





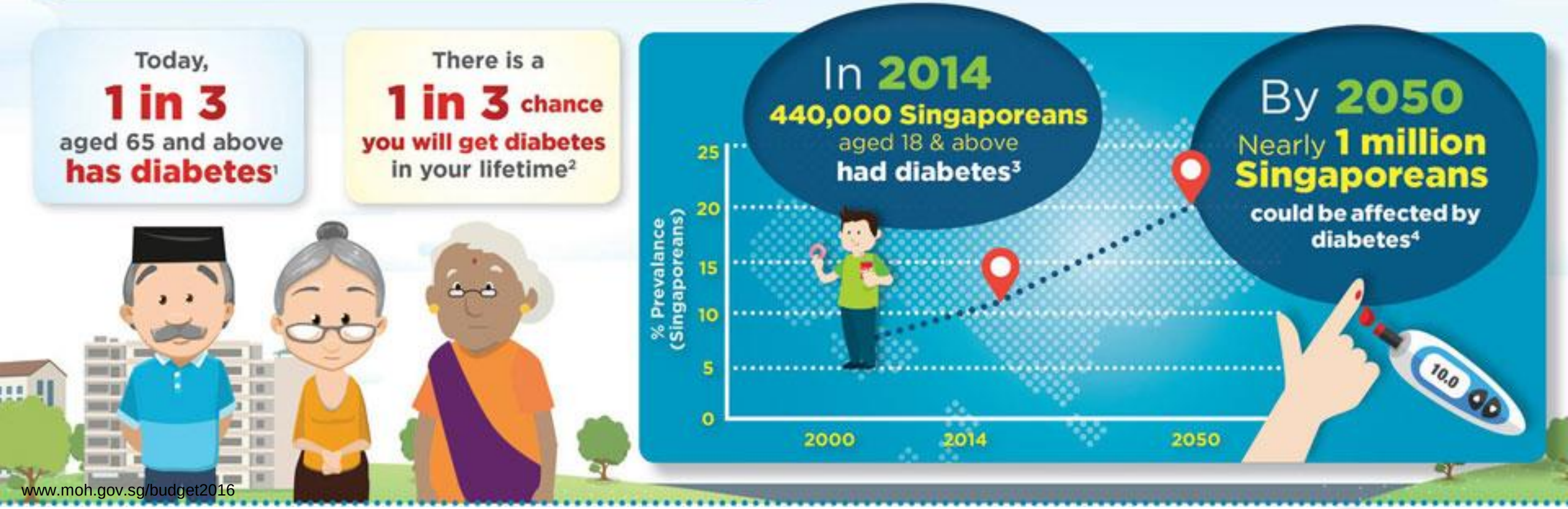
# Food Image Recognition by Deep Learning

Assoc. Prof. Steven HOI

School of Information Systems
Singapore Management University

# National Day Rally 2017: Singapore's War on Diabetes

MINISTRY OF HEALTH SINGAPORE




*"Four simple ways to fight diabetes: Go for regular medical check-ups; Exercise more; **Watch your diet; and Cut down on soft drinks**."*

- PM Lee Hsien Loong

# Traditional Food Journal


https://www.womenshealthmag.com/sites/womenshealthmag.com/files/images/food-journal-1_0.jpg

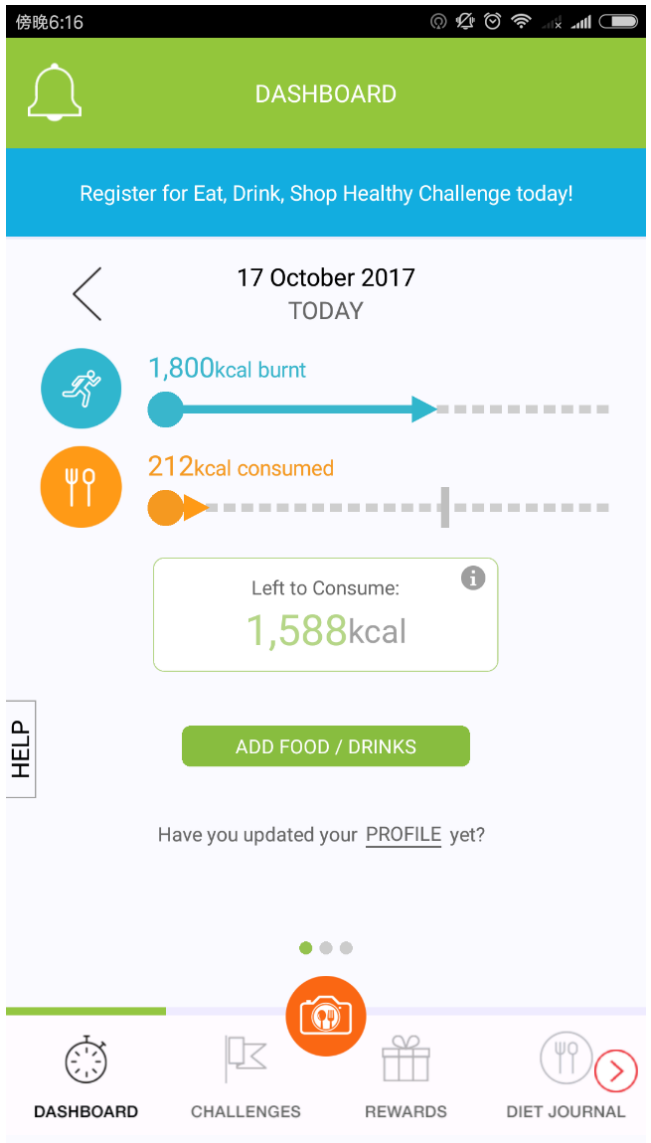
- ✖ Tedious
- ✖ Non-efficient
- ✖ Non-effective

