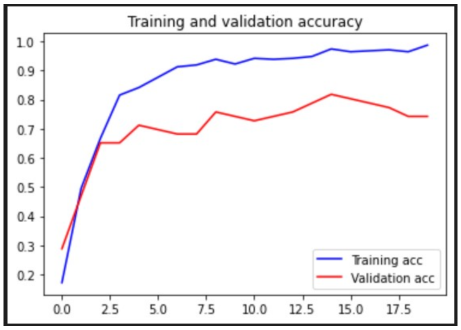


Fig. 7. MobileNet accuracy

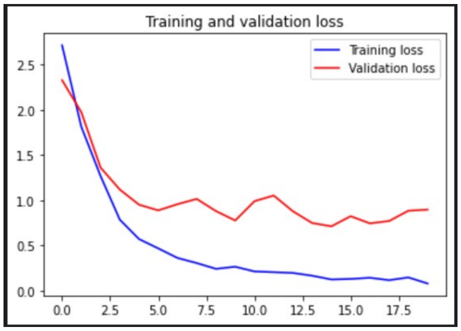


Fig. 8. MobileNet Loss

- Collected the campus images data by considering 18 prominent landmarks. Custom Dataset contains 700 images with intra-class variability, equally distributed.
- Performed Data Augmentation to increase the data size and data diversity.
- Explored various SOTA models(Resnet, EfficientNetB7, VGG, MobileNet and soon). And further explored top performing SOTA model EfficientNetB7 and MobileNet on our dataset with more epochs.

VIII. Future Work

- Expanding Dataset.
- Can integrate low light image enhancement methods for better classification of images taken at night/when no light.
- Exploring various data preprocessing steps for proper feature extraction from the existing data, making it work better even with the low quality images.

References

[1] Christof Henkel and Philipp Singer. *Supporting large-scale image recognition with out-of-domain samples*. 2020. arXiv: 2010.01650 `[cs.CV]`.

[2] Sivapong Nilwong et al. "Outdoor Landmark Detection for Real-World Localization using Faster R-CNN". In: Oct. 2018, pp. 165–169. DOI: 10.1145/3284516.3284532.

[3] Filip Radenović et al. *Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking*. 2018. arXiv: 1803.11285 `[cs.CV]`.

[4] Tobias Weyand et al. "Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[5] Cheng Xu et al. *3rd Place Solution to Google Landmark Recognition Competition 2021*. Oct. 2021.