Food Image Recognition using Convolutional Neural Networks

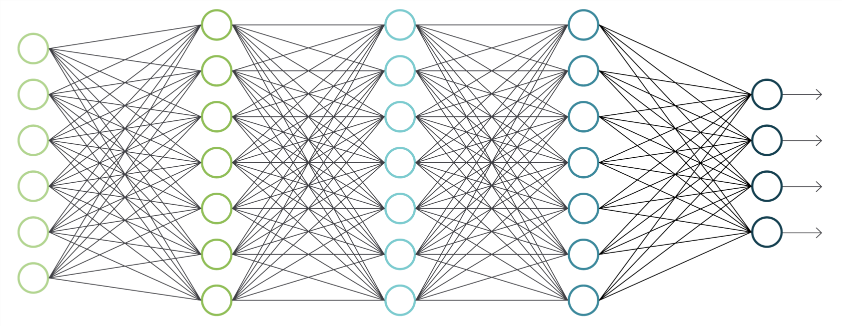
Richa Sethi

March 2021

Introduction

- Food is a form of social currency
- Photo-driven digital experience
- Image classification with CNNs for image labeling
- Food image classifier built using fastai
- Input: food image, output: correct label

Food Image Recognition



Macaron?

Input

Model

Label

Dataset Overview

- Dataset obtained from https://www.kaggle.com/kmader/food41
- Consisted of 101 classes of food with 1000 images for each class
- 750 noisy training images, 250 clean test images
- Due to GPU and constraints, 10 international food classes were chosen
- Classes: Bibimbap, Chicken curry, Macarons, Pad thai, Paella, Peking duck, Pho, Ramen, Samosa, Takoyaki

Mislabeled/anomalous Images

Bibimbap

Chicken curry

Macaron

Peking duck

Ramen

Samosa

Dataset preparation

- **Data Loading**
  - Data Loader built using Data Block API
  - convert 20% of train data into validation set

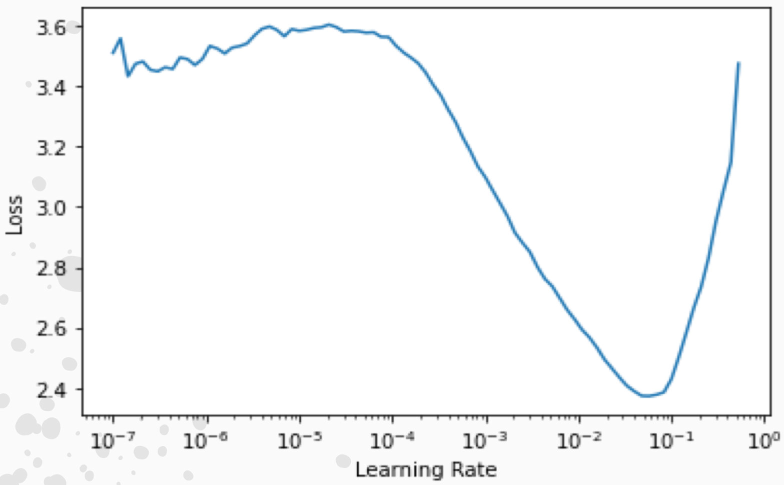
- **Data Transformation and Augmentation**
  - Resized all images to the same size on CPU
  - Batch transformation of 64 images on GPU
  - Horizontal and vertical flipping, zooming, rotating, lighting, warping, etc. performed on the train set

Transfer learning

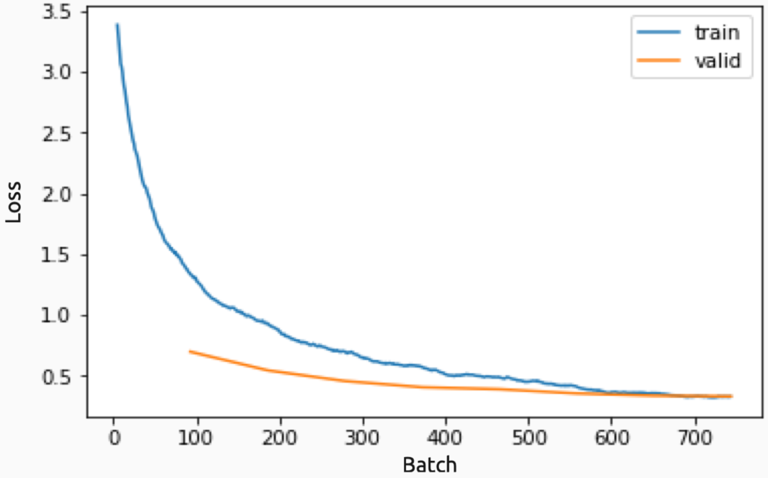
- ResNet50 as pretrained model
- Transfer learning implemented by loading the ImageNet and freezing the base layers while removing the top layers
- Top layers then replaced with trainable layers
- Data normalized to equalize the pixel intensities of RGB channels

