# Machine Learning Engineer Nanodegree

## Capstone Proposal

Ming Cheng Yen
April 12th, 2018

## Proposal

### Domain Background

As shoppers move online, it'd be a dream come true to have products in photos classified automatically. But, automatic product recognition is challenging because for the same product, a picture can be taken in different lighting, angles, backgrounds, and levels of occlusion. Meanwhile different fine-grained categories may look very similar, for example, ball chair vs egg chair for furniture, or dutch oven vs french oven for cookware. Many of today's general-purpose recognition machines simply can't perceive such subtle differences between photos, yet these differences could be important for shopping decisions.[1]

### Problem Statement

This is a competition of automatic image classification come from Kaggle, for this competition we have a dataset of furniture images, each image has one ground truth label, and our goal is classificating the furniture correctly, even they are in similarly. For this problem, we will need to build a CNN model to do image classification.

### Datasets and Inputs

All the data described below are txt files in JSON format.

#### Overview

train.json: training data with image urls and labels
validation.json: validation data with the same format as train.json
test.json: images of which the participants need to generate predictions. Only image URLs are provided.
sample_submission_randomlabel.csv: example submission file with random predictions to illustrate the submission file format

#### Training Data

The training dataset includes images from 128 furniture and home goods classes with one ground truth label for each image. It includes a total of 194,828 images for training and 6,400 images for validation and 12,800 images for testing. Train and validation sets have the same format as shown below:

```
{
"images" : [image],
"annotations" : [annotation],
}
image{
"image_id" : int,
"url": [string]
}
annotation{
"image_id" : int,
"label_id" : int
}
```

#### Testing data and submissions

The testing data only has images as shown below:

```
{
"images" : [image],
}
image {
"image_id" : int,
```

"url" : [string],
}

## Solution Statement

Since our Images have no object bounding box or part annotation,we need to build a end to end model, and the key of a good solution is fine-grained image classification, hence I plan to build a bilinear CNN model.

## Benchmark Model

According Lin and others work[2], their bilinear CNN models include classification of birds and aircrafts and cars, they got over 80% of accuracy roughly, compare with the condition of above, this competetion we have 128 classes and 194,828 images for training, for the model that I plan to build, 80% of accuracy it could be reasonable.

## Evaluation Metrics

For this competition each image has one ground truth label. An algorithm to be evaluated will produce 1 label per image. If the predicted label is the same as the groundtruth label, then the error for that image is 0, otherwise it is 1. The final score is the error averaged across all images. [1]

## Project Design

1.Checking for missing data and download images

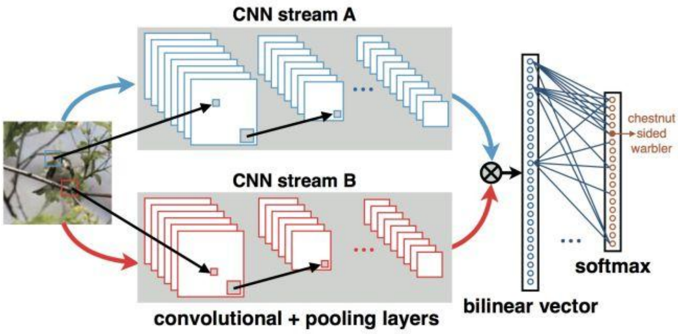
Firstable, checking all the data if they are correct,and since all the images are provided in urls, we need to download all the images.

2.Exploratory data analysis

Find the distribution of labels in training data and testing data, and since our images are download from website, we can extract the website name to see the top occurrences of websites in the data.

3.Build a bilinear CNN model

As below, a bilinear model B for image classification consists of a quadruple B = (fA, fB,P, C). Here fA and fB are *feature functions*, P is a *pooling function* and C is a *classificationfunction*. [2]



4.Tune hyper-parameter

Initial weights,add dropout layer,try different optimizer(adam,adadelta,sgd...etc)

5.Ensemble

Beside the bilinear CNN model as above, I would like to build some different architecture models , and to do ensemble selection to see if the outcome better.

## References

[1] https://www.kaggle.com/c/imaterialist-challenge-furniture-2018

[2] T.-Y. Lin, A. RoyChowdhury, and S. Maji. Bilinear CNN models for fine-grained visual recognition. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1449–1457, Sandiago, Chile, Dec. 2015.