

FOOD IMAGE RECOGNITION

with Neural Network

DSI Capstone Project

Jetnipat Sarawongsuth (Boss)

DIABETES

DIABETES IS
ON THE RISE



422 MILLION
adults have diabetes

Risk factors for type 2 diabetes

Genetics, age and family history of diabetes can increase the likelihood of becoming diabetic and cannot be changed.
But some behaviours that increase risk can:



Unhealthy diet



Physical inactivity



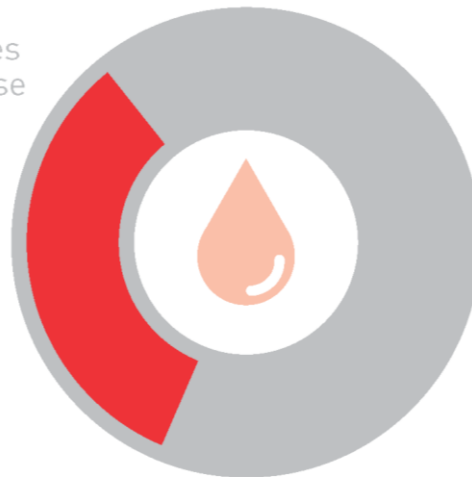
1 in 3 is overweight



1 in 10 is obese

3.7 MILLION
deaths due to diabetes
and high blood glucose

1.5 MILLION
deaths caused
by diabetes



Consequences

Diabetes can lead to complications in many parts of the body and increase the risk of dying prematurely.

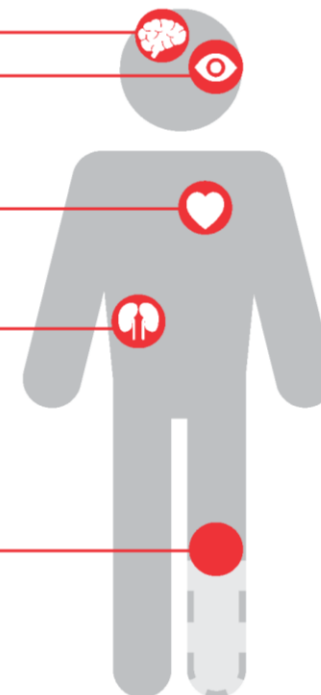
Stroke

Blindness

Heart attack

Kidney failure

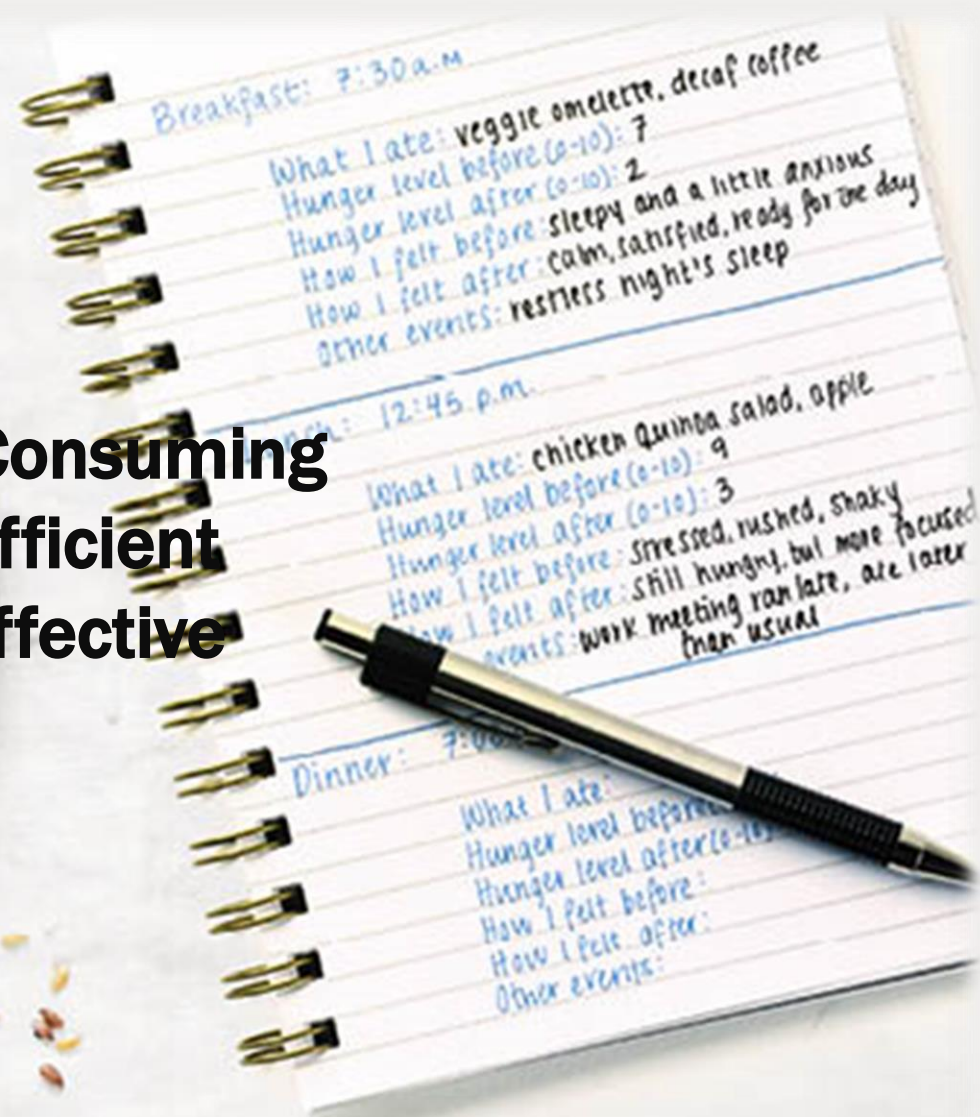
Amputation



FOOD JOURNALLING



Time Consuming
Inefficient
Ineffective



Connectify.ai





Health score

This week
13-19 September

100

Meal story

Last week's
data
+56



Weekly diary

Top 10 food 3 of 10



Add meal



Connectify.ai Goals

THIS PROJECT

Goal

1

Mobile App

User manually enters what they eat into the app after each meal



THIS PROJECT

Goal

2

Food Classification

Instead of entering food manually, user takes a picture of the food and the model identifies the food



Goal

3

Nutrition Data Retrieval

User is then given the nutrition facts (Calorie, Carb, Fat, Protein) about the food identified in the image



Goal

4

Personalised Meal Recommendation

User is provided with personalized healthy meal recommendations



DATASET

Food Images

Food Images

Dataset	# Total Images	# Images per class	Source
Training Set	~26000	~900	Food 101 (Kaggle)
Validation Set	~2900	~100	Food 101 (Kaggle)
Testing Set	580	20	Web Scraping (Google/Bing)

29

Food Classes

fried_calamari



fresh_spring_roll



pho



omelette



bibimbap



steamed_mussels



Pad_thai



samosa



Carrot_cake



Fish and chips



Carrot_cake



bibimbap



Fish and chips



Carrot_cake



Steamed_mussels



Image Label Verification

samosa



Chicken curry



ramen



Caesar salad



Miso soup



Spring_rolls



hummus



pho



Pad thai



ramen



Peking_duck



Spring_rolls



Miso_soup



Peking_duck



Carrot_cake



Chicken curry



Carrot_cake



The dataset contains image which are irrelevant or ambiguous.
These images were manually reviewed and removed accordingly.

