Computer Vision & Food Recognition

Simon Bedford
September, 2016

- Food as content

- Opportunities for computer vision techniques + food

- Models

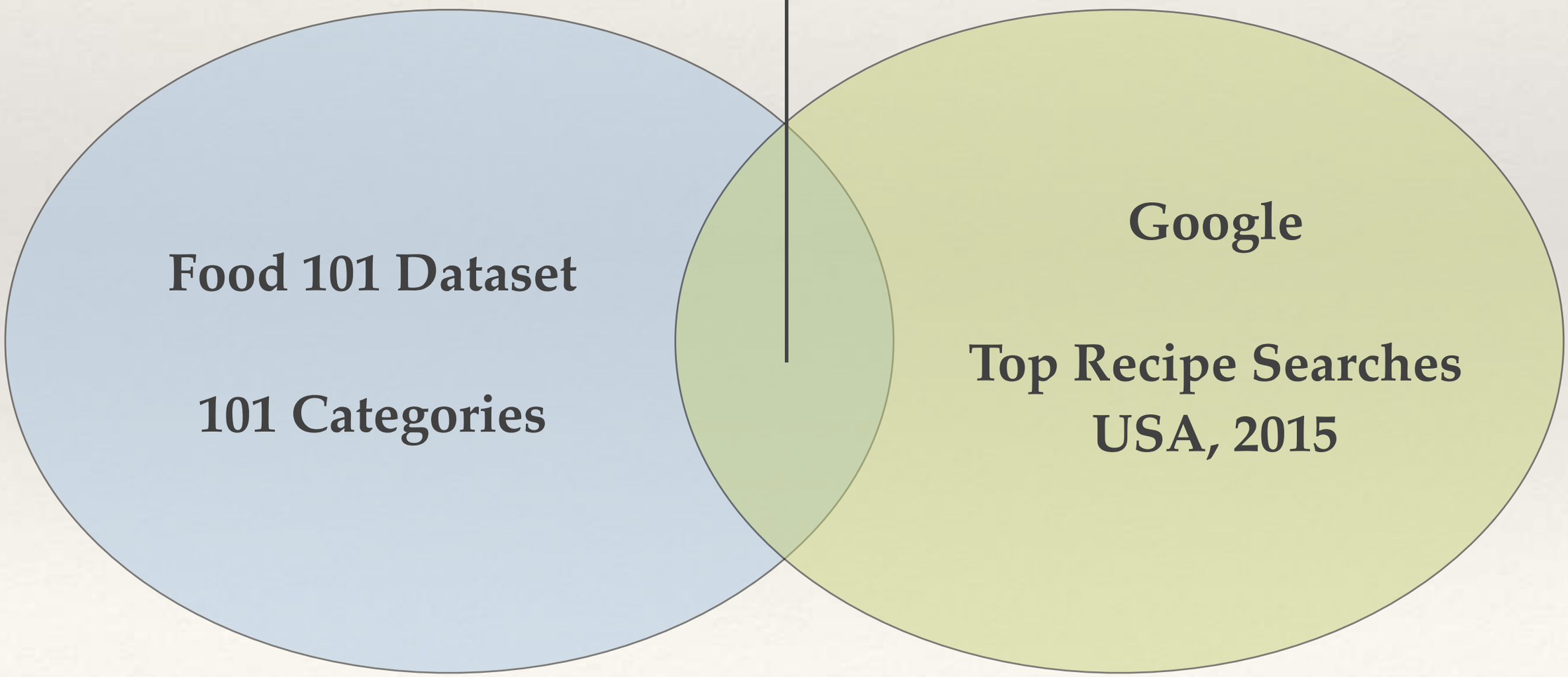
- Metrics

The dataset

- Food-101 Data Set from the ETH Zurich Computer Vision Laboratory
- 101 Categories
- 1,000 images per category
- Most common shape (512, 512)

Focus on 12 categories

- Pork Chop
- Guacamole
- Hamburger
- Chocolate Cake
- Lasagne
- Apple Pie
- Fried Rice
- Steak
- French Toast
- Cheesecake
- Carrot Cake
- Pizza



Mixed data quality

Differences in RGB histograms between image categories

Machine-learning: models + features

Models:

