Food: It's all about presentation?
Predicting Yelp ratings using photos of food

Team Members: Emily Brahma (ebrahma1), Nirmal Krishnan (nkrishn9), and Jonathan Liu (jliu118)

April 15, 2018

Project Proposal

Problem Definition

A recent study from Oxford University has showed empirically that "making something look good makes it effectively taste better too." The researchers found that "even with such basic dishes, thoughtful presentation meant diners found the food more flavourful." In this study, we would like to test this theory rigorously, using photos of food to predict the rating of a specific restaurant on popular food-rating app Yelp.

Potential Solution to Problem

We plan on using convolutional neural networks for this task. We will vary the size of various convolutional parameters including kernel, stride, and number of layers. Additionally, we would like to test the effects of transfer learning towards this problem, so we also plan on using the initial layers of resnet18 with pre-trained weights. We may add additional layers and solely train on these added layers to reduce computational costs.

In addition to pictures of food, as a stretch goal, we would also like to integrate pictures of the restaurant to see if our model can pick up on restaurant ambiance and appearance. We would also like to integrate features about the business such as latitude/longitude and category of cuisine (Italian, Mexican, etc).

Dataset

For our dataset, we are using the Yelp Open Dataset. This consists of each business's data, review data, general user data, user checking data, and photos with associated metadata. We will be primarily focusing on the photos and their metadata, and the reviews for each business.

Team Member Assignments

For this project there are three main parts: data collection and pipelining, data filtering and verification, and model design. However, since these parts are sequential, it is not easy or obvious to divide the project amongst us. Therefore, we all plan on working together at each step, with one of us leading each one week sprint.

- Week 1 (April 15- April 21): Data Collection and pipelining led by Jon
- Week 2 (April 22- April 28): Data filtering and verification led by Nirmal
- Week 3 (April 29- May 5): Model Design led by Emily