Image Data Augmentation

Step

1

Random Flip

Images are randomly
flipped horizontally



Step

2

Random Rotation

Images are randomly
rotated clockwise/
anti-clockwise



Step

3

Random Translation

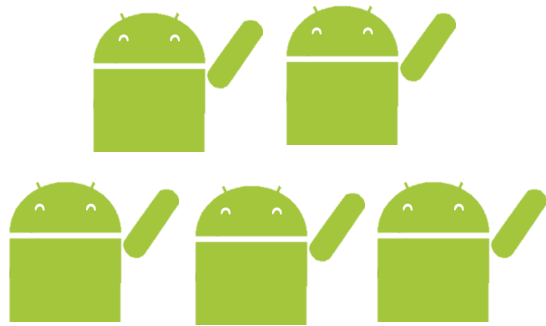
Images are randomly
shifted left/right



Why?

More Data

Model benefits from learning from a larger dataset



More Robust

Model becomes more robust to the real life images taken at different angles



Augmented Image Examples



After Image Data Augmentation...

Dataset	# Total Images	# Images per class	Source
Training Set	~26000	~900	Food 101 (Kaggle)
Validation Set	~2900	~100	Food 101 (Kaggle)
Testing Set	580	20	Web Scraping (Google/Bing)

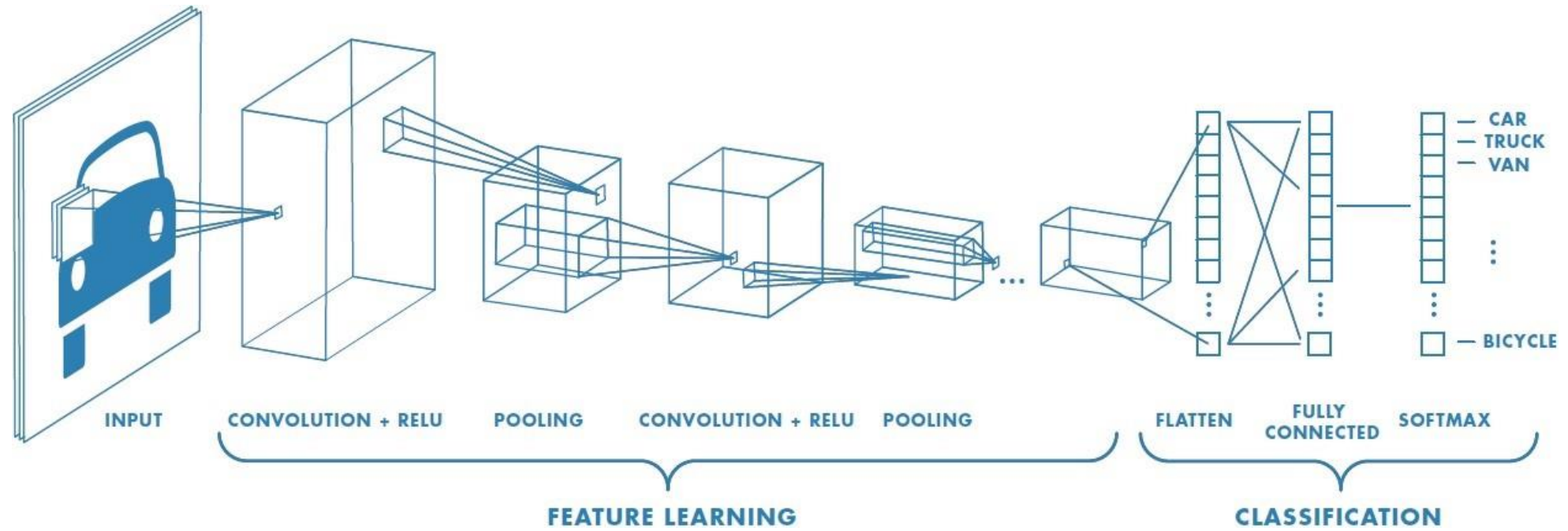


Removing irrelevant Images &
Performing Image Data Augmentation (two augmentations per image)

Dataset	# Total Images	# Images per class	Source
Training Set	~76000	~2600	Food 101 (Kaggle)
Validation Set	~8700	~300	Food 101 (Kaggle)
Testing Set	580	20	Web Scraping (Google/Bing)

Modelling

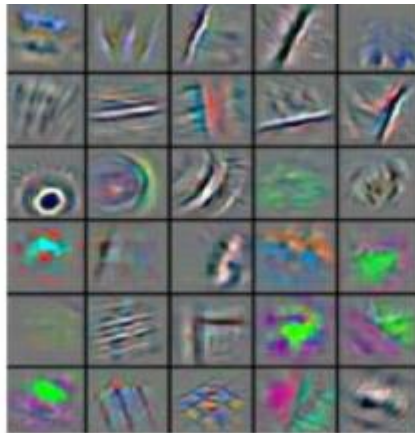
How does CNN work?