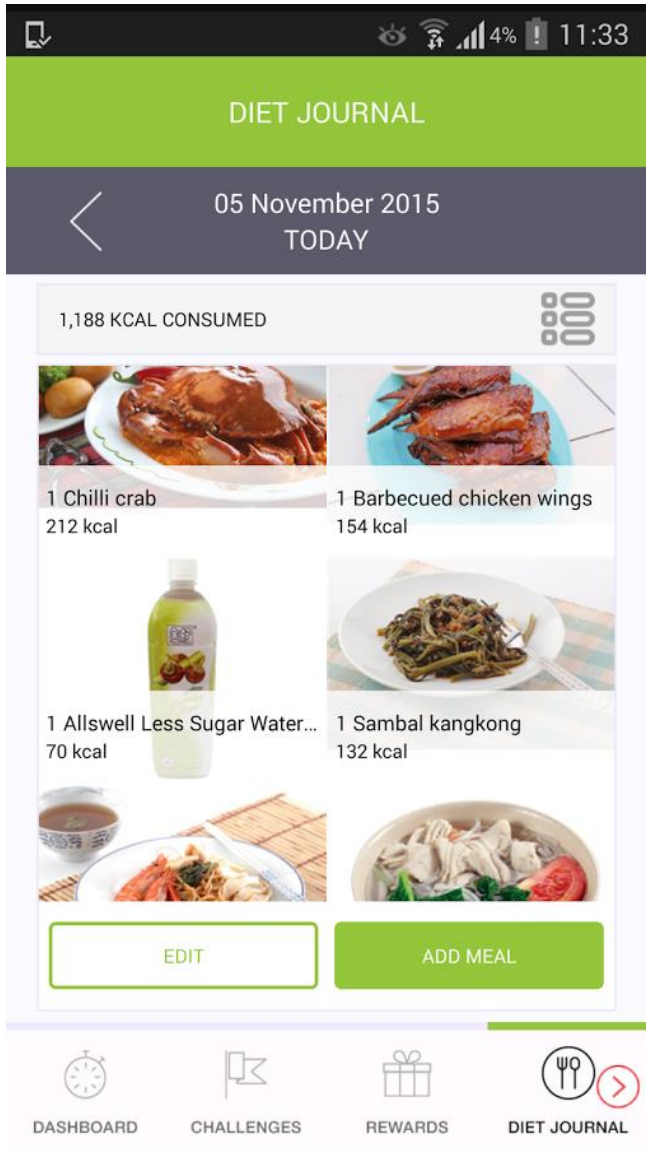
# Smart Food Logging

Health Promotion Board

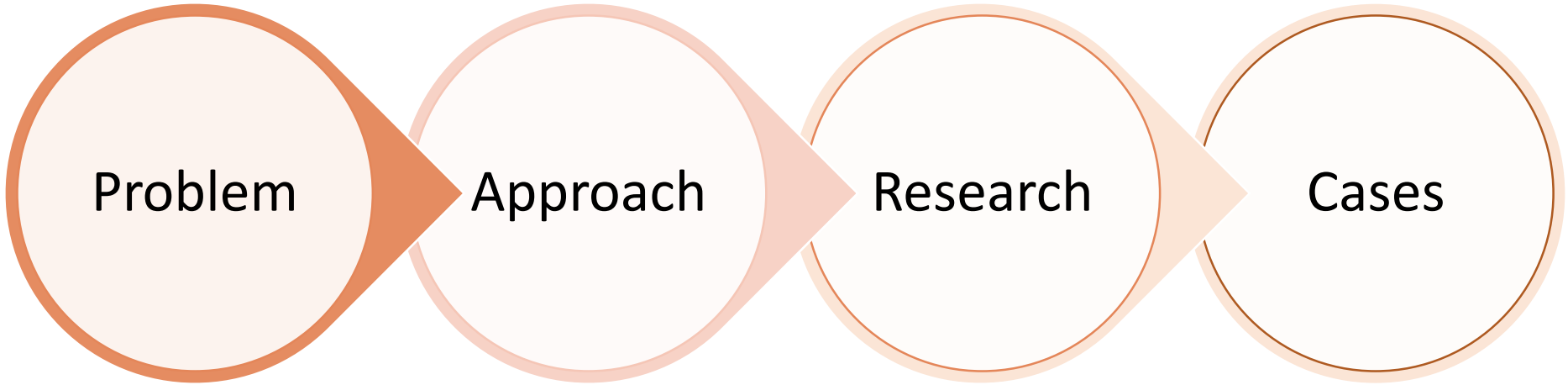
Healthy 365

Powered by FOODAi



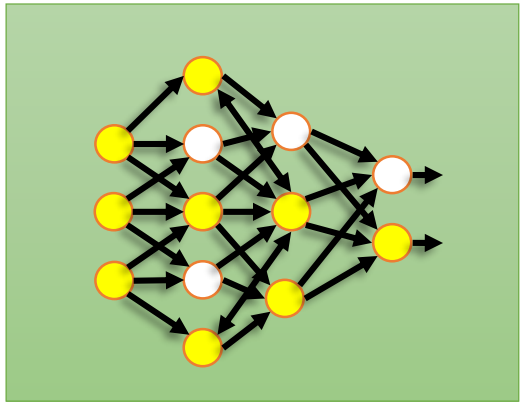



# Roadmap



# Food Image Recognition

- Visual Recognition



Machine Learning

Laksa?

# Food Image Recognition

- Could be very challenging…

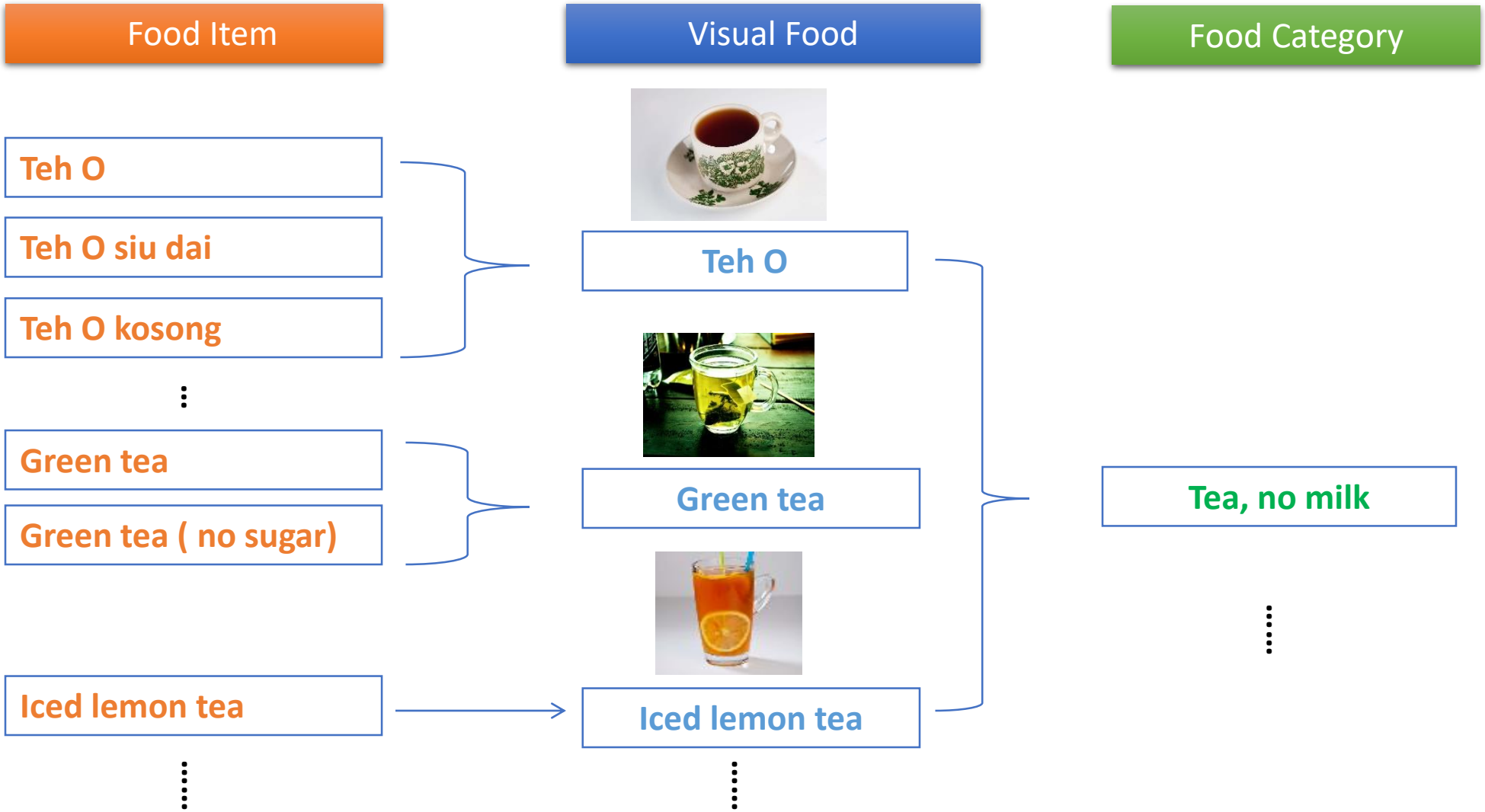
http://supermerlion.com/wp-content/uploads/2010/04/madnesskopiteh.jpg