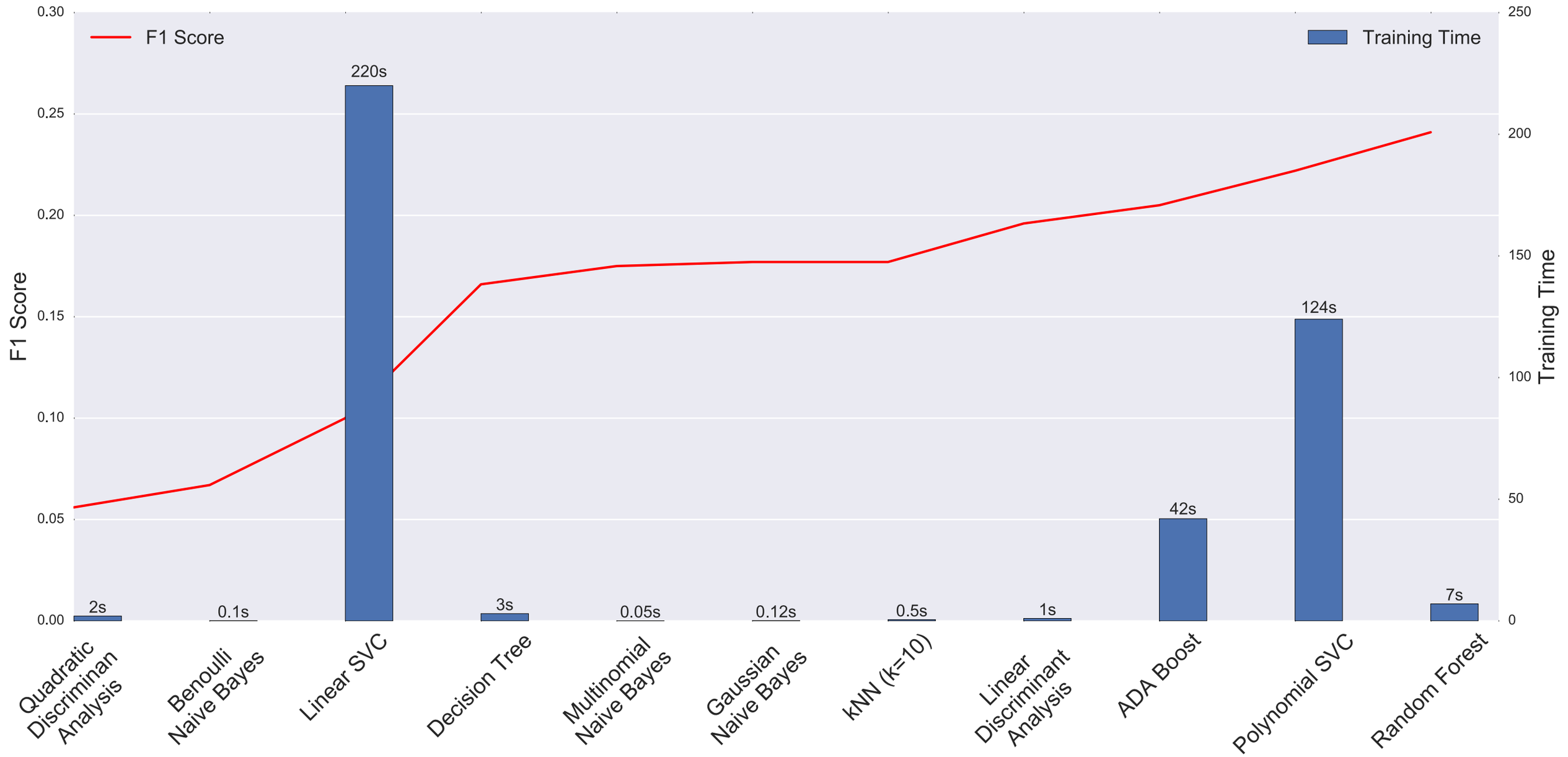
- k-Nearest Neighbours
- Support Vector Machines
- Decision Trees
- Random Forests
- ADA Boost Classifier
- Naive Bayes Classifiers
- Linear & Quadratic Discriminant Analysis

+

Features:

