Google Landmark Recognition - Kaggle

To
Prof. Sithu Aung

Presented by
Anusha
Nikhil
Purva
Sachin
Sahana

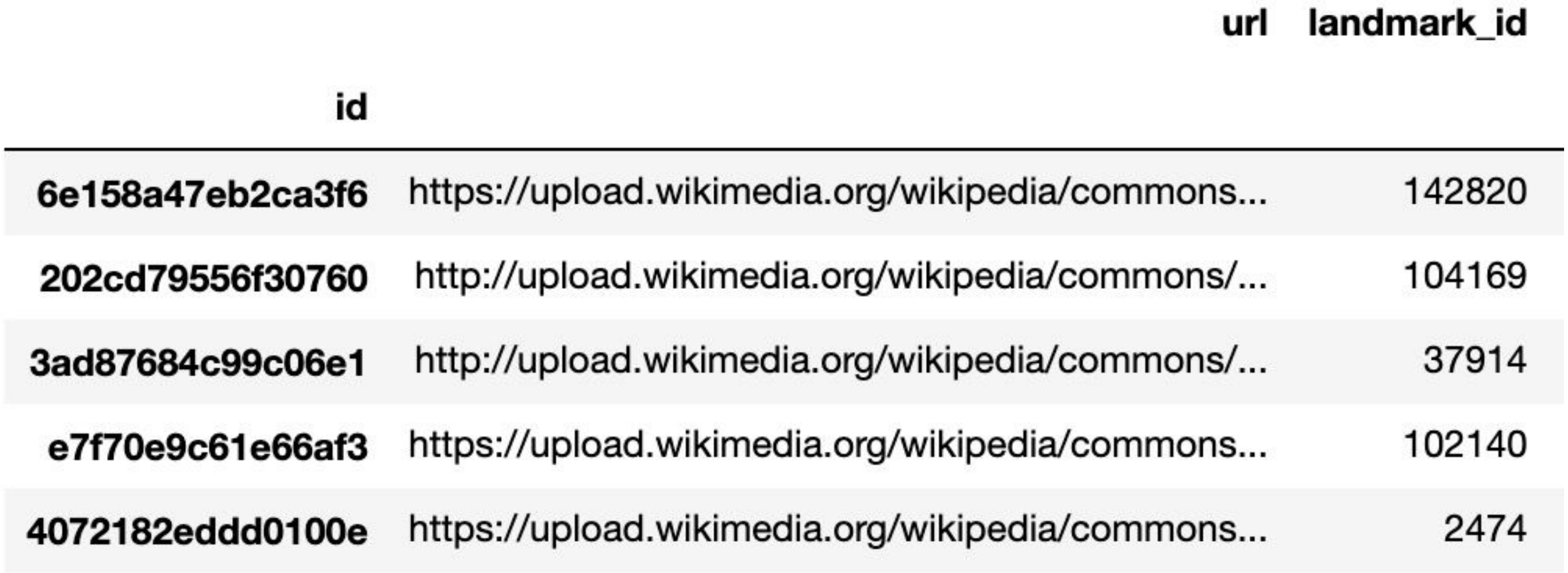
Kaggle Competition

- Google Landmark Recognition 2019 (2nd edition of the challenge)
- Build models that recognize the correct landmark

Dataset

- Dataset consists of 5,000,000 images for training.
- Nearly 200K classes.
- Test dataset consists of images for which the model should predict the corresponding landmark id.

Pandas DataFrame



	url	landmark_id
id		
6e158a47eb2ca3f6	https://upload.wikimedia.org/wikipedia/commons...	142820
202cd79556f30760	http://upload.wikimedia.org/wikipedia/commons/...	104169
3ad87684c99c06e1	http://upload.wikimedia.org/wikipedia/commons/...	37914
e7f70e9c61e66af3	https://upload.wikimedia.org/wikipedia/commons...	102140
4072182eddd0100e	https://upload.wikimedia.org/wikipedia/commons...	2474

Approach Used

- Download all the images in local system
- Pre-processing - resize to 299x299
- Data Augmentation
- Train the model (used pre-trained Xception deep CNN model)
- Predict the class of images

