# Google Landmark Recognition 2021:
# Benchmark for Instance-Level Recognition

Ariuntuya Altanzaya, Chingun Khasar
aa4928@nyu.edu, ck3411@nyu.edu

New York University
Tandon School of Engineering

December 21, 2021

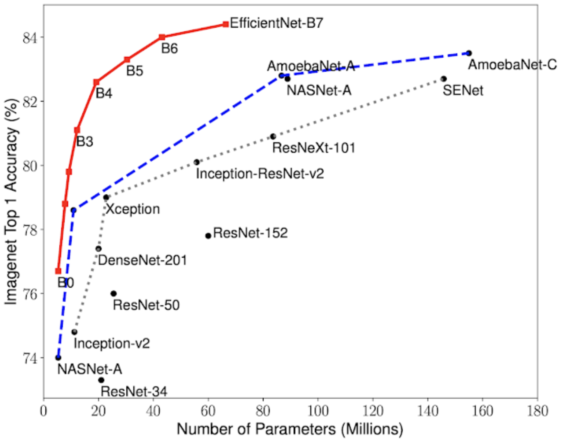
## Abstract

We present an efficient model for large-scale landmark recognition and retrieval. Through our research, we show how to combine and experiment with concepts from recent research and Kaggle competition winners and introduce our own architecture suited for large-scale landmark identification. As a start, we implemented and trained ResNet-50. We observed that EfficientNet models achieve higher accuracy and better efficiency over existing CNNs while reducing parameter size. Compared with ResNet-50, EfficientNet-B4 improved accuracy from 76.3 percent of ResNet-50 to 82.6 percent (+6.3). Embeddings of images are extracted via EfficientNet architecture which are optimized by RAdam algorithm. However, due to computation constraints considering the size of our dataset (+100 GB), we chose the simplest model EfficientNet-B0. With these settings, our solution scores a maximum 0.02 GAP score which scores a Bronze Award in the Kaggle competition.

## 1. Introduction

Google Landmark Recognition 2021 Competition is the fourth landmark recognition competition on Kaggle. Competition entries are evaluated using Global Average Precision (GAP). The goal of the challenge is to recognize a landmark presented in a query image. This year, specifically, the host introduced more diversity in the test images to measure global landmarks more fairly. To facilitate recognition by retrieval approaches, the private training set contains only a 100k subset of the total public training set. This 100k subset contains all of the training set images associated with the landmarks in the private test set.

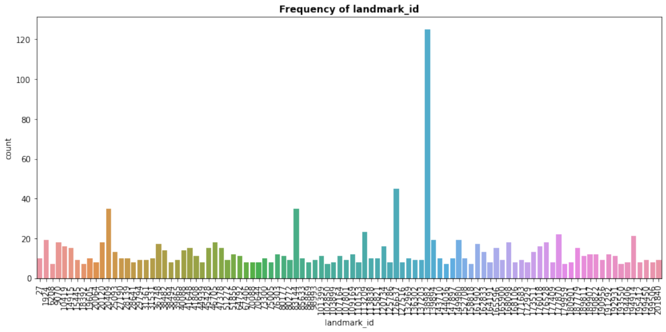
The cleaned subset of GLDv2 (GLDv2 CLEAN) consists of approximately 1.5 million images with 81,313 classes. Both GLDv2 and GLDv2 CLEAN can be used for training in this competition. Our training code is available online. Github.