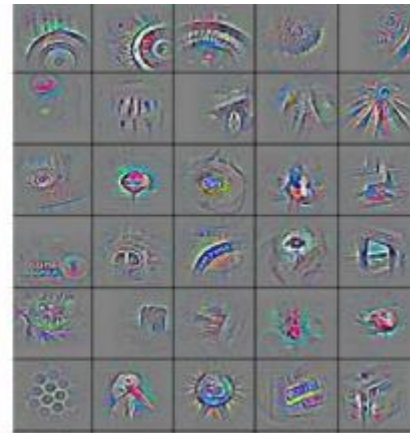
Level 1 Convolution



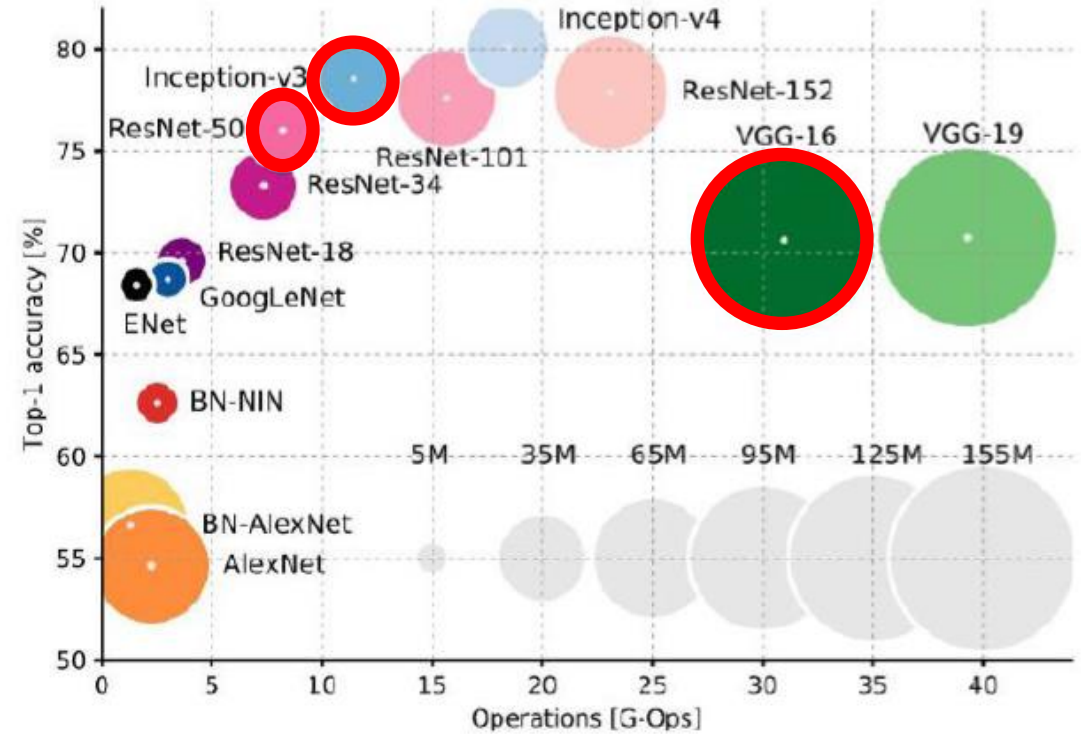
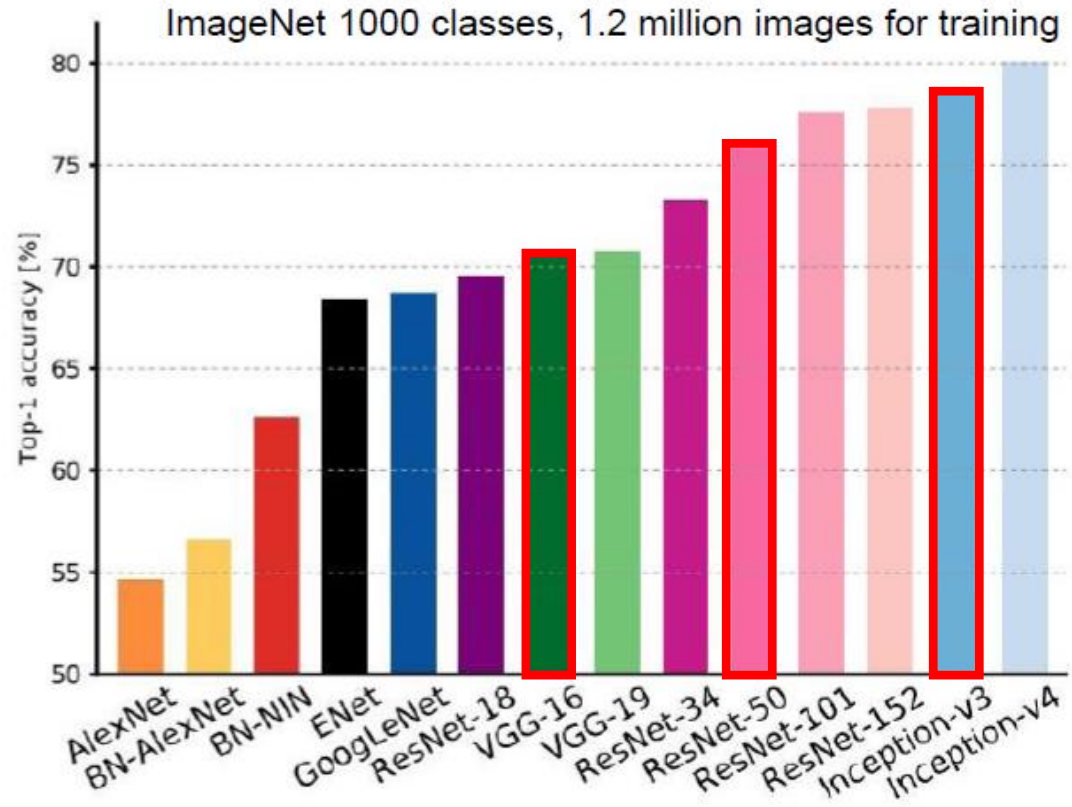
Level 2 Convolution