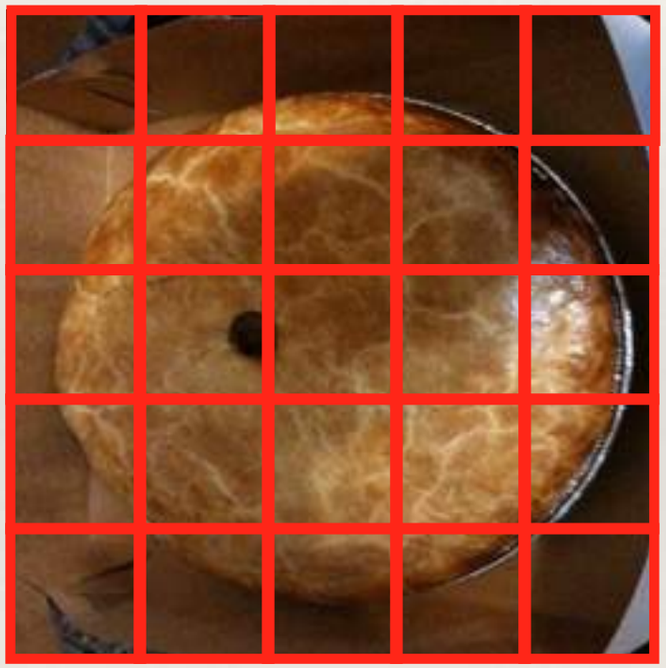
- RGB Histograms
- Individual Pixel Values
- Number of Edges
- Number of Corners
- Unsupervised methods (e.g, Principal Component Analysis, k-Means Clustering)

Comparison of classifiers



Best machine learning procedure

Image divided into 32 x 32 grid (256 cells in total)



For each cell calculate:

- Average red pixel value
- Average green pixel value
- Average blue pixel value
- Number of edges
- Number of corners

Chain all features together into one vector (..1,280 features..)