## 2. Method

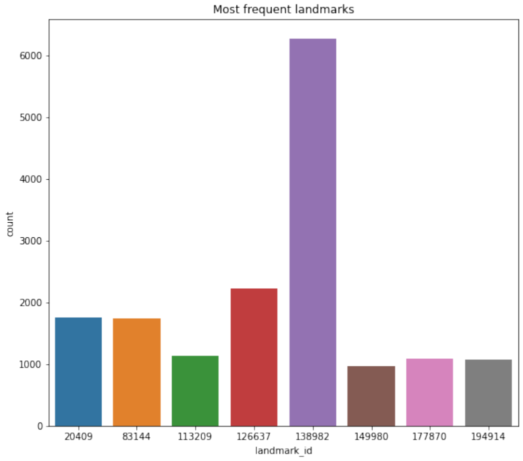
### 2.1 Overview

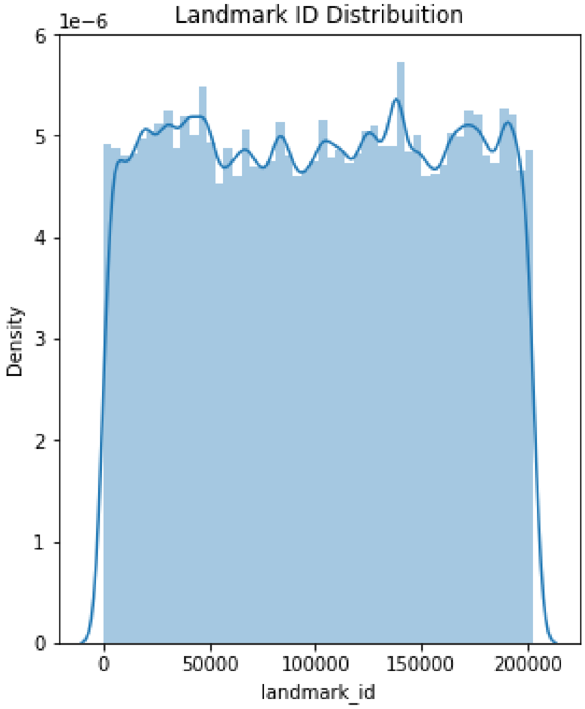
The Google Landmarks Dataset (GLDv2) is the largest dataset to date by a large margin which consists of over 5M images and 200k distinct instance labels, split into 3 sets of images: train, index, and test. There are 4132914 images in train set, 761757 images in index set. The test set consists of 118k images. The Google Landmarks Dataset v2 is designed to simulate real-world conditions therefore, poses several hard challenges. It is a large scale dataset with millions of images of hundreds of thousands of classes. All the images are given in a link form that the python script has to retrieve from the link to train and test. The outputs should represent the ID associated with each of the images, and the ID corresponds with the name of the landmarks. Considering that the v2 dataset was complied by mining web landmark images, the intra-class variability is very high considering same class images can include indoor and outdoor views.

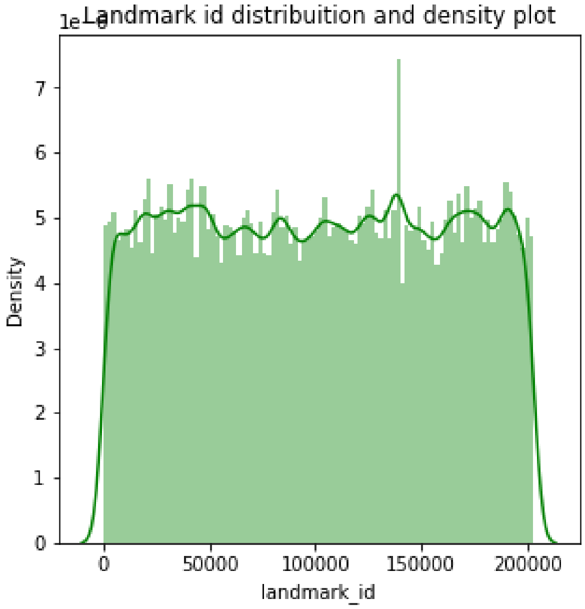


## 2.3 Data Prepossessing

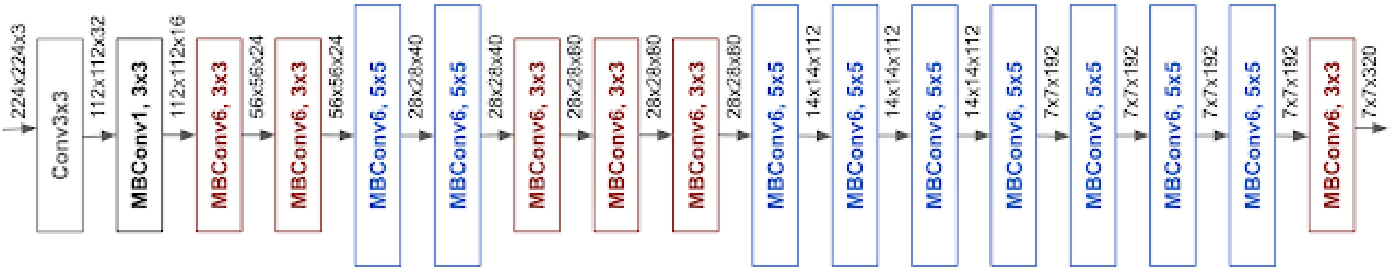
The images in our data were taken at various angles, horizon shifts, zoom etc. Similar variations are expected to be found in test data as well. The kind of parameters to use were chosen by eyeballing the images and specific values for these parameters were decided based on the accuracy received upon model training. Because our data set was large compared to other challenges, we did extensive data prepossessing steps to fully understand our data as shown below we looked at the images by the class and id.







### 2.4 Model Training



We train all our models on GLDv2 CLEAN data only. Similar to last year's solution, we planned to implement Ensemble of different architecture models including ResNet-50, EfficientNet-B5, EfficientNet-B7, and EfficientNet-B0. However, only one of the models satisfied both our efficiency and accuracy requirements. ,First, we implemented ResNet-50 and our mode overfit our data. In addition, as we did more research, EfficientNet showed a better performance in CNN, therefore, we experimented with EfficientNet-B5 and EfficientNet-B7 which did not converge within the time scope of our project and required greater computer power. Competition entries are evaluated using Global Average Precision (GAP) as metric. For each test image, we will predict one landmark label and a corresponding confidence score. The evaluation treats each prediction as an individual data point in a long list of predictions, sorted in descending order by confidence scores, and computes the Average Precision.