**Singapore Tea or Teh**

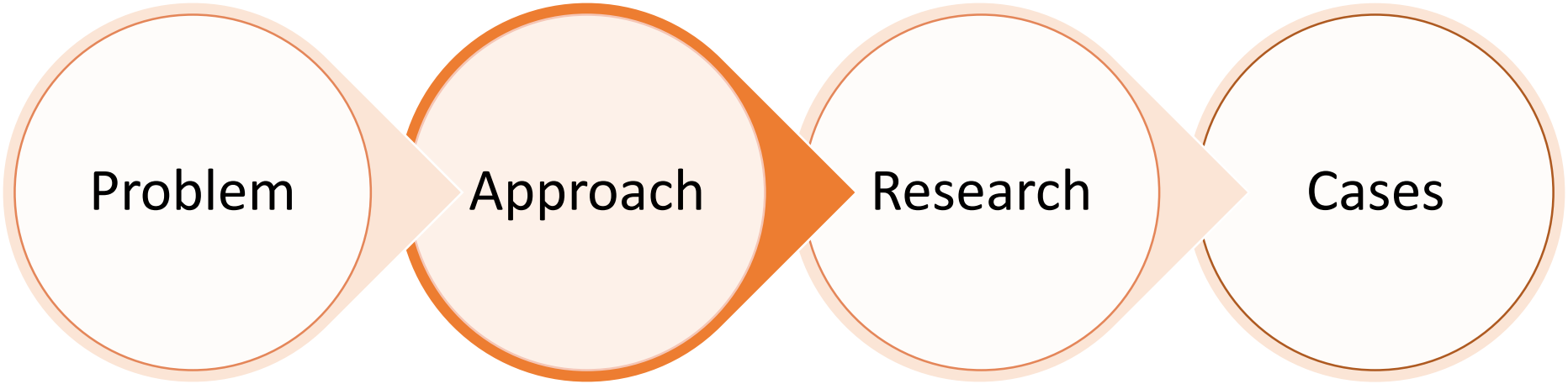
- *Teh*, tea with milk and sugar
- *Teh-C*, tea with evaporated milk
- *Teh-C-kosong*, tea with evaporated milk and no sugar
- *Teh-O*, tea with sugar only
- *Teh-O-kosong*, plain tea without milk or sugar
- *Teh tarik*, the Malay tea
- *Teh-halia*, tea with ginger water
- *Teh-bing*, tea with ice, aka *Teh-ice*
- *Teh-siu-dai*, tea with less sugar
- *Teh-gah-dai*, tea with extra sweetened milk

……

# Food Name Hierarchy

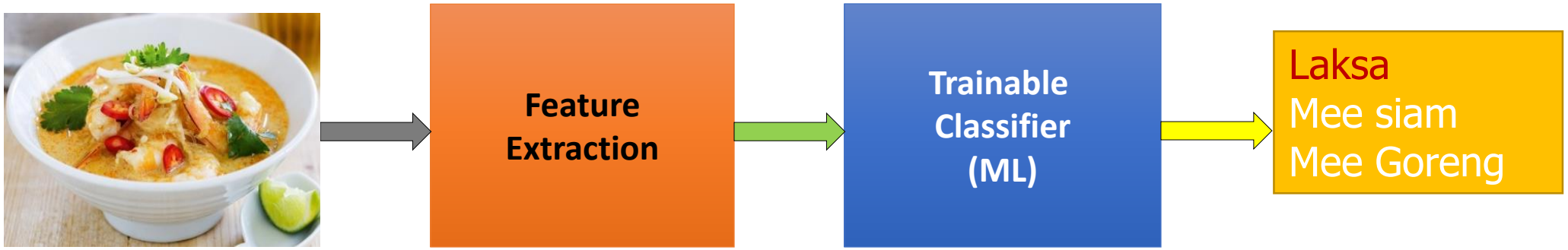


# Roadmap



# Visual Recognition

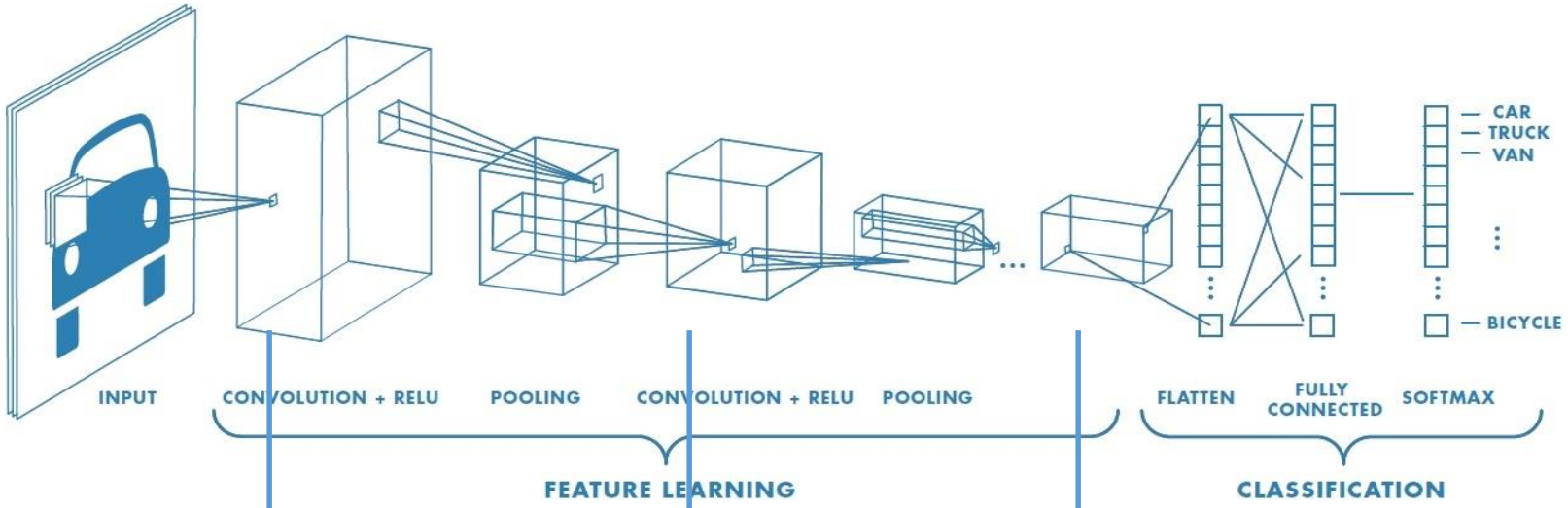
- Classical Computer Vision Pipeline




- Deep Learning Approach

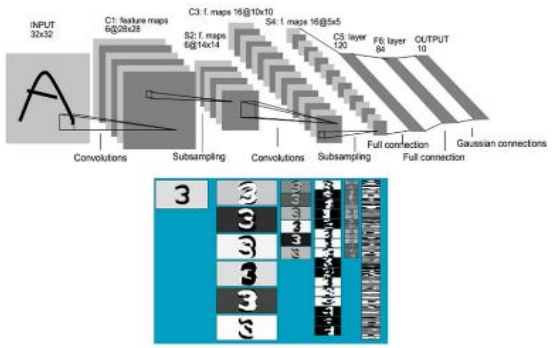
# Deep Convolutional Neural Networks (CNN)

- Convolutional Neural Networks (CNN)





Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]