Random Forest

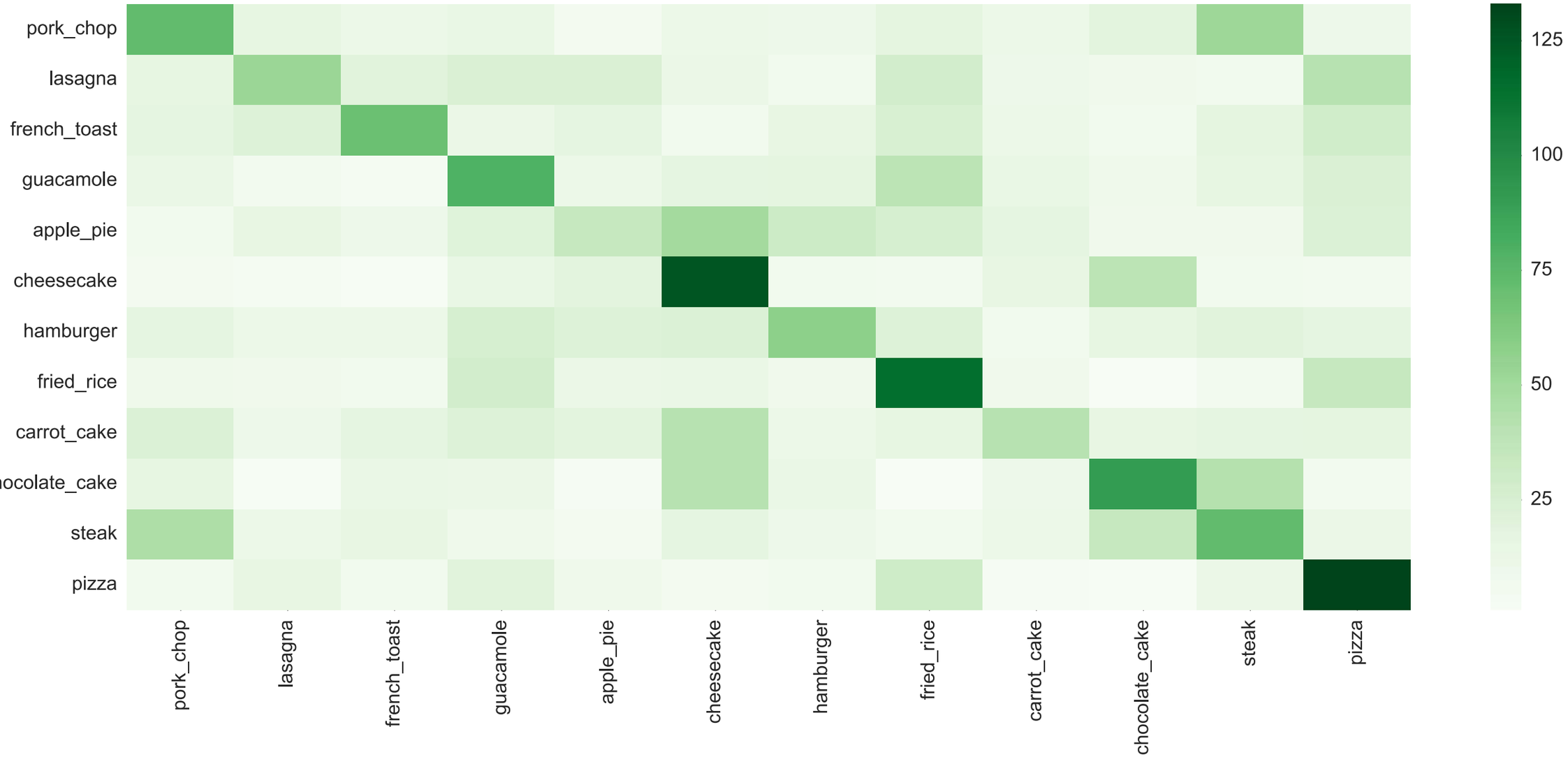
+

Grid Search for Hyperparameter Optimization

Per-class results

	Precision	Recall	F1-Score	Support
Pork Chop	0.29	0.29	0.29	250
Lasagne	0.31	0.21	0.25	250
French Toast	0.37	0.28	0.32	250
Guacamole	0.28	0.32	0.30	250
Apple Pie	0.19	0.14	0.16	250
Cheesecake	0.35	0.50	0.41	250
Hamburger	0.30	0.23	0.26	250
Fried Rice	0.35	0.46	0.39	250
Carrot Cake	0.26	0.16	0.20	250
Chocolate Cake	0.36	0.36	0.36	250
Steak	0.26	0.29	0.28	250
Pizza	0.38	0.53	0.44	250

Confusion Matrix



- CNNs

Feature extraction approach

AlexNet architecture

FC Layer 6

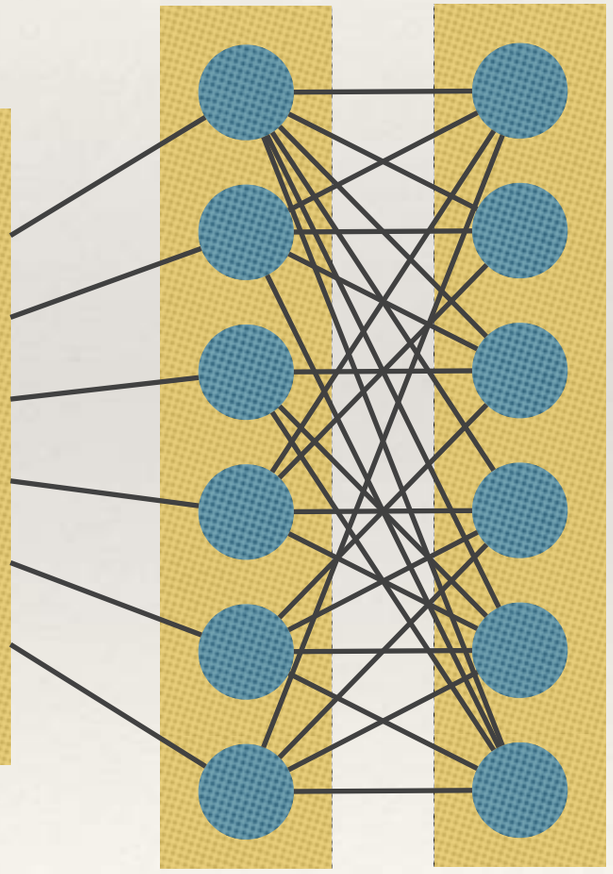
FC Layer 7

'Break' the network architecture at this point and use the outputs of FC layer 7 as features

