Your Food Friends: Food Plate Segmentation and Classification

Winston Li, John Ryan Byers, Nathan DePiero

BROWN

CSCI 1430

Summary

YourFoodFriends is a segmentation and multi-class classification ensemble for detecting foods on a plate.

Masks of potential food candidates were generated via MetaAI's Segment-Anything-Model. With the masks, cropped images were inputted to a fine-tuned InceptionResNet V2 convolutional neural network for labeling.

From these outputs, the original food plate image is labeled with predicted food names.

Problem Statement

Millions of Americans are visually impaired and might have trouble identifying food placed in front of them. Tourists in new countries might not know a dish and are too embarrassed to ask what it is.

With a simple picture, YourFoodFriends offers an aid to accurately identify the foods on a plate or tray. Trained on 101 foods ranging from all cultures and cuisines, our classification ensemble seeks to help the visually impaired and tourists in need.

Data Used For Classification

Food-101 (Bossard et al.)

- Contains 101 different labels with 1000 images each
- Pre-split into 75,750 training and 25,250 validation samples
- Training samples contain deliberate mis-labelings

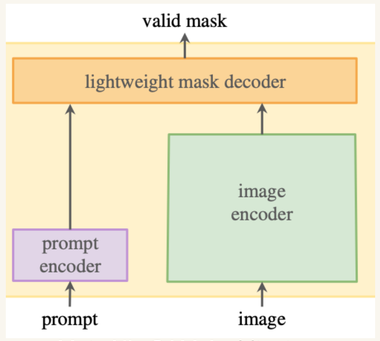
Dumplings

Pancakes

Ice Cream

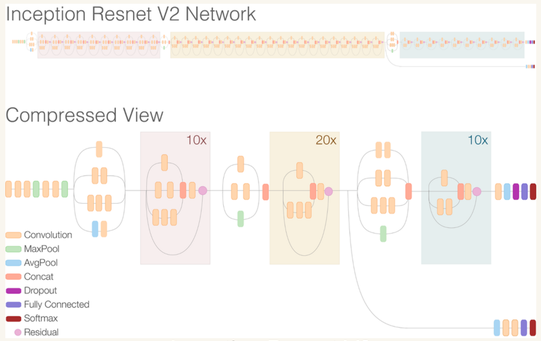
Algorithm Specifications

Segmentation



MetaAI's SAM Architecture

Classifier



Inception Resnet V2

Architecture

Input RGB image of food tray

Segmented images

SAM

Our Classifier

Output food labels

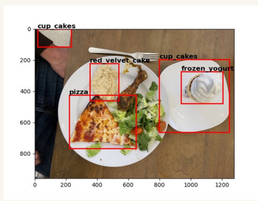
- Hotdog
- French Fries
- Pizza

Results

Original Ratty plate of food

SAM mask unfiltered output

Output classified food tray

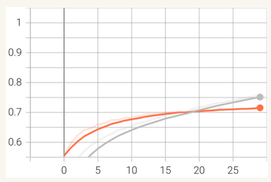


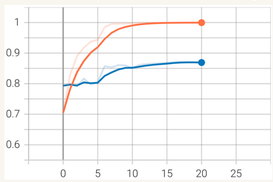
Classifier Accuracy

Model #5 with IRNv2 Base

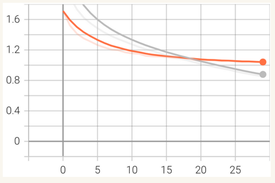
Model #5 Fine Tuning

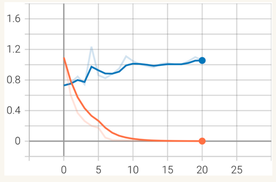
Accuracy for Train and Test per Epoch





Sparse Categorical Loss per Epoch





Gray: Training Accuracy
Orange: Validation Accuracy

Orange: Training Accuracy
Blue: Validation Accuracy

Various Classification Model Performances

Model #	Base	Head	Train Acc	Test Acc
1	VGG16	f,4096,d05,4096,d05	0.4645	0.5023
2	Xcept	g	0.7007	0.6727
3	FT MNv2	g	0.9999	0.7895
4	FT ENB3	g	0.9998	0.8265
5	FT IRNv2	g,2048,d05	0.9997	0.8713

Key: f=flatten; g=GlobalAveragePooling2D(); # = Dense(#)
d05 = dropout(0.5); FT = fine-tuning

References

[1] Martin Abadi et al. TensorFlow: Large-scale machine learn-ing on heterogeneous systems, 2015. Software available from tensorflow.org
[2] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In European Conference on Computer Vision, 2014
[3] Alexander Kirillov et al. Segment anything. arXiv:2304.02643, 2023
[4] Karen Simonyan and Andrew Zisserman. Very deep convolu-tional networks for large-scale image recognition, 2015
[5] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. Inception-v4, inception-resnet and the impact of residual con-nections on learning. CoRR, abs/1602.07261, 201

Acknowledgements

We would like to thank: Professor Tompkin, our mentor TA Anh Duong, Oscar CCV, and Brown Grid Engine