Food Image Recognition by Deep Learning

Assoc. Prof. Steven HOI

School of Information Systems
Singapore Management University
“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

❌ Tedious

❌ Non-efficient

❌ Non-effective
Smart Food Logging

Healthy 365

Powered by

FoodAi
Roadmap

Problem → Approach → Research → Cases
Food Image Recognition

• Visual Recognition

Machine Learning

Laksa?
Food Image Recognition

• Could be very challenging...

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

Food Item

Teh O
Teh O siu dai
Teh O kosong

Visual Food

Teh O

Food Category

Tea, no milk
Green tea
Green tea (no sugar)
Iced lemon tea
Roadmap

Problem → Approach → Research → Cases
Visual Recognition

- Classical Computer Vision Pipeline

- Deep Learning Approach
Deep Convolutional Neural Networks (CNN)

• Convolutional Neural Networks (CNN)


LeNet [LeCun et al. 1998]
Deep CNN for Visual Recognition

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

[ Krizhevsky et al. 2012 ]

ImageNet Classification top-5 error (%)
Why Deep Learning?

Machine Learning

HPC (GPU)

Data

Product

Accuracy

Small data

Big data

Data Size

Deep Learning

Traditional Learning
GPU for High Performance Computing

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

<table>
<thead>
<tr>
<th>SGFood724 Dataset</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td># total images</td>
<td>361,676</td>
<td>7,240</td>
<td>36,200</td>
</tr>
<tr>
<td># Image per class</td>
<td>~500</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

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

Histogram of #visual foods (724 visual food classes)
FoodAI: Open API Services

http://www.foodai.org

Smart Food Recognition with the state-of-the-art Visual Recognition technology
FoodAI System Architecture

Frontend
- App
- Web

Backend
- API Service
- MODEL INFERENCE ENGINE
- DATABASE
- Apache Thrift™

Offline
- MODEL TRAINING
- ANNOTATION SYSTEM
- EXTERNAL DATA COLLECTION

Tools:
- Caffe
- TensorFlow
- NVIDIA® DGX-1™
- Flask
- Python
- C++
- mongoDB
Roadmap

Problem → Approach → Research → Cases
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

<table>
<thead>
<tr>
<th>Models (SGFOOD)</th>
<th>Top-1 Accuracy (%)</th>
<th>Top-5 Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogleNet</td>
<td>71.5</td>
<td>91.0</td>
</tr>
<tr>
<td>ResNet-18</td>
<td>71.2</td>
<td>91.5</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>76.1</td>
<td>93.3</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>73.2</td>
<td>91.9</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>74.7</td>
<td>92.7</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Models (IMAGENET)</th>
<th>Top-1 Accuracy (%)</th>
<th>Top-5 Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-50</td>
<td>77.1</td>
<td>93.3</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>78.2</td>
<td>93.9</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>78.6</td>
<td>94.3</td>
</tr>
</tbody>
</table>
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

- ResNet-50 (10%)
- ResNet-50(10%)+augmentation
- ResNet-50 (100%)

<table>
<thead>
<tr>
<th>Model Type</th>
<th>TOP-1 Accuracy</th>
<th>TOP-5 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-50 (10%)</td>
<td>60.0</td>
<td>82.7</td>
</tr>
<tr>
<td>ResNet-50(10%)+augmentation</td>
<td>76.1</td>
<td>93.3</td>
</tr>
<tr>
<td>ResNet-50 (100%)</td>
<td>58.0</td>
<td>83.6</td>
</tr>
</tbody>
</table>
Roadmap

Problem → Approach → Research → Cases
Case Studies: Food logging photos from users
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)

[Images of various hard cases examples]
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

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