Standard machine learning model

Input Image

Convolution layers
Non-linearities

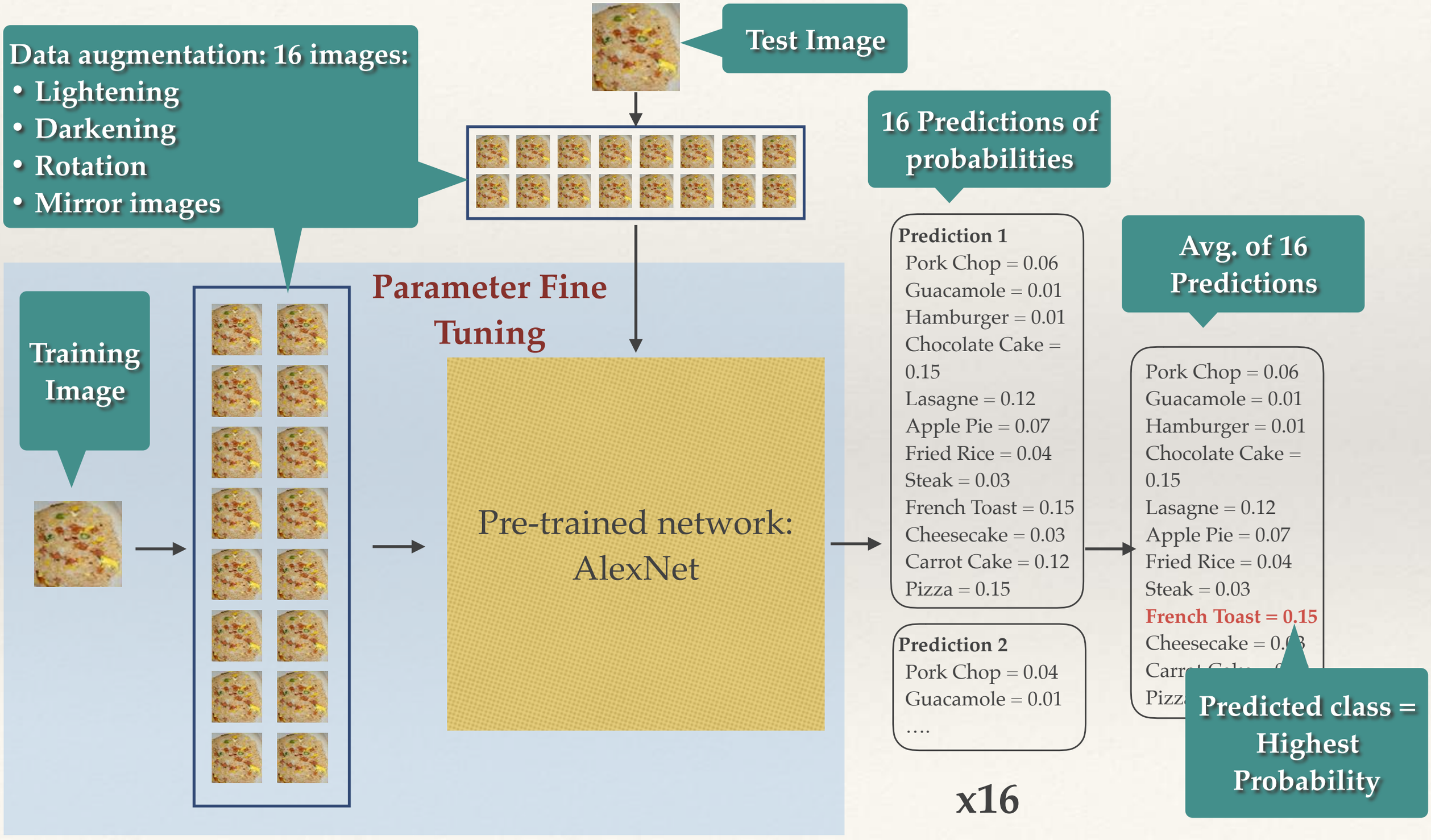


Linear Support Vector Machine Classifier

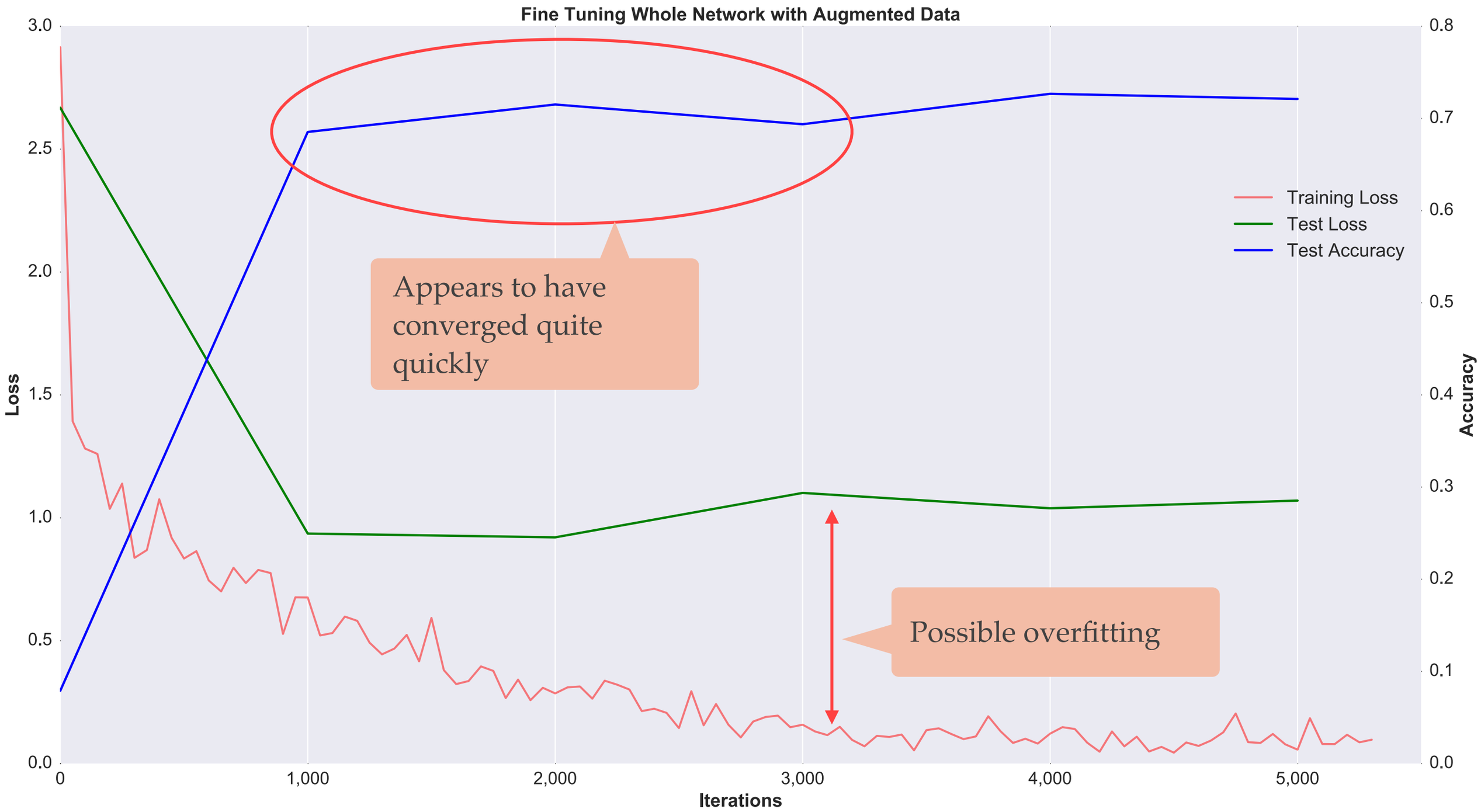
Per-class results

	Precision	Recall	Old F1-Score	New F1-Score	Support
Pork Chop	0.55	0.51	0.29	0.53	211
Lasagne	0.67	0.61	0.25	0.64	187
French Toast	0.64	0.62	0.32	0.63	208
Guacamole	0.90	0.89	0.30	0.90	189
Apple Pie	0.55	0.60	0.16	0.57	194
Cheesecake	0.70	0.67	0.41	0.68	206
Hamburger	0.71	0.76	0.26	0.73	200
Fried Rice	0.81	0.88	0.39	0.84	213
Carrot Cake	0.66	0.65	0.20	0.66	205
Chocolate Cake	0.70	0.76	0.36	0.73	178
Steak	0.54	0.51	0.28	0.52	211
Pizza	0.82	0.78	0.44	0.80	198