Model training

- Model training was performed by first training the network using Leslie Smith's 1cycle policy



lr_finder on ResNet 50 model



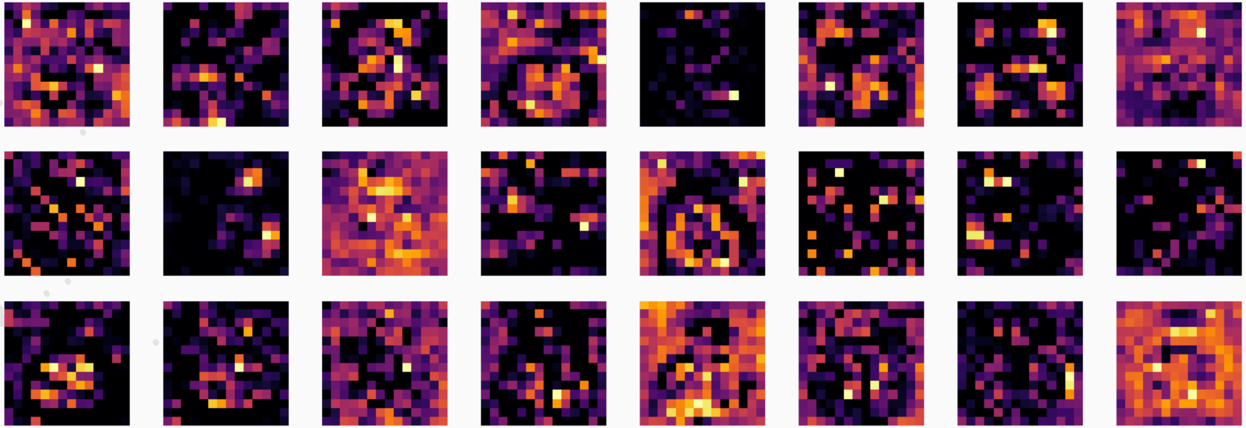
Loss convergence after 8 epochs

Top Losses


Prediction/Actual/Loss/Probability
Pad thai/Pho / 11.76 / 1.00
Ramen/Bibimbap / 10.69 / 0.75
Ramen/Takoyaki / 8.84 / 0.95
Paella/Takoyaki / 8.62 / 0.81
Chicken curry/Pho / 8.49 / 0.77
Bibimbap/Chicken curry / 7.19 / 1.00
Peking duck/Chicken curry / 6.94 / 1.00
Bibimbap/Pho / 6.74 / 0.82
Peking duck/Chicken curry / 6.70 / 0.99

