What's on My Plate? Identifying Different Food Categories

Mo Islam, Surbhi Maheshwari, Nate Nunta
{moislam, surbhim, nnunta}@stanford.edu

Introduction

Balanced diet is an important component of a healthy lifestyle. Lack of awareness is a big hurdle in achieving that. Our project is an effort towards an on-demand nutrition tracker application. We take food images as input and classify them into one out of 101 food categories. To do so, we use transfer learning on four deep learning based image recognition models - InceptionV3, VGG19, ResNet50 and Xception. Our best model can predict the right category 60% across the 101 categories.

Data and Pre-processing

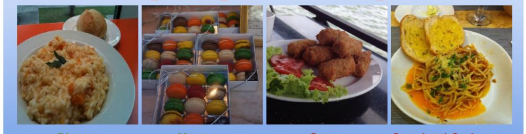
Data used: 40K images of 101 food categories like apple pie, club sandwich, ice-cream, fried rice, seaweed salad, tacos and waffles.

Original source: Food-101 dataset published by Bossard, Lukas, Matthieu Guillaumin, and Luc Van Gool. It has 101K images - 1K images of each of the 101 categories.

Distribribution:
Training: 30K images, Dev: 5K images, Test: 5K images

Pre-processing: Cropped all images to 299*299 size to match the input requirements of various models.

Sample Inputs and Outputs



Risotto | Macaron | Gyoza / Spring Rolls | Spaghetti Carbonara / Spaghetti Bolognese

References

- Bossard, Lukas, Matthieu Guillaumin, and Luc Van Gool. "Food-101–mining discriminative components with random forests." *European Conference on Computer Vision*. Springer, 2014.
- Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *arXiv* 2016.
- Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
- Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." *arXiv preprint arXiv:1605.07678* (2016).

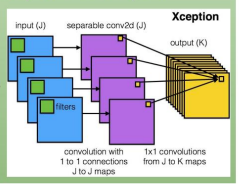
Model and Hyperparameters

We applied transfer learning to four different models to test three different families of models:

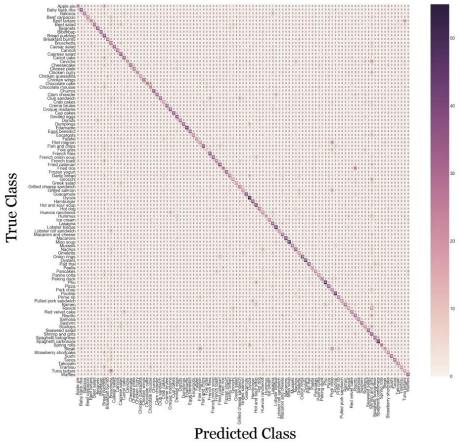
- InceptionV3 - 48 layers - Highly efficient computations
- VGG19 - 19 layers - Known for its simple 3*3 convolutions
- ResNet50 - 50 layers - Bypasses 2 layers to feed input ahead
- Xception (an independent extension of Inception family) - Depth wise separable convolutions on InceptionV3

Step 1: Architecture search on four models with 6 additional layers that we trained, dropout of 0.5 and 10 epochs each. Xception performed the best.

Step 2: Hyperparameter search on the Xception model across different parameters and early stop.



Confusion Matrix of 101 Food Classification



Results and Discussion

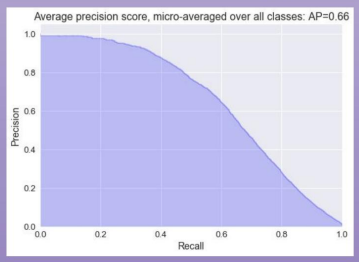
When calculating accuracies, we used "Top-1" accuracy, counting an image as accurately predicted if the top returned softmax percentage matched with the input image.

Model	Training accuracy	Training loss	Val. accuracy	Val. loss
InceptionV3	36%	2.62	36%	2.83
VGG19	35%	2.63	37%	2.84
ResNet50	62%	1.40	0.9%	12.26
Xception - base model	58%	1.58	44%	2.85
Xception - final model	**63%**	**1.38**	**60%**	**1.73**

Our final best performing model has following features:

- 7 additional layers, Categorical Loss Entropy, Adam Optimizer
- Two dropout layers of 0.7 and 0.7
- Pass 1 of 30 epochs on 7 layers
- Pass 2 of 20 epochs on 7+4 layers
- Pass 3 of 8 epochs on 7+8 layers
- Precision - 0.68, Recall - 0.60, F1 score - 0.62

These is a promising result as it is in-line with the expectation given Xception is the best performing model in literature. Hyperparameter search improved accuracy and reduced overfitting significantly..



Future work: The model can be made more accurate by training it on more examples. Additionally, a lookup table can be use to map the food to its nutritional data. Lastly, the model can be made to learn new food categories as it sees them.