Fine Tuning Approach



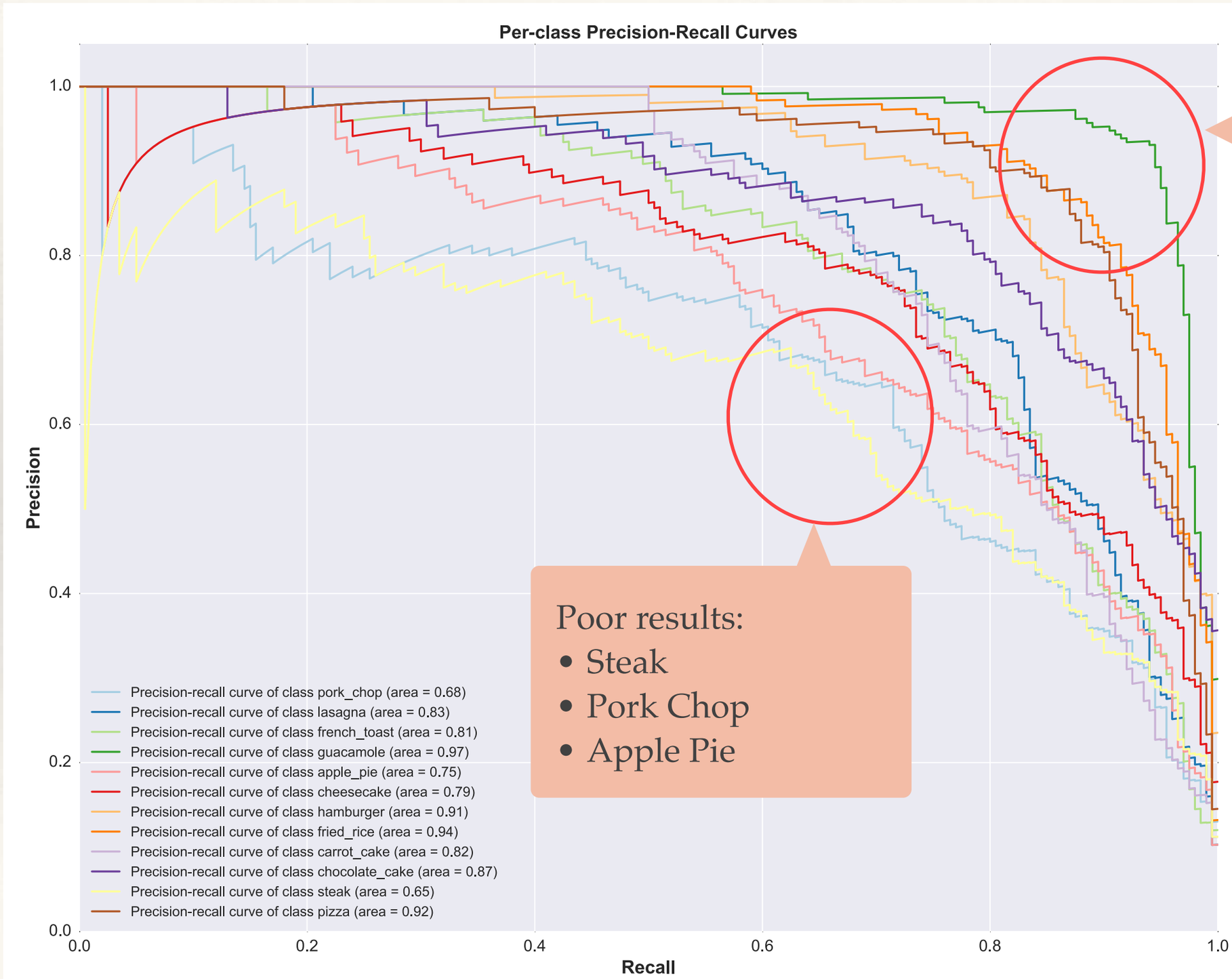
Training Curve



Per-class results

			F1 Scores		
	Precision	**Recall**	**Machine Learning**	**Extracted Features**	**Fine Tuning**
Pork Chop	0.70	0.54	0.29	0.53	**0.61**
Lasagne	0.73	0.76	0.25	0.64	**0.74**
French Toast	0.72	0.71	0.32	0.63	**0.72**
Guacamole	0.90	0.93	0.30	0.90	**0.92**
Apple Pie	0.61	0.74	0.16	0.57	**0.67**
Cheesecake	0.73	0.72	0.41	0.68	**0.72**
Hamburger	0.76	0.86	0.26	0.73	**0.80**
Fried Rice	0.71	0.92	0.39	0.84	**0.80**
Carrot Cake	0.74	0.74	0.20	0.66	**0.74**
Chocolate Cake	0.86	0.73	0.36	0.73	**0.79**
Steak	0.70	0.55	0.28	0.52	**0.62**
Pizza	0.89	0.82	0.44	0.80	**0.86**