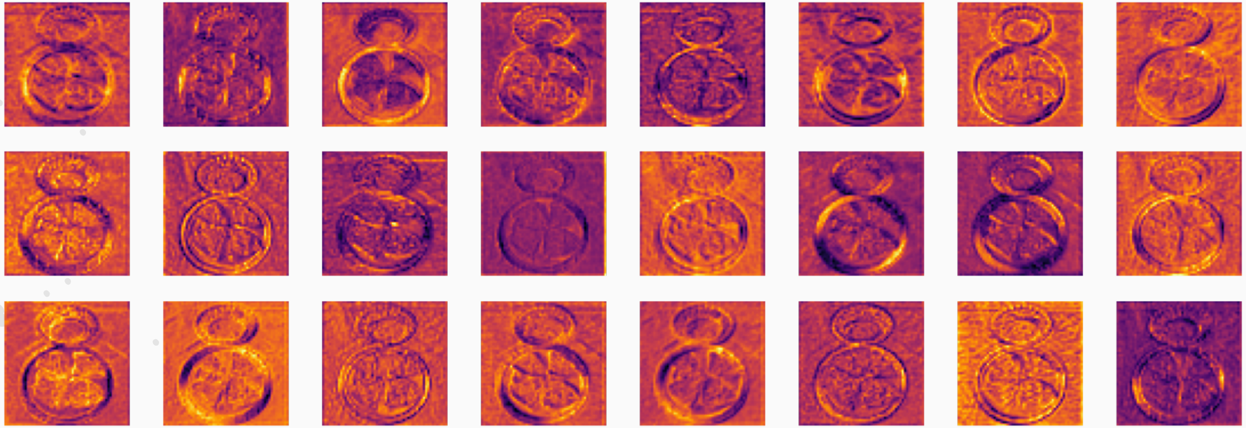
Activation Layer Visualization

Deepest layers



Activation Layer Visualization

Intermediate layers



Activation Layer Visualization

Top layers

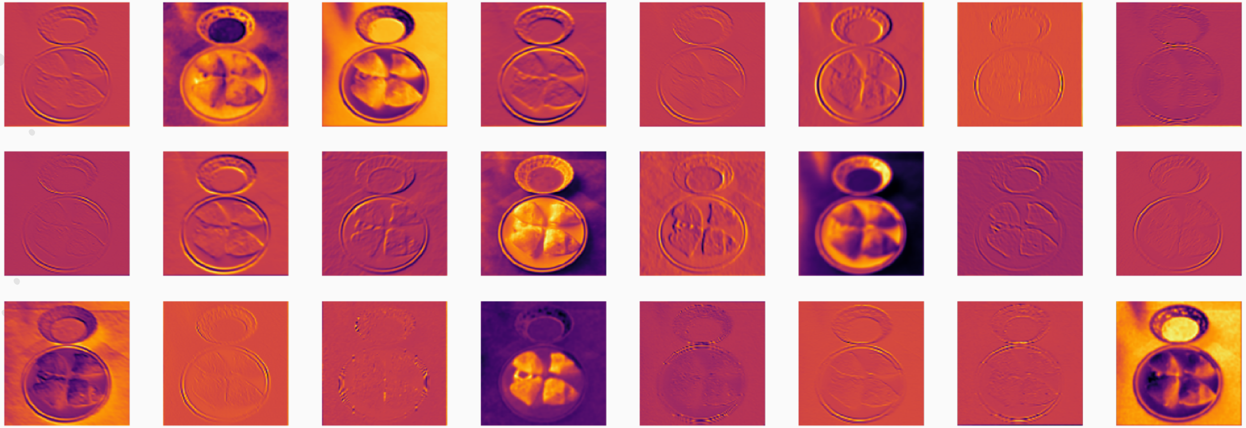


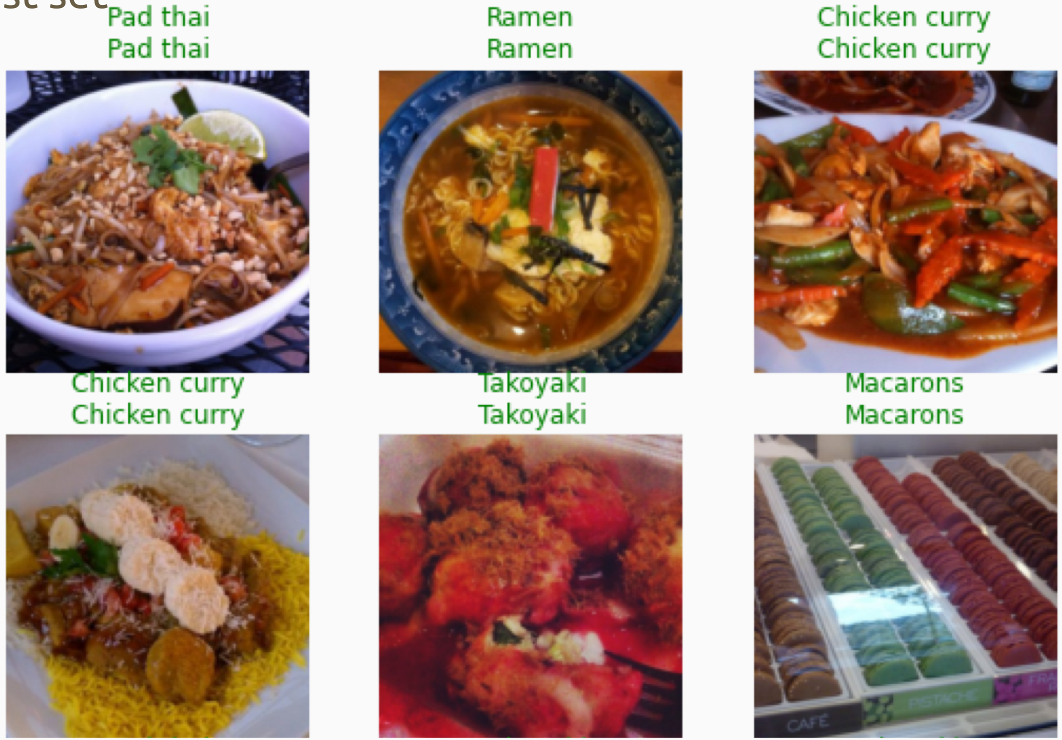
Image heatmap

Original

Heatmap

Inference

- 94% accuracy on test set



Results

Select your food!

Upload (1)

Classify

```
[('Paella', 0.9999978542327881),
 ('Bibimbap', 1.2092289125575917e-06),
 ('Ramen', 4.115964316042664e-07)]
```

Enter image url!

https://thegirlonbloor.com/wp-content/uploads/20

Classify

```
[('Pad thai', 0.5531837344169617),
 ('Bibimbap', 0.4448730647563934 3),
 ('Ramen', 0.0009309540037065744)]
```

Conclusions

- Transfer learning was used to identify and classify 10 classes of
- ResNet 50 with image transformation and augmentation showed the best results with accuracy as high as 94%
- Transfer learning was successful because the earlier pre-trained layers had already learned a lot of the general features needed to identify food images
- Intermediate activation layers were visualized to understand how the model works on a granular level and which features help the model identify the food class
- GUIs were build around to result to predict the food class along with the probability from the local machine as well as from web.

Future Work

- Use deeper network or newer structure, such as ResNet200 or EfficientNet
- Use an exhaustive hyperparameter search
- Improve model performance by adding bounding boxes to the images
- Train models to recognize images within a subset of food (e.g. vegetables vs fruits vs bread vs noodles vs. cakes etc
- Extend the model to out ingredients, recipe and nutritional information along with the food prediction