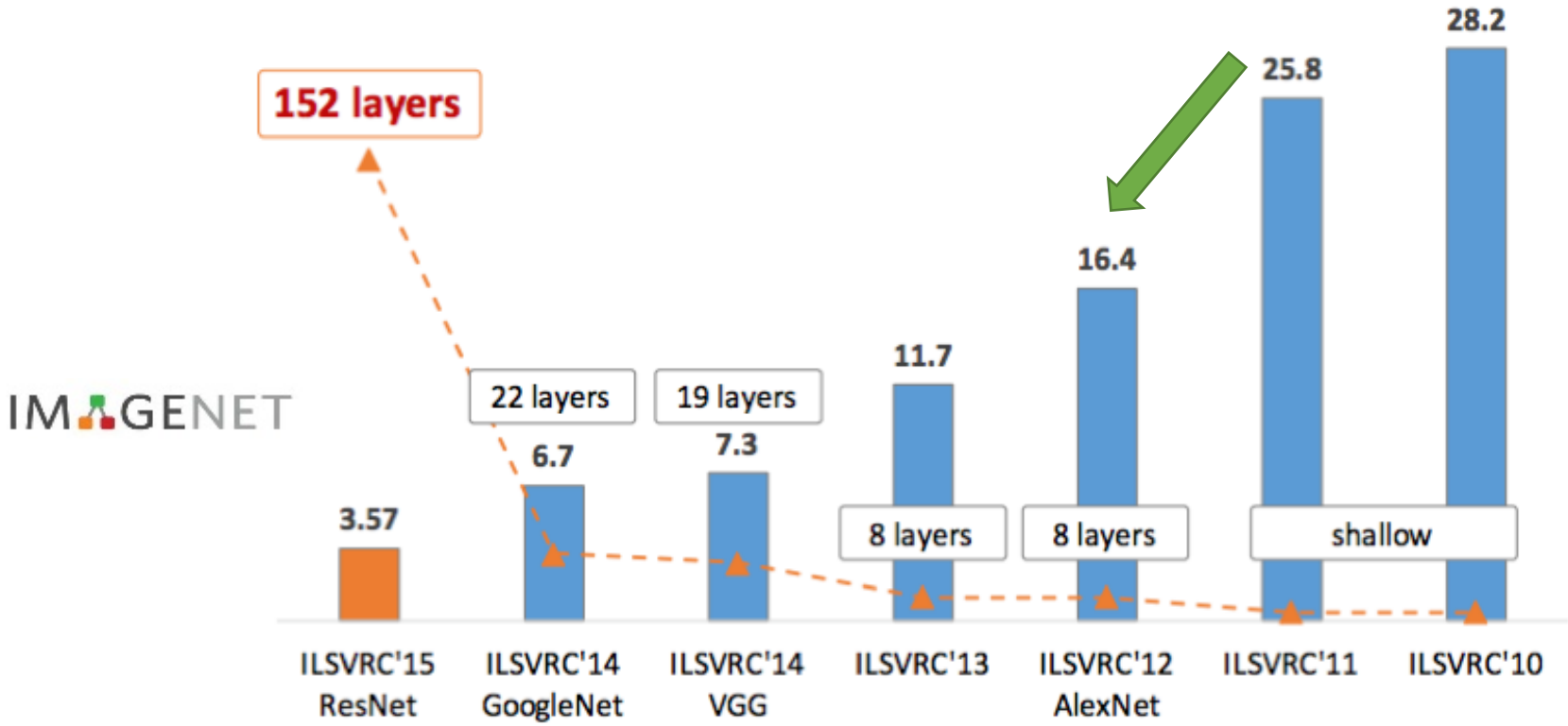
LeNet [LeCun et a. 1998]

Photos taken form https://www.mathworks.com/discovery/convolutional-neural-network.html

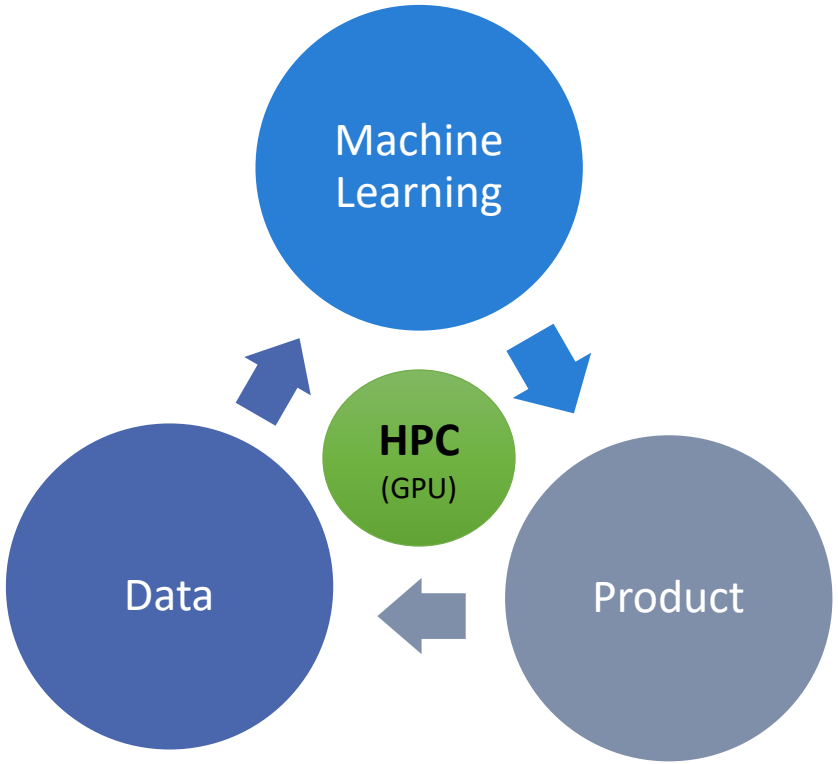
# Deep CNN for Visual Recognition

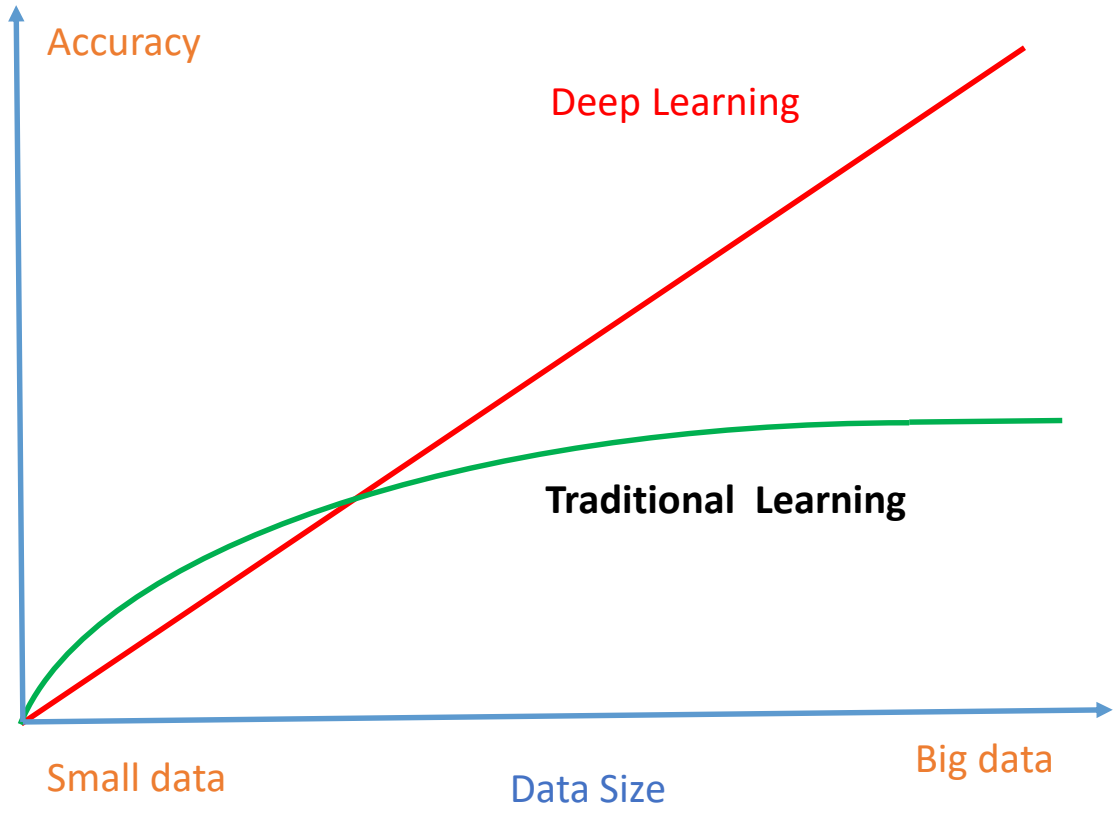
- Revolution of Depth
  - From AlexNet (8-layers) in 2012