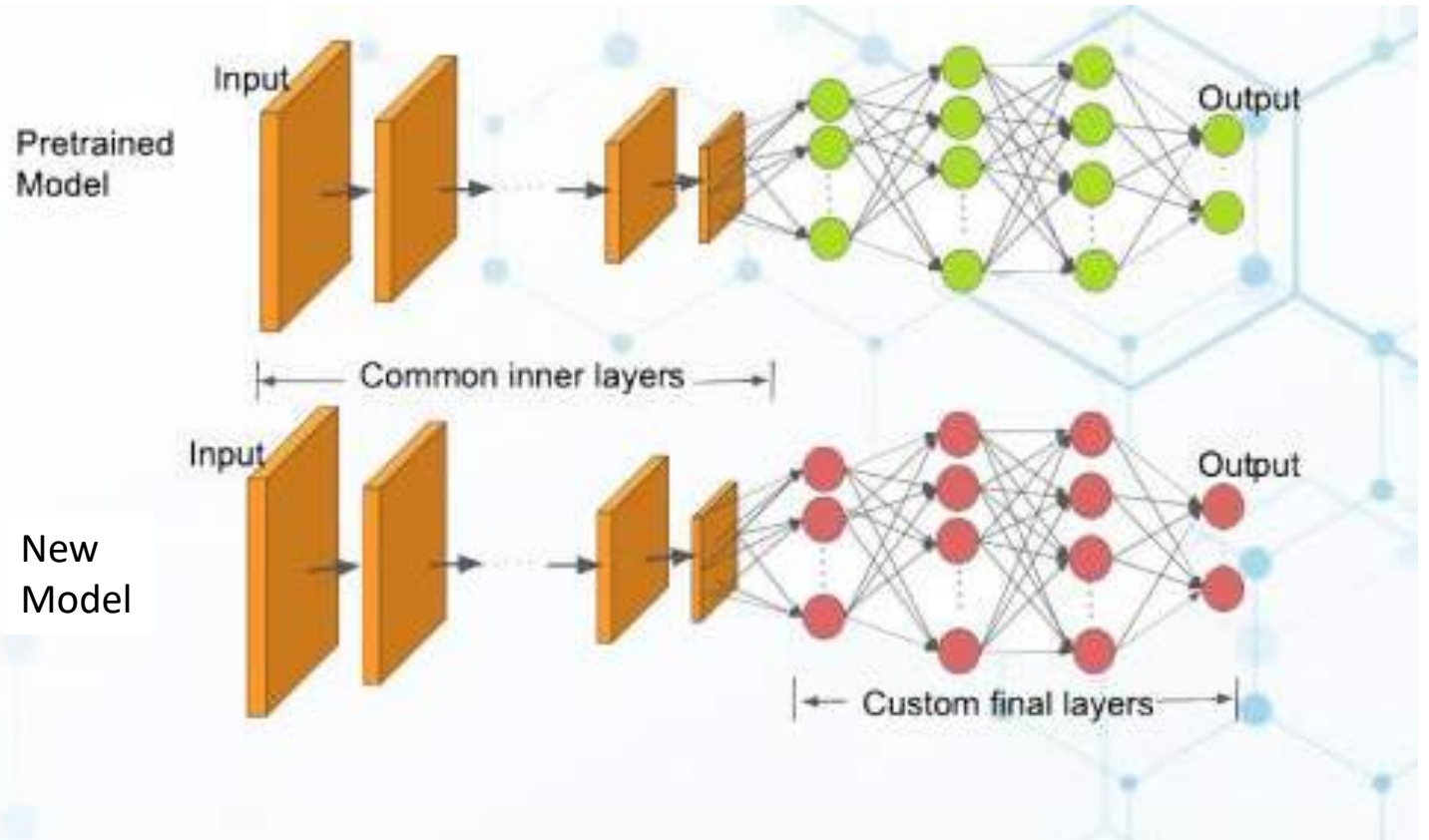
Level 3 Convolution



State of the Art CNN Models



Transfer Learning



Models Setups

CNN Models

- 1 x Custom Model (From Scratch)
- 2 x Inception V3 (Transfer Learning)
- 1 x ResNet 50 (Transfer Learning)
- 2 x Inception-ResNetV2 (Transfer Learning)
- 2 x VGG16 (Transfer Learning)

Configurations:

- Optimizer: Adam
- Regularization:
 - Dropout
 - Early Stopping (pat = 3)
- Metric: Accuracy

Model Specs

Model	# Layers	# Total Params	# Trainable Params
Custom	15	~5m	~5m
VGG16	16	~134m	~120k
VGG16 Dropout	15	~15m	~555k
InceptionV3 Dropout	49	~24m	~2m
InceptionV3 GAP	48	~22m	~60k
ResNet50	50	~24m	~60k
Inception-ResNetV2	164	~54m	~44k
Inception-ResNetV2 Dropout	165	~56m	~1.6m

Image Preprocessing

Preprocessing Techniques

- **InceptionV3**: Normalize the pixel values between -1 and 1
- **VGG16**: Each color channel is zero-centered with respect to the ImageNet dataset, without scaling.

