Interactive city guide

E6998: Data Science Project

Team:
Jing Zhong
Shiqi Fu
Yuanxu Zhu
Xinli Jia   Professor:
Lucie Xu   Prof. Houlihan

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science
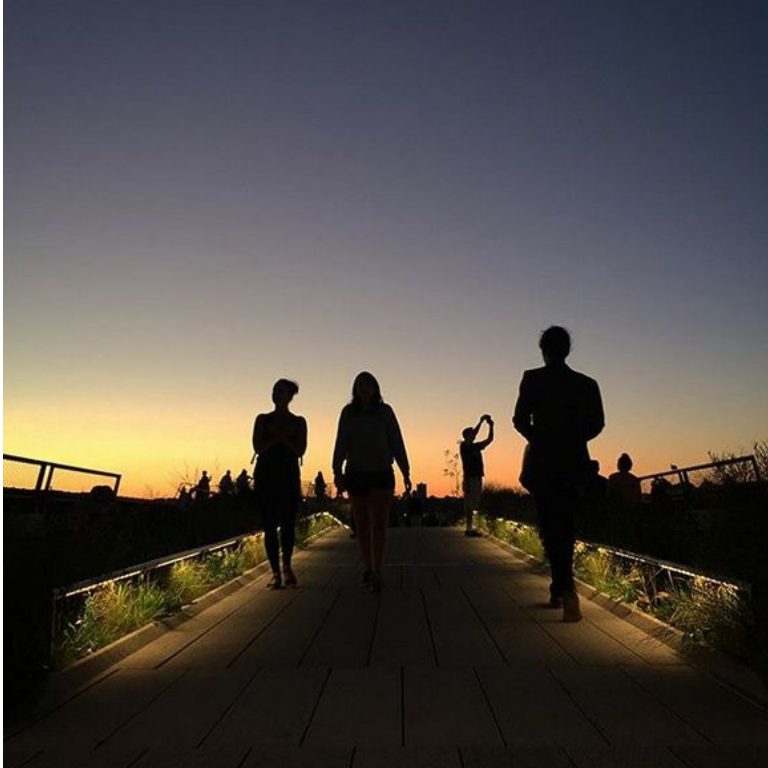
Outline

1. Context of the project

2. The data

3. The methodology

4. Prelim results

5. Next steps

Context

Context (a bit more difficult…)

Data Source

- Using Google Image Search Results
- Crawler Extension in Chrome helped downloading these image in batch
- Images in each categories were checked by hand

The data Categories

-- pictures from different places

|	
Museums	MET, MOMA, Whitney, Guggenheim, American Museum of Natural History, Frick, 911 Memorial
Skyscrapers	Empire, Rockefeller, WTC, Chrysler, Flat Iron
College	Columbia, NYU
Train Stations	Penn Station, Grand Central
Bridges	Brooklyn, Williamsburg, Manhattan
Parks	Central Park, Bryant Park, Battery Park, Dumbo, Prospect Park, Washington Square
Time Square	
Wall Street	
Statue of Liberty	

Why TensorFlow

- Popular and with Large community
- Edge tech in Machine Learning area
- Easy Python API
- Awesome other projects to stand on

What TensorFlow does
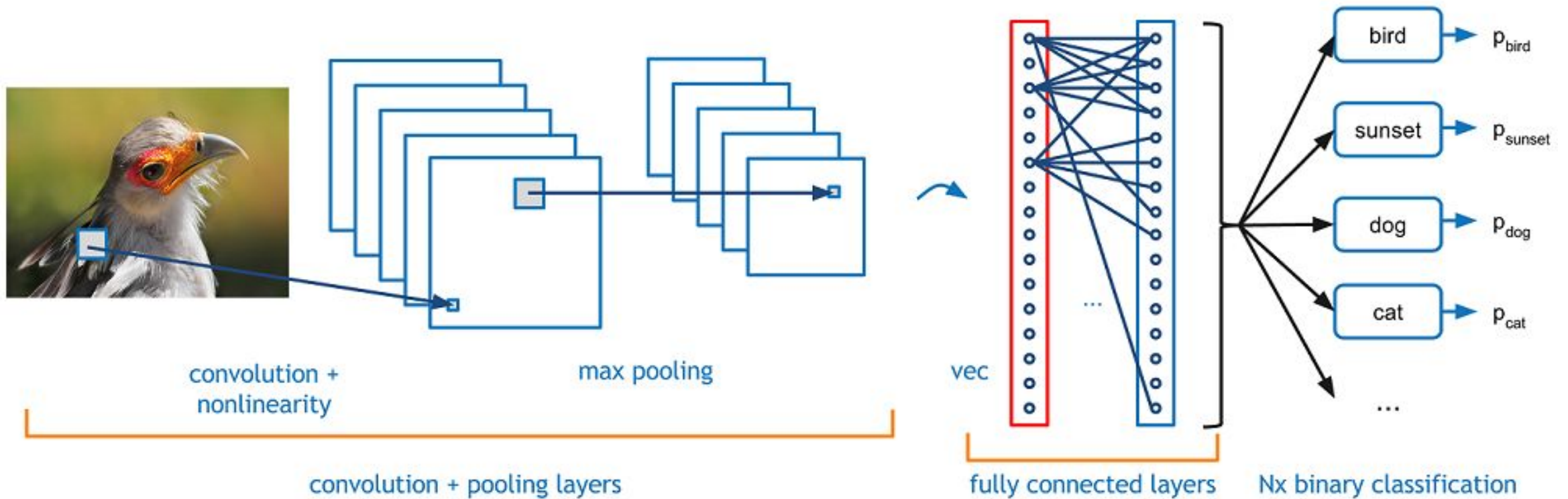
Convolutional Neural Network (CNN)