[ Krizhevsky et al. 2012 ]



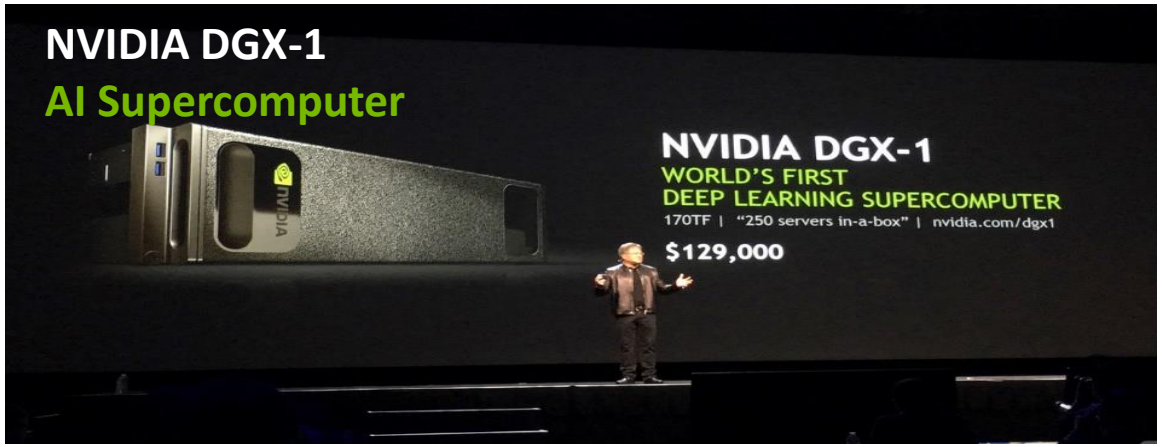
# Why Deep Learning?

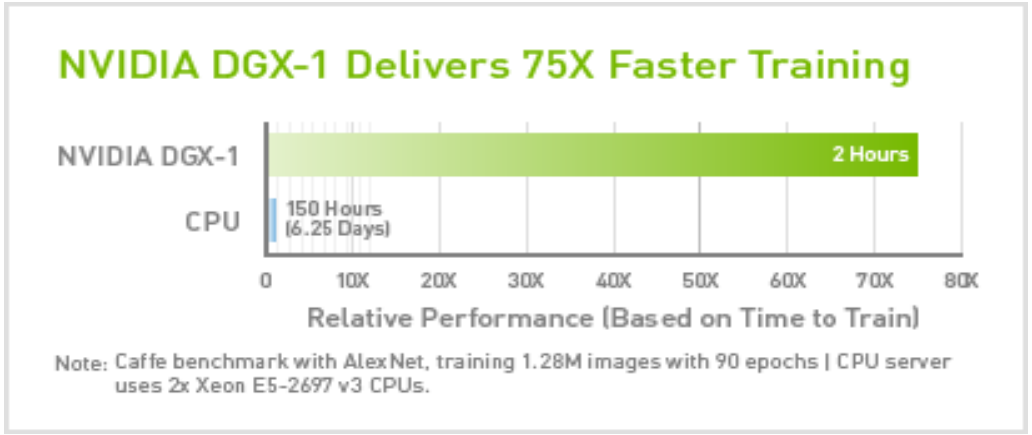


# GPU for High Performance Computing

- Deep Learning on GPU Clusters
- DGX-1: NVIDIA Pascal™-powered Tesla® P100
- Performance equal to 250 conventional servers.




LARC Living Analytics Research Centre

Singapore 1st DGX-1 Deep Learning Supercomputer (with P100 GPUs)


SG-FOOD

# SGFOOD Data Statistics

| SGFood724 Dataset | Training | Validation | Test |
|---|---|---|---|
| # total images | 361,676 | 7,240 | 36,200 |
| # Image per class | ~500 | 10 | 50 |

#Food Items: 1038 **#Visual Food: 724** #Food Category: 158


Histogram of #visual foods (724 visual food classes)

