LANDMARK RECOGNITION

Thanapat Katapornsiri Woraprat Saelo Sorawit Sunthawatrodom Waritphon Sriphrom

Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand

ABSTRACT

In this paper we describe our solutions to Google landmark recognition challenge 2019 Held on Kaggle. With a noisy data we've tried to create the best performing model to predict the landmark in the given picture. First, we explore the data and rescale it to be practical for our scale of work. then we approach this data by trying what we've tried in our homework and then try the solution that win the contest. Finally, we try other model and found inceptionV3 to be the best performer

1. INTRODUCTION

To foster progress in landmark recognition, the Google Landmarks Dataset (GLD) was released last year together with two competitions (Google Landmark Recognition and Retrieval Challenges) on Kaggle. For the recognition challenge, the training set consists of 1,225,029 images belonging to 14,951 classes whereas the test set consists of both landmark and non-landmark/distractor images for a total of 117,703 images. For each test image that depicts a landmark, one is asked to predict the correct landmark class together with a confidence score whereas for distractors one should leave an empty prediction in the submission. The evaluation metric for the competition is GAP, so that it is important to make sure distractors, if predicted, should have lower confidence scores than real landmark images. This year, Google released the second version of the dataset known as Google-Landmarks-Dataset-v2 (GLD2) together with two new competitions (Google Landmark Recognition and Retrieval 2019 Challenges) on Kaggle. For the recognition challenge, same evaluation metric is used as last year, and the test set is of similar size with 117,577 images in total of both landmark and distractor images. The training set, on the other hand, is much larger with 4,132,914 images belonging to 203,094 classes. Moreover, unlike last year, the training set was released without any data cleaning step and hence is much more diverse. In this project we decided to solve a simplified version of Google Landmark Recognition 2019 Challenge with data set consisting of 81,200 images in 100 classes.

1.1. Motivation

In pattern recognition course many student prefer to use clock drawing as a topic but every student are free to decide their own topic. In our case we'd like to try something different so we decided to choose landmark recognition as our project.

1.2. Previous Work

Everyone of us are a student in pattern recognition course but while all of us did the homework most of us didn't finish it. In addition to that some of us have done other computer vision project before.

1.3. What I Am Going to Do

First we're going to scale down the landmark recognition data set to only consist of 100 class that have the most data. which result in 81,200 images on our data set. then we're going to create a model from those data set using CNN and pre-trained model such as ResNet and Inception.

2. PROPOSED METHOD

We'll now describe our approach in detail, which compose of the following step

2.1. Dataset Cleaning

The dataset consist of 203,094 classes and 4,132,914 images which is very impractical for us to use. In our submission we choose the top 100 class sorted by how many data they have per class. even then we've still got over 80,000 image to use. we decide to initially use 10% of the data. You can go to experimental result section for performance comparison between both dataset.

2.2. Dataset Preparation

After many discussions we've decided to use 60-20-20 partitioning of training-validation-test set as opposed to the more

Model	Accuracy
CNN	0.079
ResNet50	0.178
ResNet101	0.170
InceptionV3	0.278

Table 1. Accuracy comparison on each model

traditional 80-20 partitioning of training-test set.

2.3. Model

Initially, we decided to start with a simple CNN which is one of the most basic image classification algorithm. Then we moved on to ResNet50 as our base model since it's the model that won the competition. In addition to that we'll also try other image classification model such as InceptionV3 and ResNet101. We'll run each model for 10 epoch. Using accuracy as our main comparison factor and use categorical cross entropy as our main loss function.

3. EXPERIMENTAL RESULTS

We'll now goes over our experimental result in many different method.

3.1. The Dataset

Our dataset is similar to many other image classifier dataset (eg. ImageNet) consist of the Image file and the label.However this challenge is a recognition of the landmark in the picture and not the object in the picture and we have very noisy dataset for the job containing everything from a scenery all the way to a selfie. which will surely lead to a bad accuracy.

3.2. Impact of the Size of Dataset

As we've mentioned earlier we've got 2 dataset, The full one and the 10% one. The result is very comparable (see Fig.1) and because of this result we decided to use the 10% set as our base dataset to facilitate training.

3.3. Models

We decided to try ResNet101 and InceptionV3 in addition to ResNet50 which won the competition and CNN which is our most basic algorithm. At first we expect the 3 models aside from CNN to perform comparably but the result is quite different from what we've expected. InceptionV3 performed the best followed by ResNet101 and ResNet50 which performed similarly and CNN is the worst performer. see Table 1. for detailed comparison.

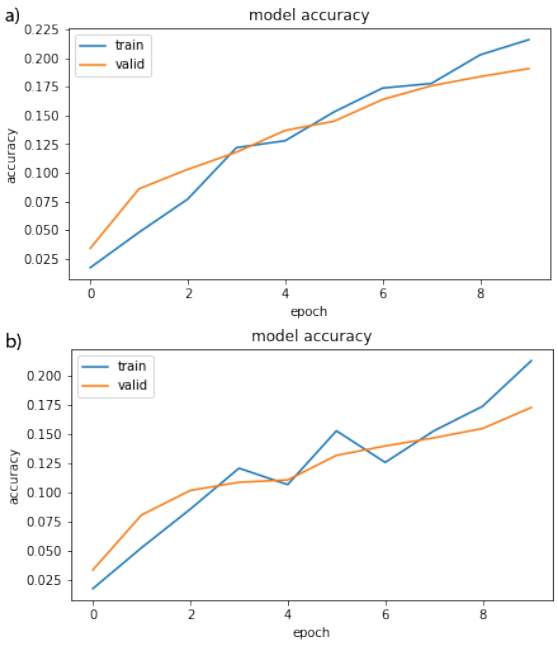

Fig. 1. a) represent model accuracy trained on 10% dataset b) represent model accuracy trained on full dataset.

4. CONCLUSIONS

In conclusion we've tried many different method of model training. Since our feature only consist of image, the model will impact the result the most. Even so, due to the noisy data the best accuracy we can manage is 27% in 10 epochs. In the future we would try other model such as Inception-ResNet and filter the dataset.