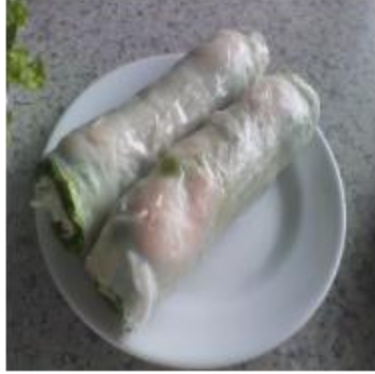


Original Images

fried_calamari



fresh_spring_roll



pho



omelette



bibimbap



steamed_mussels



pad_thai



samosa



carrot_cake



fish_and_chips



carrot_cake



bibimbap



fish_and_chips



carrot_cake

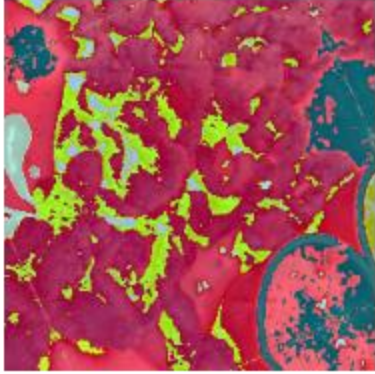


steamed_mussels

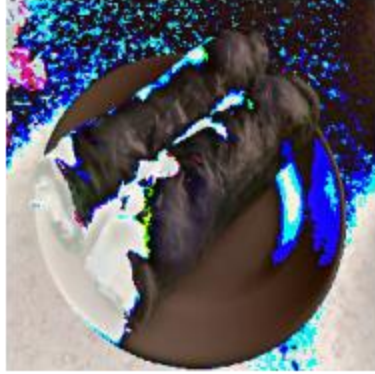


VGG16 Preprocessed Images

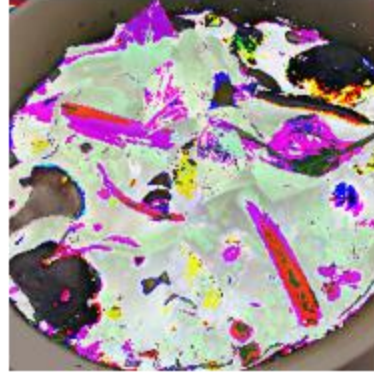
fried_calamari



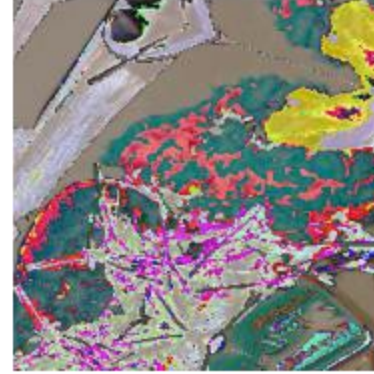
fresh_spring_roll



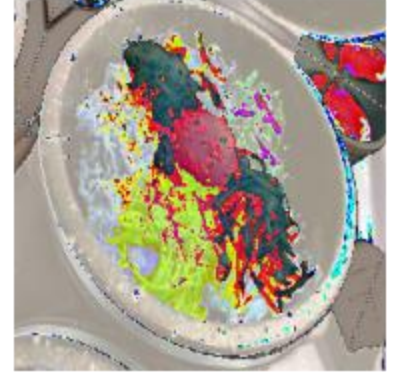
pho



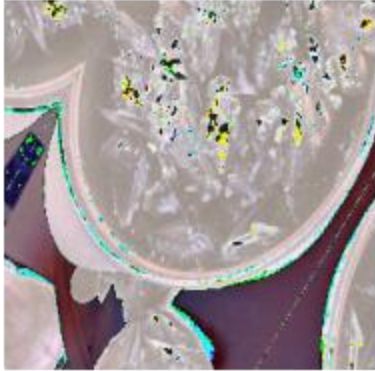
omelette



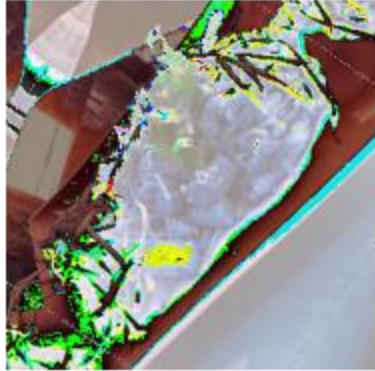
bibimbap



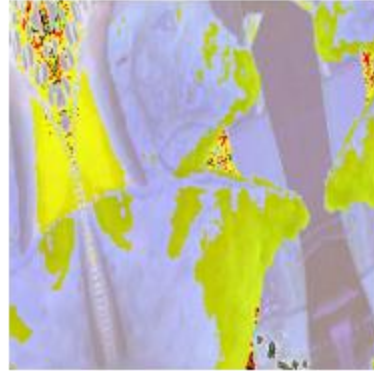
steamed_mussels



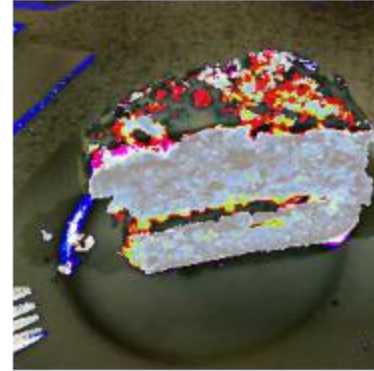
pad_thai



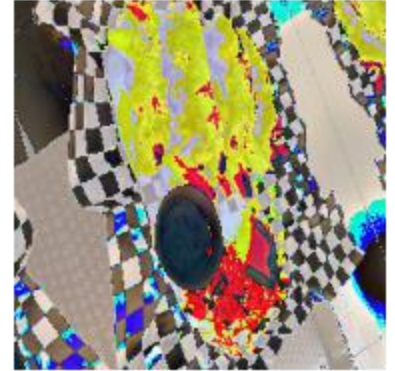
samosa



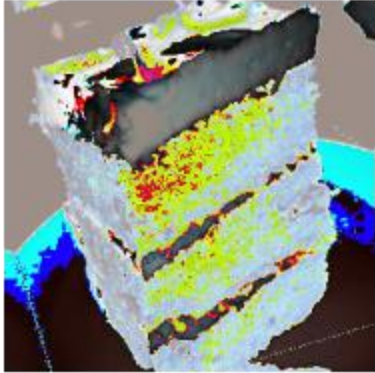