# FoodAI: Open API Services

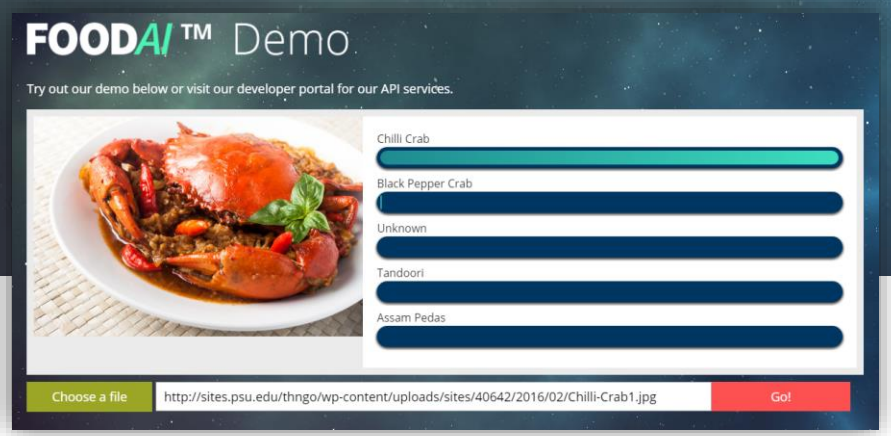
Home About Demo Contact Log in

http://www.foodai.org

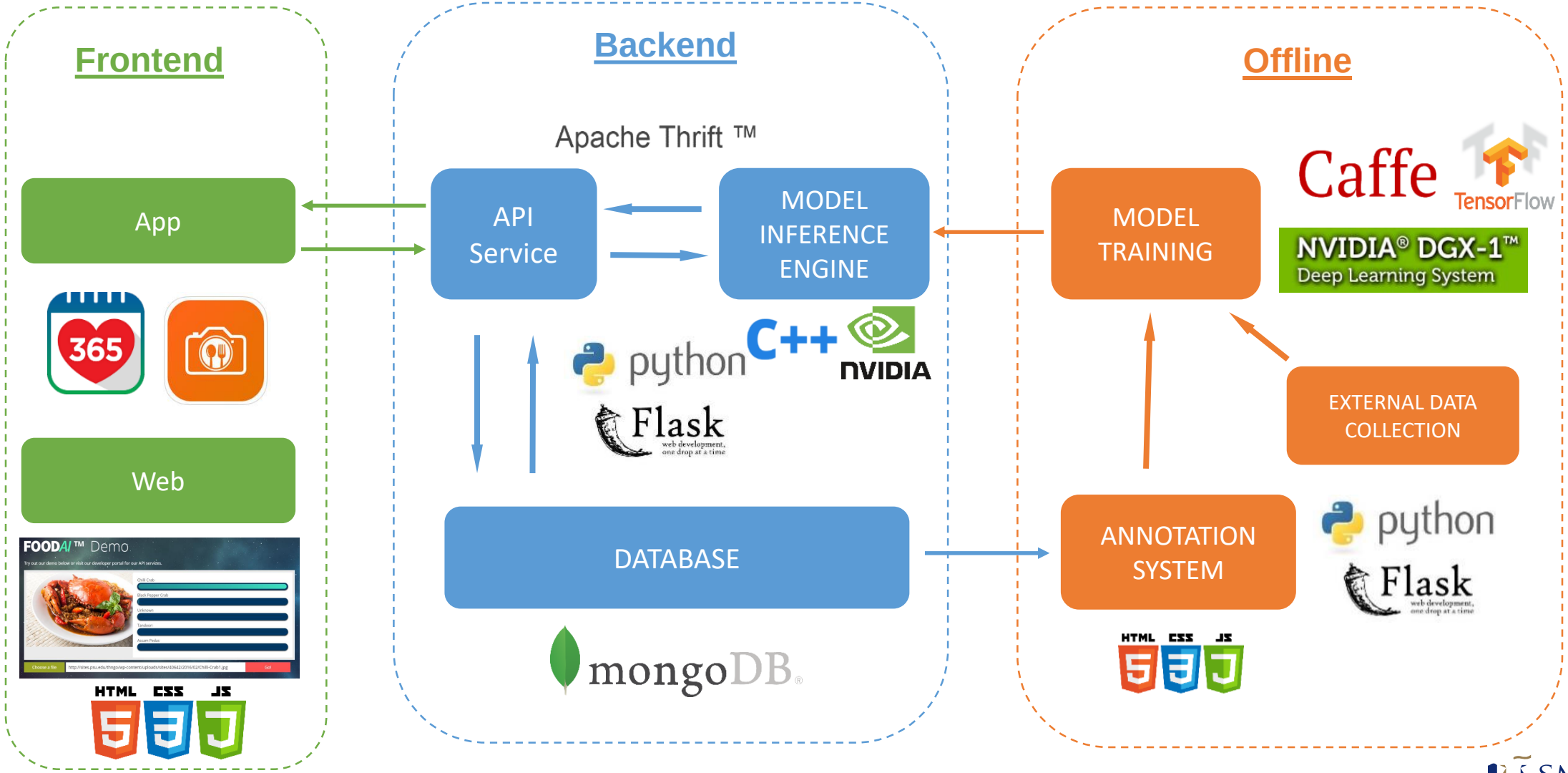
FOODAi

Smart Food Recognition with the state-of-the-art Visual Recognition technology

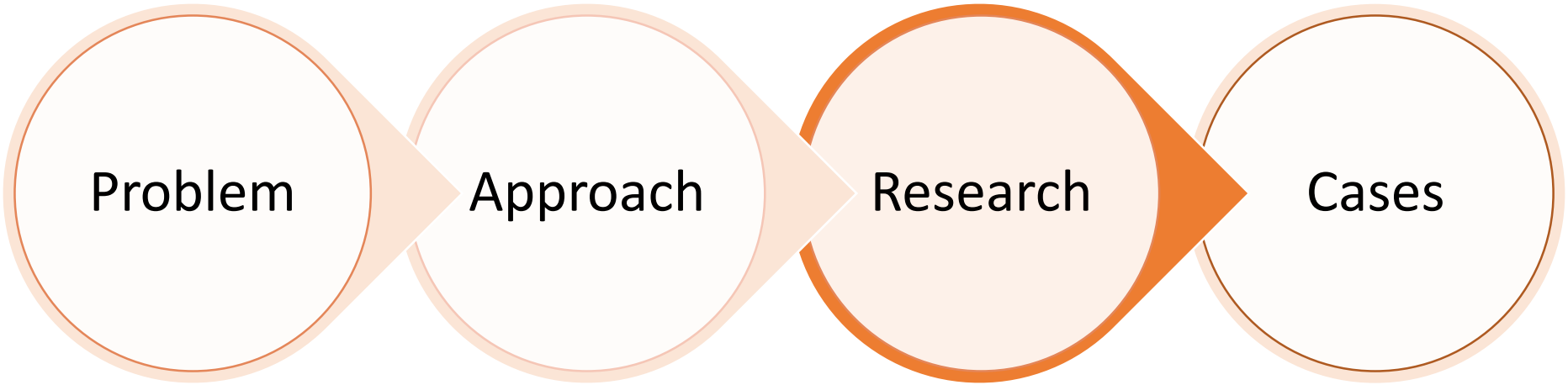
Try our Demo




# FoodAI System Architecture



# Roadmap

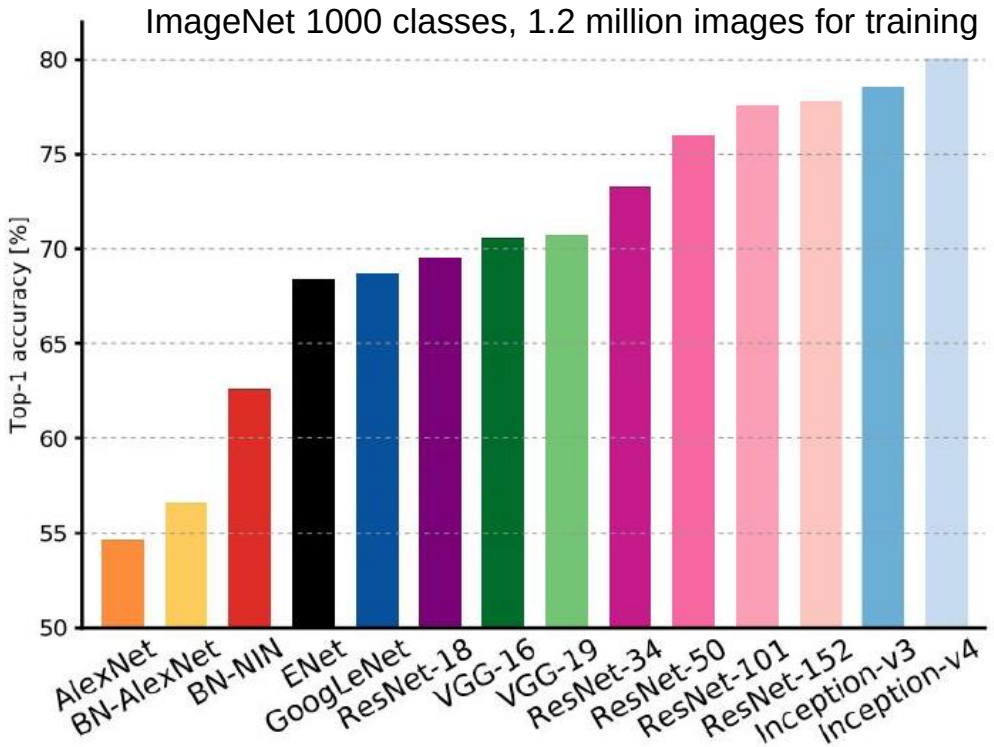


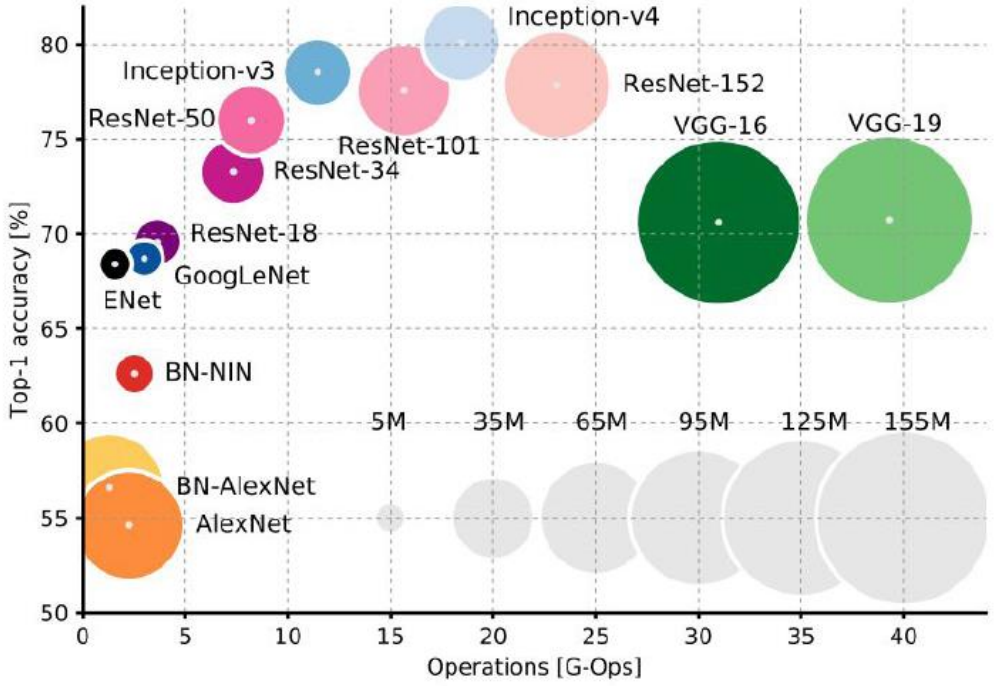
# Research Challenges

- How to train a good CNN model?
- How to deal with new food?
- How the labeled data size affects the accuracy?

