

# How to use deep learning on user's photos for classification on Yelp

Xiaoman Dong

[xd71@georgetown.edu](mailto:xd71@georgetown.edu)

Yigao Li

[yl942@georgetown.edu](mailto:yl942@georgetown.edu)

Taoran Yu

[ty232@georgetown.edu](mailto:ty232@georgetown.edu)

## **Abstract:**

Our group wants to improve the yelp review system, by using several deep learning methods such as Convolutional Neural Network with the yelp dataset.

Yelp is a very popular restaurant review application worldwide. There are tons of photos uploaded by users each day. Hence the goal of this project is to build a model which could automatically create tags for each user's uploaded image. For example, when user putted an image into the review system, then the system would automatically provide the most relating tag for the input image rather than asking users to manually click their preferred tags. On Yelp.com, each uploaded photo is labeled as one of the 5 categories. It can be an image of food, drink, menu, restaurant interior decoration or storefront.

## **1 Expected Result**

To put a tag for each photo, such as Food, Inside, Outside, Drink or Menu

## **2 Data**

- Data source: Yelp

Reference: <https://www.yelp.com/dataset/download>

- Data dimensions

photo\_id/food/business\_id/caption/label

- Limitations of data

Some photos may have several reasonable tags if food, menu and drink appear in the same picture, therefore it may be hard to identify which tag should be put for the image, depending on the main part of the picture or size of each part within the same picture.

- Improvement of future data

To collect more data for training set and use input image including just one item in each image. If possible, cleaning the data manually is essential.

- Input feature vector

Images with different labels

Output feature

Tags with Food/Inside/Outside/Drink/Menu

### **3. Assessment Metrics**

Our baseline is a “Caffe Classifier” that runs the CNN by means of Caffe. it’s a special form of an abstract classifier that can take different signals and perform different classification algorithms. Our current “facade” classifier is an ensemble that takes the weight average of classification results from two independently trained Caffe Classifiers. It would be quite straightforward if we decide to further incorporate new classifiers relying on other signals, such as photo captions. [1]

### **4 Approach**

- Our group would use deep CNN in form of “AlexNet” to recognize different classes
- Different methods from PyTorch, Tensorflow, Keras
- Run different models on datasets such as AlexNet, GoogLeNet, ResNet, NASNet
- To test the testing set, we would like to use k-fold cross validation for accuracy

## 5 Uniqueness

We want to try different CNN methods and hopefully to build our own neural network model for training the images. The current precision of AlexNet based on the Yelp's review was 94% and we would like to increase a little bit precision if possible

## 6 Reference

[1] "Yelp." *How We Use Deep Learning to Classify Business Photos at Yelp*,  
engineeringblog.yelp.com/2015/10/how-we-use-deep-learning-to-classify-business-photos-at-yelp.html.

[2] "Review of Deep Learning Algorithms for Image Classification ",  
<https://medium.com/comet-app/review-of-deep-learning-algorithms-for-image-classification-5fdbca4a05e2>

[3] "Yelp Restaurant Photo Classification" ,  
[http://cs231n.stanford.edu/reports/2016/pdfs/014\\_Report.pdf](http://cs231n.stanford.edu/reports/2016/pdfs/014_Report.pdf)