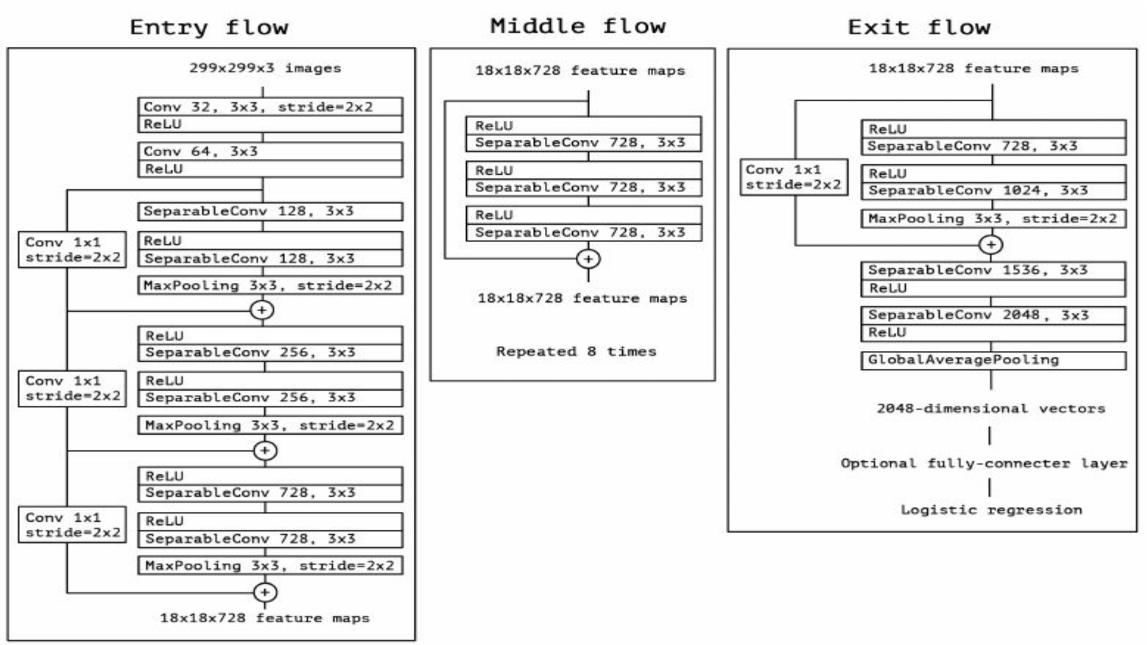
System Specifications

- For high performance computing and graphics processing:
- Trained over HPC
- GPU used: NVIDIA Tesla P100

Pre-processing:

- Normalized the input images
- Resized to 299x299 as per the input of xception architecture
- Increasing the data size by doing image augmentation using rotation, left and right shift

Xception Model Architecture



Entry flow

299x299x3 images

Conv 32, 3x3, stride=2x2
ReLU

Conv 64, 3x3
ReLU

SeparableConv 128, 3x3

Conv 1x1
stride=2x2

ReLU
SeparableConv 128, 3x3

MaxPooling 3x3, stride=2x2

ReLU
SeparableConv 256, 3x3

Conv 1x1
stride=2x2

ReLU
SeparableConv 256, 3x3

MaxPooling 3x3, stride=2x2

ReLU
SeparableConv 728, 3x3

Conv 1x1
stride=2x2

ReLU
SeparableConv 728, 3x3

MaxPooling 3x3, stride=2x2

18x18x728 feature maps

Middle flow

18x18x728 feature maps

ReLU
SeparableConv 728, 3x3

ReLU
SeparableConv 728, 3x3

ReLU
SeparableConv 728, 3x3

18x18x728 feature maps

Repeated 8 times

Exit flow

18x18x728 feature maps

ReLU
SeparableConv 728, 3x3

Conv 1x1
stride=2x2

ReLU
SeparableConv 1024, 3x3

MaxPooling 3x3, stride=2x2

SeparableConv 1536, 3x3
ReLU

SeparableConv 2048, 3x3
ReLU

GlobalAveragePooling

2048-dimensional vectors

Optional fully-connecter layer

Logistic regression

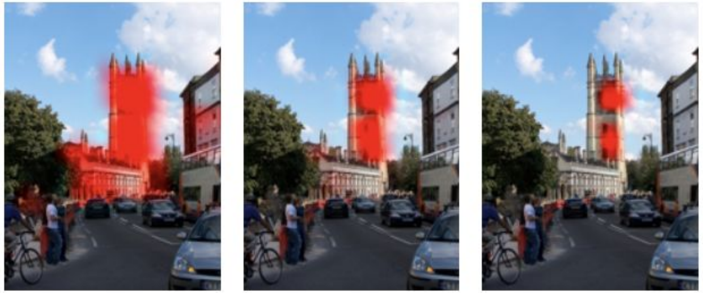
Xception Module

- Extreme Inception module.
- Depthwise Separable Convolution in Xception
- Same number of parameters as Inception V3 module, used in a different order.
- Requires no bounding box or annotation and weights are learned by taking mean of discriminative ability of each activation function in image matching, to recognize most of the images.
- Outperforms Inception v3 on Imagenet dataset.
- It’s architecture powering Google’s mobile version applications

Generalized mean pooling:

- Weighted Generalized Mean pooling ensures that the model is learning only informative features and not the whole region.

$$\mathrm{f}_k^{(a)} = \frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x.$$



$p = 1$ $p = 3$ $p = 10$

Challenges Faced

- Broken URLs and corrupt images while downloading
- Very large dataset ~5 million images
- Took 2 days to download ~200,000 images
- Very high computing power needed
- Time taken to run the program is high
- 200k classes available, but lot of classes have very less number of images per class

Code Demo