# Model Training

- A Family of CNN models for visual recognition




"An Analysis of Deep Neural Network Models for Practical Applications"
Alfredo Canziani, Adam Paszke, Eugenio Culurciello Published 2016 in ArXiv

# Experimental Setups

- CNN Models
  - GoogleNet
  - ResNet: 18, 50, 101, 152
- Settings
  - Toolbox: Caffe & TensorFelow
  - Finetuned from ImageNet pretrained models
  - Batch Size: From 16 to 128
  - Optimizer: SGD with momentum/RMS Prop/Adam
  - Learning rate: Fixed/multi-step/exponential decay
  - Dropout/Batch Normalizations

# Benchmark of FoodAI

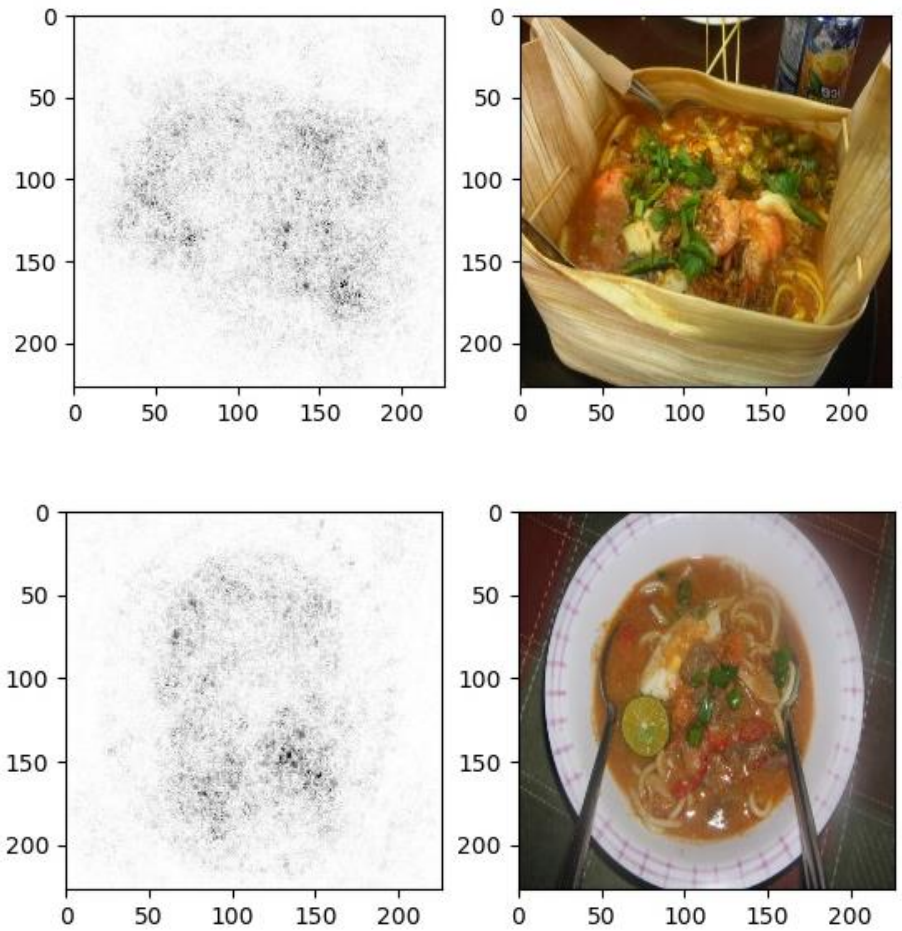
724 visual food classes, 361,676 images for training, ~500 images per class

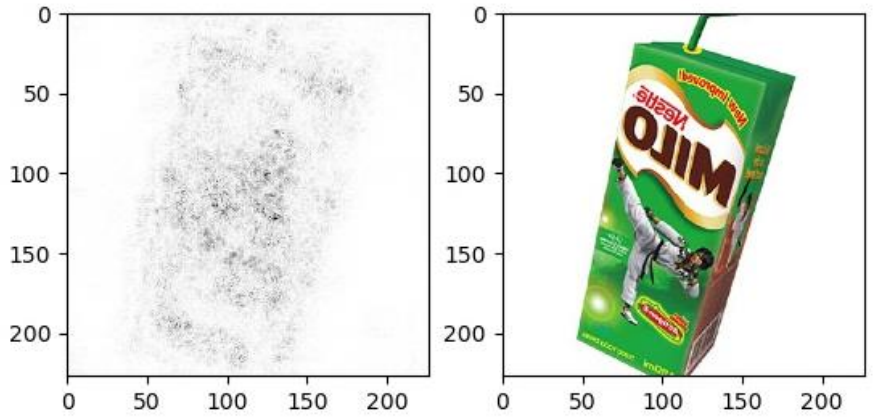
| Models (SGFOOD) | Top-1 Accuracy (%) | Top-5 Accuracy (%) |
|---|---|---|
| GoogleNet | 71.5 | 91.0 |
| ResNet-18 | 71.2 | 91.5 |
| ResNet-50 | 76.1 | 93.3 |
| ResNet-101 | 73.2 | 91.9 |
| ResNet-152 | 74.7 | 92.7 |

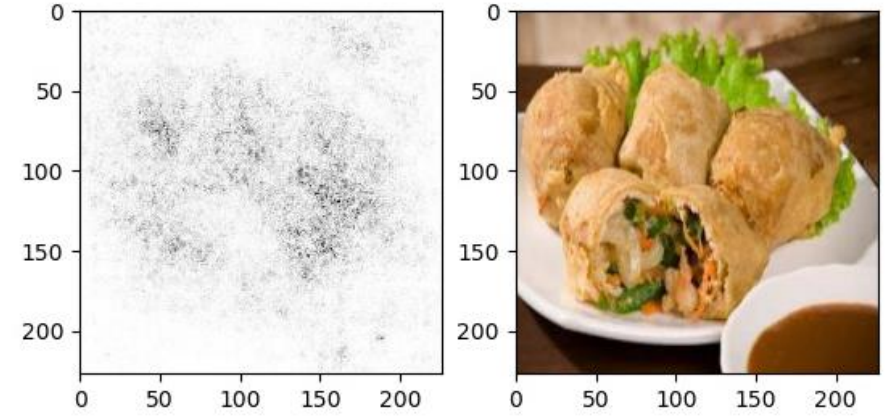
1000 object classes, 1.2 million images for training, 1200 images per class

| Models (IMAGENET) | Top-1 Accuracy (%) | Top-5 Accuracy (%) |
|---|---|---|
| ResNet-50 | 77.1 | 93.3 |
| ResNet-101 | 78.2 | 93.9 |
| ResNet-152 | 78.6 | 94.3 |