How we use TensorFlow

- Use TensorFlow to build image classifiers
- 1st classifier: used to determine the main category of the place (museum, park, skyscraper etc...)
- 2nd classifier: used to determine the exact place

Prelim Results

Empire_State_Building.jpg

JPEG image - 60 KB
Created Today, 12:44 AM
Modified Today, 12:44 AM

gcs.jpg

JPEG image - 15 KB
Created Today, 9:35 PM
Modified Today, 9:35 PM


```
zhongjing at MacBookPro in ~/desktop/DS_project on master*
$ python ./tf_files/label_image.py ./tf_files/test_imgs/Empire_State_Building.jpg
2017-10-30 16:50:57.709914: W tensorflow/core/platform/cpu_feature_guard.cc:45] The
ions, but these are available on your machine and could speed up CPU computations.
2017-10-30 16:50:57.709944: W tensorflow/core/platform/cpu_feature_guard.cc:45] The
s, but these are available on your machine and could speed up CPU computations.
2017-10-30 16:50:57.709950: W tensorflow/core/platform/cpu_feature_guard.cc:45] The
ns, but these are available on your machine and could speed up CPU computations.
2017-10-30 16:50:57.709955: W tensorflow/core/platform/cpu_feature_guard.cc:45] The
s, but these are available on your machine and could speed up CPU computations.
2017-10-30 16:50:58.902109: W tensorflow/core/framework/op_def_util.cc:333] Op Batch
to work in GraphDef version 9. Use tf.nn.batch_normalization().
empire state building (score = 0.99533)
times square (score = 0.00209)
central park (score = 0.00125)
grand central station (score = 0.00075)
statue of liberty (score = 0.00058)

This is an image of empire state building!!!
```

```
zhongjing at MacBookPro in ~/desktop/DS_project on master*
$ python ./tf_files/label_image.py ./tf_files/test_imgs/gcs.jpg
2017-10-30 21:35:47.817644: W tensorflow/core/platform/cpu_feature_guard.cc:45] The Tens
 these are available on your machine and could speed up CPU computations.
2017-10-30 21:35:47.817672: W tensorflow/core/platform/cpu_feature_guard.cc:45] The Tens
ese are available on your machine and could speed up CPU computations.
2017-10-30 21:35:47.817677: W tensorflow/core/platform/cpu_feature_guard.cc:45] The Tens
hese are available on your machine and could speed up CPU computations.
2017-10-30 21:35:47.817680: W tensorflow/core/platform/cpu_feature_guard.cc:45] The Tens
ese are available on your machine and could speed up CPU computations.
2017-10-30 21:35:48.979930: W tensorflow/core/framework/op_def_util.cc:333] Op BatchNorm
n GraphDef version 9. Use tf.nn.batch_normalization().
grand central station (score = 0.78431)
empire state building (score = 0.10570)
times square (score = 0.07518)
central park (score = 0.02128)
statue of liberty (score = 0.01353)

This is an image of grand central station!!!
```
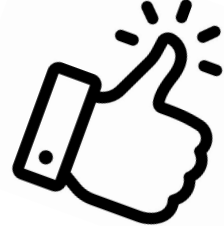
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Next Steps

- Implement the 1st layer classifier.
- Improve the performance. Maybe combine different algorithms
- Build different models for colored or black/white pictures
- Add more categories (rooftop bars, restaurants etc...)
- Build an APP(Web or Mobile) to do the prediction anytime

- Use images from Instagram to determine the hottest spots in NYC for each category

Conclusion

80% 👍
Accuracy