carrot_cake



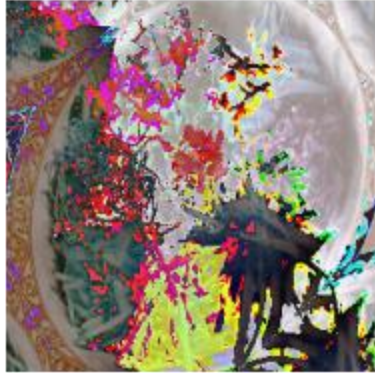
fish_and_chips



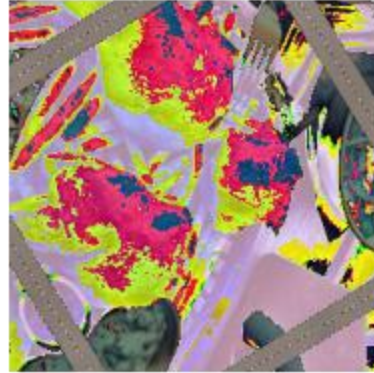
carrot_cake



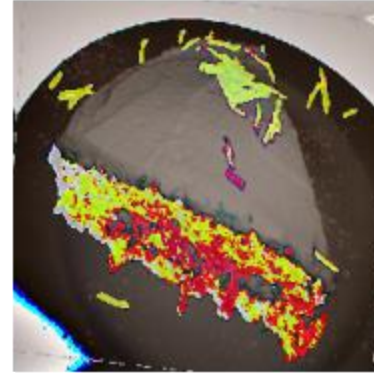
bibimbap



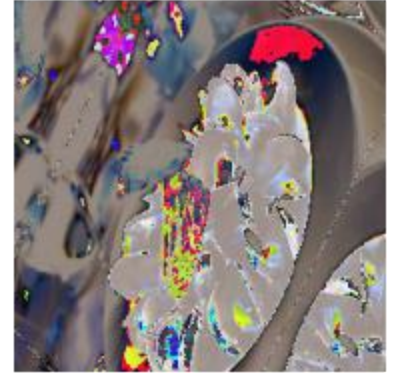
fish_and_chips



carrot_cake

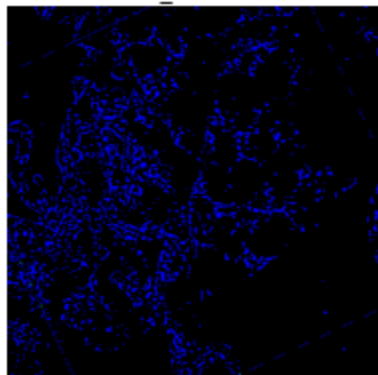


steamed_mussels



Inception V3 Preprocessed Images

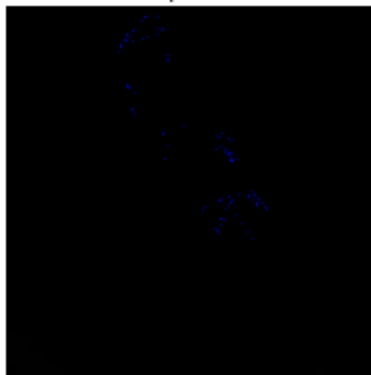
fried_calamari



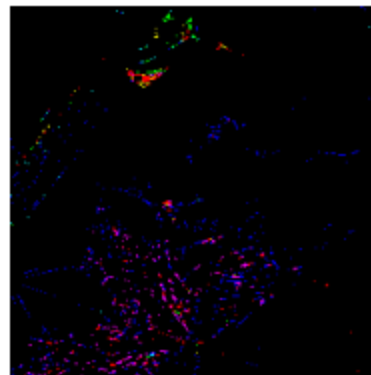
fresh_spring_roll



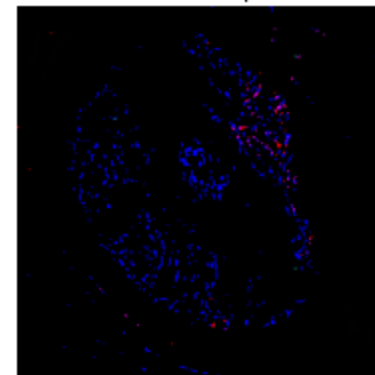
pho



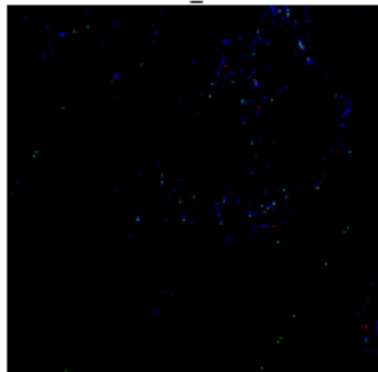
omelette



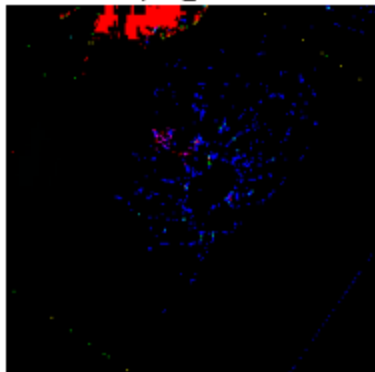
bibimbap



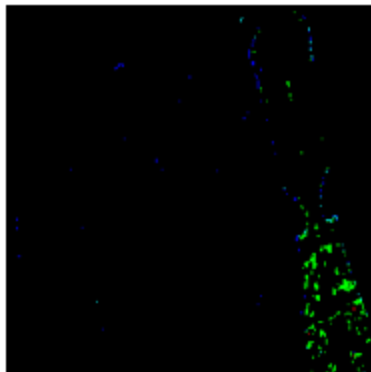
steamed_mussels



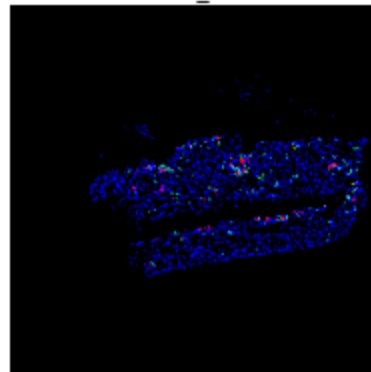
pad_thai



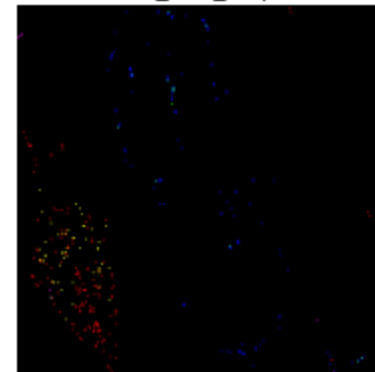
samosa



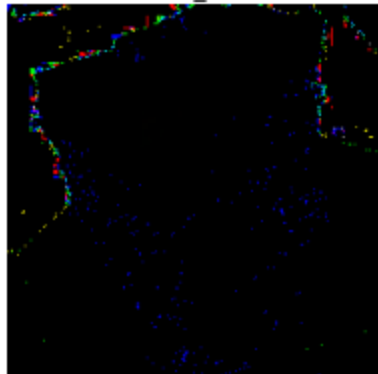
carrot_cake



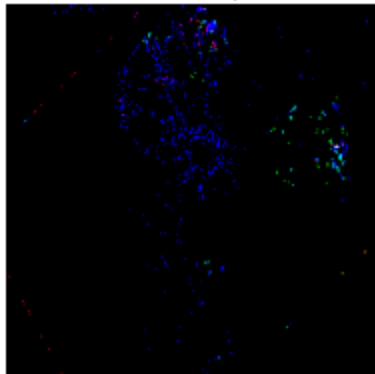