# Food Saliency Map

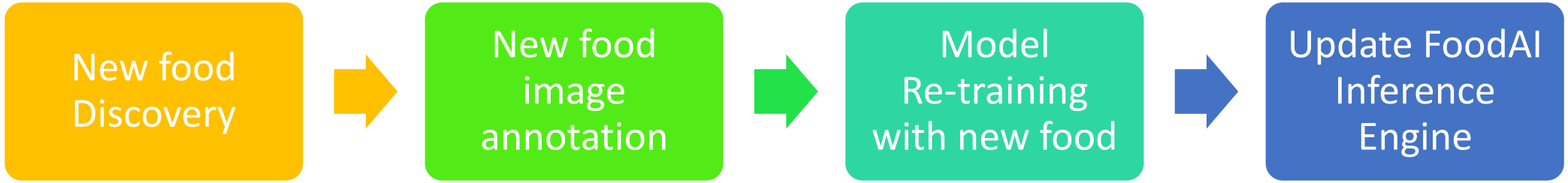


# How to handle NEW food?

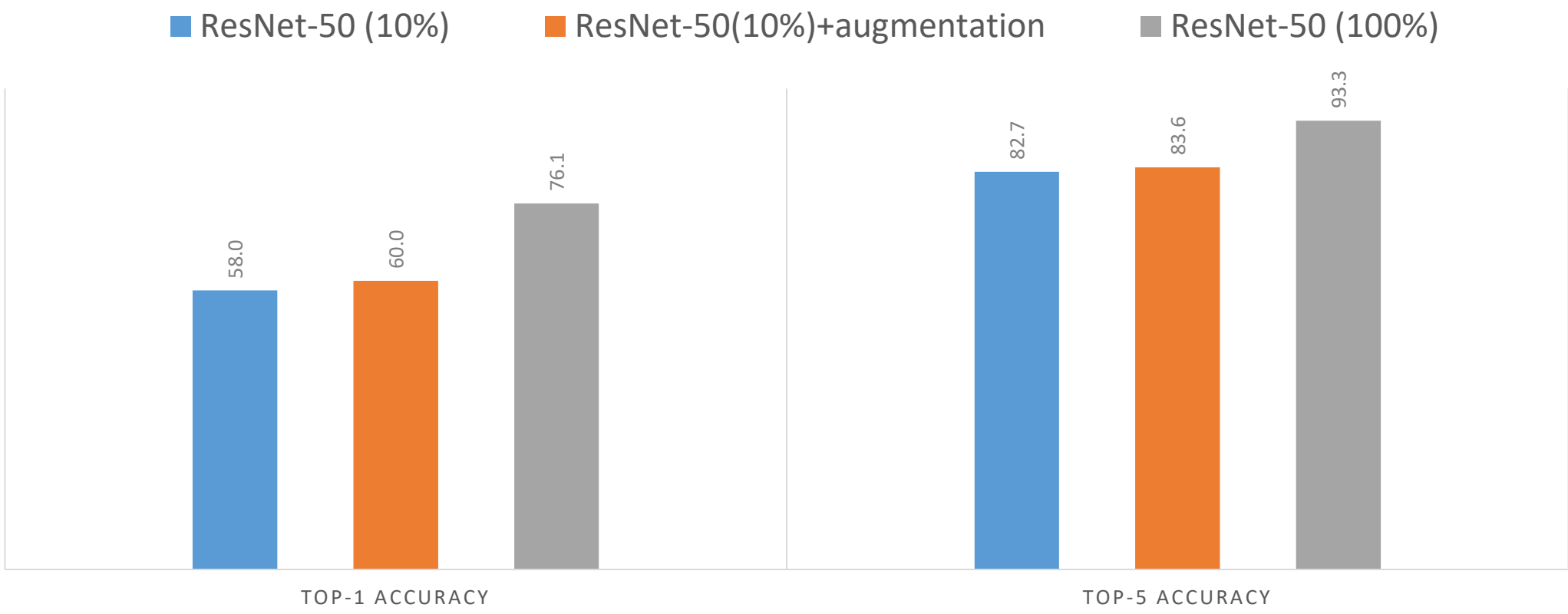
- Too many possible food items in the market
- Only consider popular food for majority of users



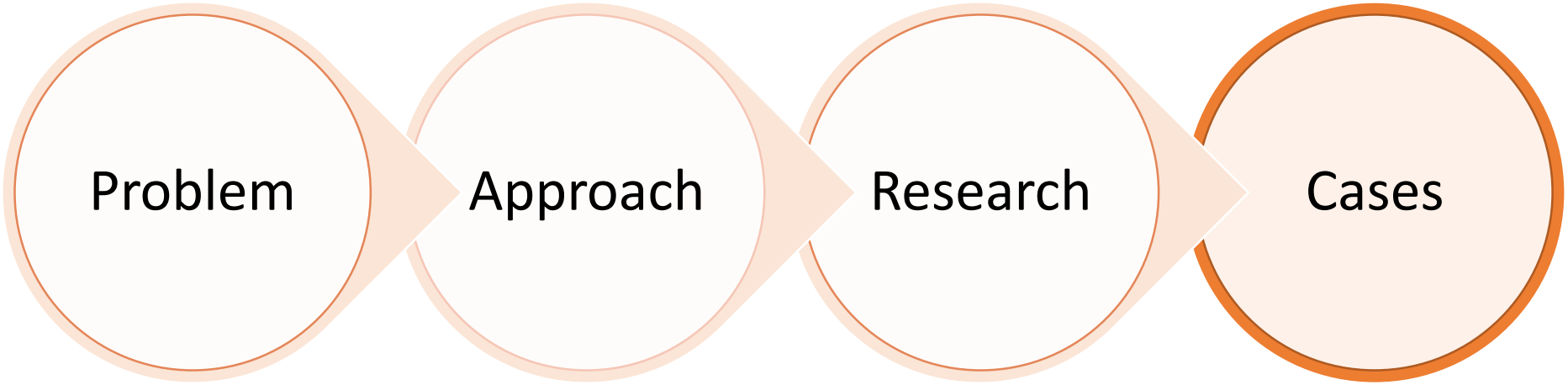

- New food has few images available at the beginning

What if only **10x less amount of labeled data** is available to train an CNN model?

# Training on 10x less labeled data



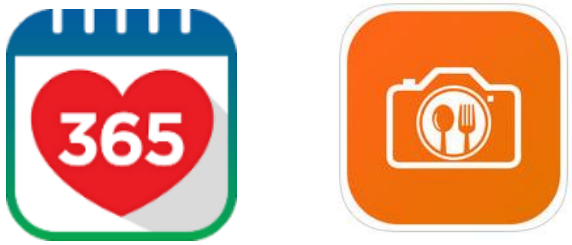
# Roadmap



# Case Studies: Food logging photos from users

Mobile App

Web




Powered by

FOODAI

# Case Studies: Easy Cases
















# Case Studies: Hard Cases

Large inter-class similarity (e.g., drinks)

## Kopi O






## Americano





# Case Studies: Hard Cases

## Large inter-class similarity (e.g., drinks)

### Instant Coffee



### Teh C / Teh


### Plain Porridge



### Soya milk


# Case Studies: Hard Cases

Large inter-class similarity (e.g., drinks)

Instant Coffee

Teh O

Teh / Teh C

# Case Studies: Hard Cases

Large intra-class diversity (e.g., Economy rice)

# Case Studies: Hard Cases

## Incomplete Food










# Case Studies: Hard Cases

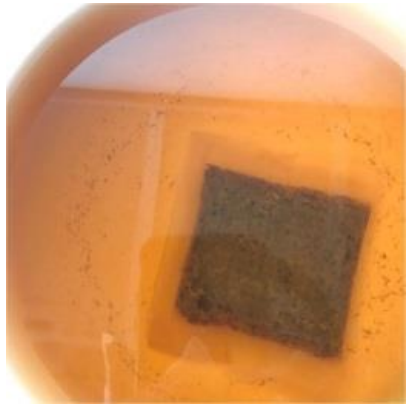
## Non Food






# Case Studies: Hard Cases

Poorly taken photos (illumination, rotation, occlusion, etc)










# Case Studies: Hard Cases

## Multiple food items










# Case Studies: Hard Cases

## Unknown food / food not in our list


# How to build a more sustainable solution?

## Better Learning

Go beyond supervised CNN

## Crowdsourcing

Combined with human wisdom

# Thank You!


http://www.foodai.org

Acknowledgements

National Research Foundation

Health Promotion Board

http://www.larc.smu.edu.sg