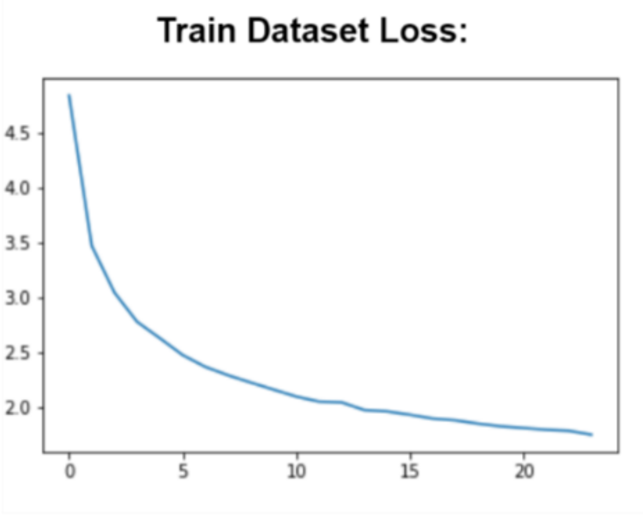
If there are (N) predictions, formatted as label/confidence pairs, sorted in descending order by their confidence scores, then the Global Average Precision is computed as:

$$GAP = \frac{1}{M}\Sigma_{i=1}^{N} P(i)rel(i)$$

We track the GAP score using softmax predictions. Properly ranking and re-ranking predictions is crucial to improve the GAP metric that is sensitive to how landmarks and non-landmarks are ranked respectively. Tailored method for differentiating between landmark and non-landmark is required to significantly improve the GAP score. We use RAdam optimizer with maximum learning rate of 1e-3 and weight decay of 1e-4 across all models. We resize the images keeping aspect ratio as well as resize to a fixed size.

## 4. Results

Considering that we trained our model for limited number of epochs on small batch size, we achieved a GAP score of 0.O2 and we plan that our GAP score should improve by increasing our epoch and batch$_{s}ize.$



## 5. Challenges

Considering that our dataset is a very expensive and large by margin, even though, it could be interpreted as a benefit for training purposes, also comes as a challenge to us as this is the first time for us to work with a dataset this scale. The Google Landmarks Dataset v2 is designed to simulate real-world conditions therefore, poses several hard challenges. It is a large scale dataset with millions of images of hundreds of thousands of classes.

We experimented with different layers for our backbone for feature extraction and computing power is a current challenge we are facing to get the best result so far. We would like to add head layers for classification after finalizing our backbone layers in order to add diversity to our solution.

# 6. Conclusion

In this paper, we presented our solution to the Google Landmark Recognition 2021 competition. We compared the result of several different models and at last we decided to present our most efficient pipeline. After implmenting and training our baseline EfficientNet-B0, we reached a final score of 0.02 on our GAP score which would awards us with bronze on Kaggle competition. In order to improve the score, ensembling different architecture models and training for more epochs and batch sizes are needed which requires more computational power.

# References

[1] Henkel, Christof, and Philipp Singer. ”Supporting large-scale image recognition with out-of-domain samples.” arXiv preprint arXiv:2010.01650 (2020).

[2] K. Ozaki and S. Yokoo. Large-scale landmark retrieval/recognition under a noisy and diverse dataset. arXiv preprint arXiv:1906.04087, 2019.

[3] T. Weyand, A. Araujo, B. Cao, and J. Sim. Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval. In Pro- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2575–2584, 2020.

[4] K. Chen, C. Cui, Y. Du, X. Meng, and H. Ren. 2nd place and 2nd place solution to kaggle landmark recognition and retrieval competition 2019. arXiv preprint arXiv:1906.03990, 2019.

[5] Mingxing Tan, Quoc V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946 [cs.LG]

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deep Residual Learning for Image Recognition arXiv:1512.03385 [cs.CV]

[7] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. Albumentations: Fast and flexible image augmentations. Information, 11(2), 2020.

[8] Q. Ha, B. Liu, F. Liu, and P. Liao. Google landmark recognition 2020 competition third place solution. arXiv preprint arXiv:2010.05350, 2020.

[9] T. Weyand, A. Araujo, B. Cao, and J. Sim. 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, pages 2575–2584, 2020.

[10] F. Perronnin, Y. Liu, and J.-M. Renders. A family of contextual measures of similarity between distributions with application to image retrieval. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 2358–2365. IEEE, 2009.