fish_and_chips



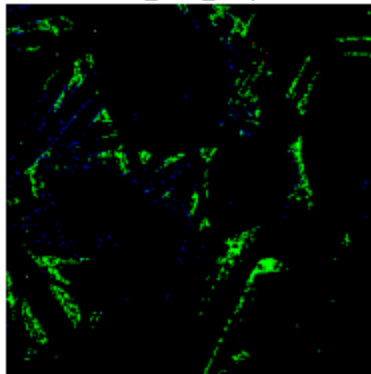
carrot_cake



bibimbap



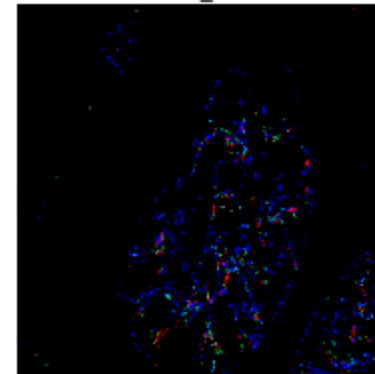
fish_and_chips



carrot_cake

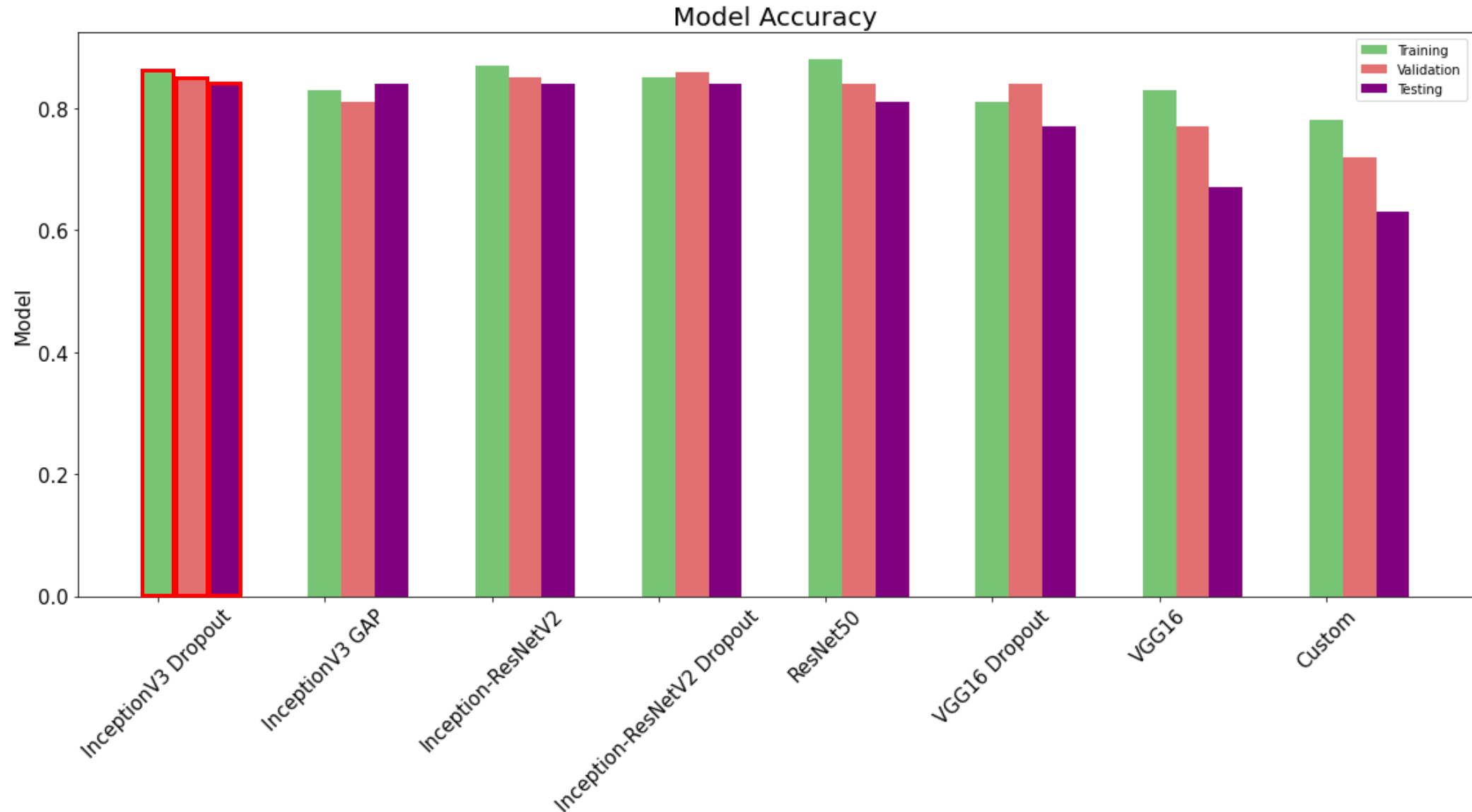


steamed_mussels



Evaluation

Model Benchmark



All models performed better than the **baseline accuracy** of 3.4%

Best Performing Model

Inception V3 Dropout

Description:

Transfer Learning on Inception V3 Model with weights from ImageNet dataset.

- Inception V3 image preprocessing (normalized between -1 and 1)
- Fully Connected Layers:
 - Dense layers (1024 filters)
 - Dropout layer (0.5)
 - Dense layer (29 filters)
- 49 layers and 2 million parameters to finetune.

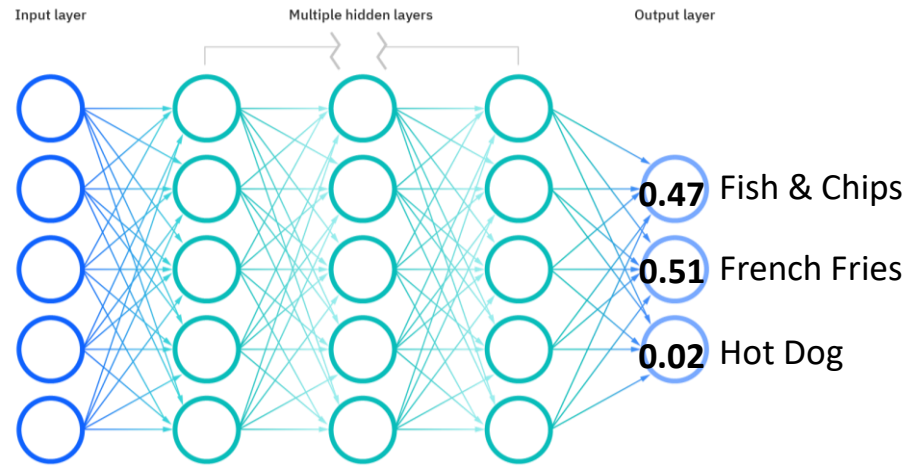
Top-N Accuracy Testing Set

Performance

Model	Top-1 Accuracy	Top-3 Accuracy	Top-5 Accuracy
Inception V3 Dropout	0.84	0.95	0.97