Precision-Recall



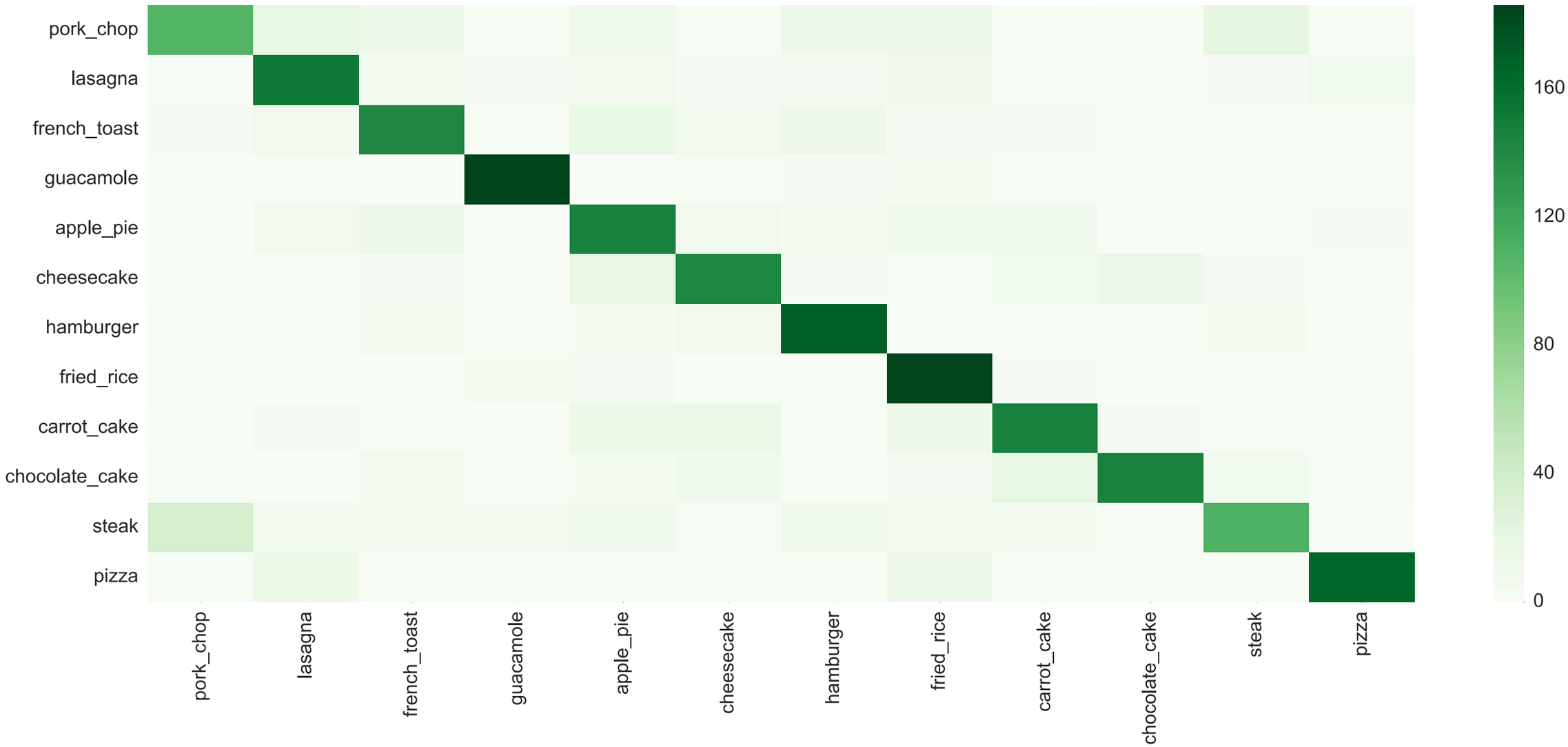
Good results:

- Guacamole
- Pizza
- Fried Rice

Poor results:

- Steak
- Pork Chop
- Apple Pie

Confusion Matrix



Fail Cases Examples

Steak Predicted as Pork Chop

Pork Chop Predicted as Steak

Future Work for Optimisation

- Fine-tuning hyper-paremeters e.g., dropout rate
- Increase training batch size (currently 150)
- More data augmentation
- Fine tune more recent models e.g., VGGNet, GoogleNet
- Ensemble of fine-tuned models

Machine Learning vs CNNs for Image Classification

	Machine learning	CNNs
Classification Accuracy		
Training Speed		
Testing Speed		
Ease of getting started		
Resources required		
Feature Selection		
Overall		

- Able to test 40+ models quite quickly
- Makes you think more about underlying images & their content
- Ultimately, not accurate enough

- Harder to get going and needed to use external computing resources
- But once started, training models was easier than expected
- Increase in accuracy makes up for everything else

Note: Personal opinions based upon experience with this project

Recommendations

1. Expand the model to include all 100 food categories from the existing dataset.
2. Seek to increase the number of images by looking for other sources of data.
3. Invest more time in optimising the model
4. Consider a pilot based on using a smaller set of 10-15 consolidated food categories.