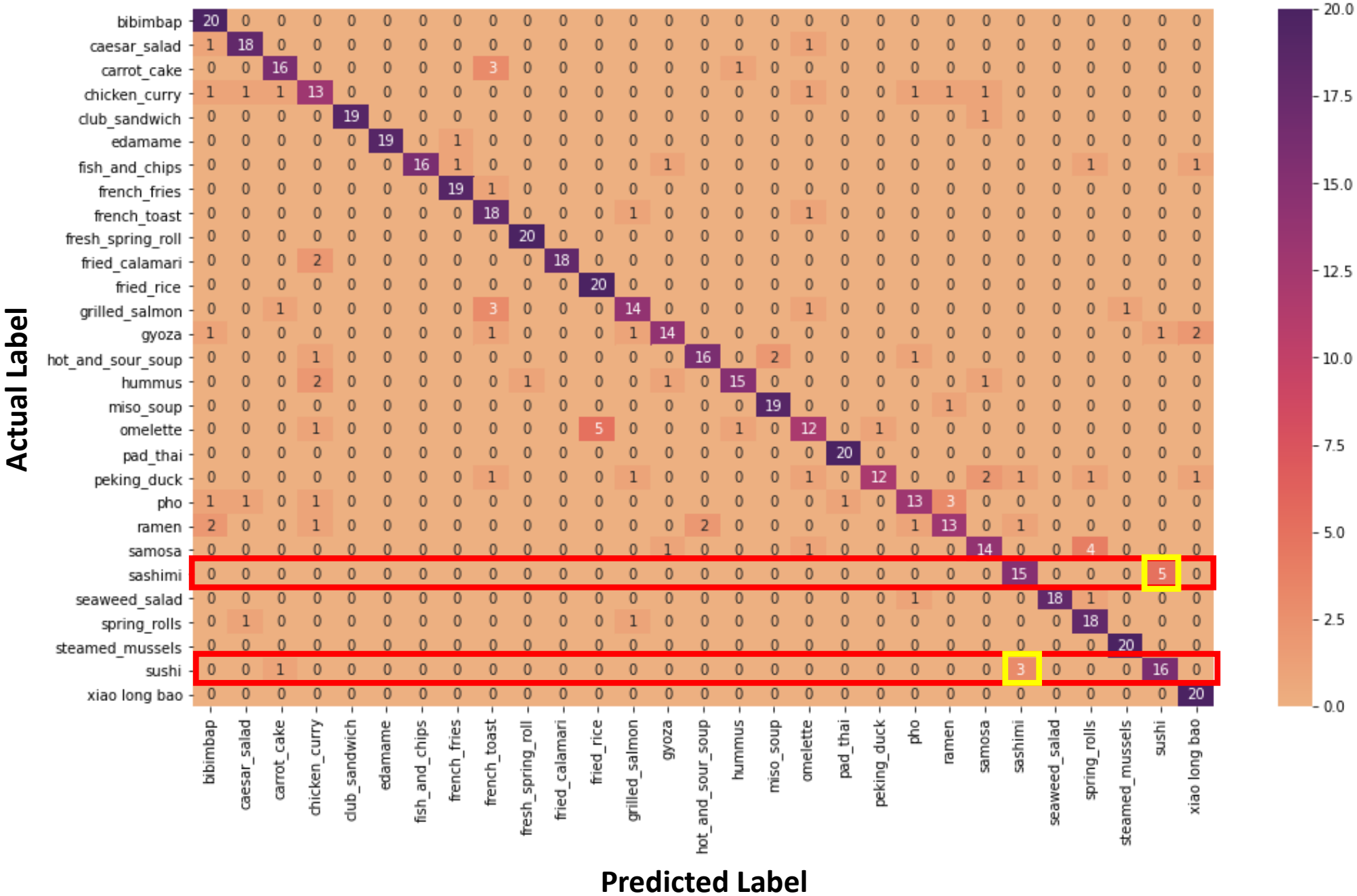
Fish & Chips



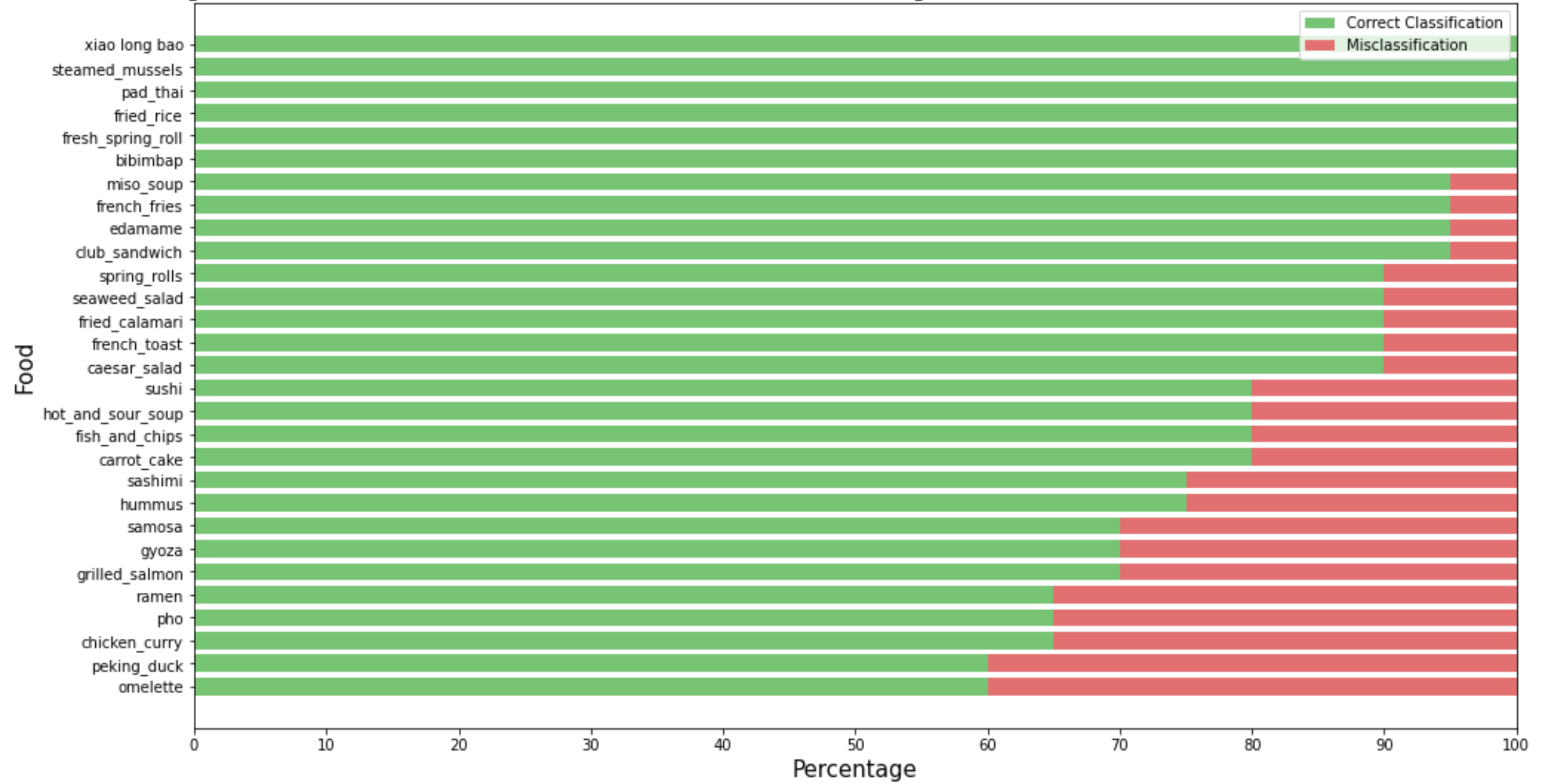
Top-1 Acc: ✗ Incorrect Prediction
Top-2 Acc: ✓ Correct Prediction

Web App Demo

Confusion Matrix



Accuracy and Misclassification Rate by Food Class



Misclassified Food Examples

True: miso_soup
Pred: ramen



True: omelette
Pred: fried_rice



True: omelette
Pred: chicken_curry



True: omelette
Pred: fried_rice



True: omelette
Pred: hummus



True: omelette
Pred: fried_rice



True: omelette
Pred: peking_duck



True: omelette
Pred: fried_rice



True: omelette
Pred: fried_rice



True: peking_duck
Pred: spring_rolls



True: peking_duck
Pred: samosa



True: peking_duck
Pred: xiao long bao



True: peking_duck
Pred: samosa



True: peking_duck
Pred: grilled_salmon



True: peking_duck
Pred: omelette



Limitations & Future Improvements

1

Multiple Food Types

Food Images can often contain multiple food classes
(eg. Fish&Chips vs French Fries)

Possible Solution

Assign class_weight when training to prioritise certain classes over others

2

Large intra-class diversity

Food Images belonging to same class might be diverse in how they look
(eg. Peking ducks – whole, sliced, duck rolls)

Possible Solution

Acquire more training data
and/or
Split classes that have large diversity

3

Large inter-class similarity

Food Images of different classes might look very similar
(eg. Sushi vs Sashimi)

Possible Solution

Acquire more training data
and/or
Group very similar classes together