FOOD COUNTING
IS THIS POSSIBLE?

Chong-Wah Ngo
City University of Hong Kong

Joint work with Alan Nguyen
How far can multimedia automate food tracking?

**Food recognition**
- Dishes
- Ingredients
- Cooking methods
- Cutting methods

**Quantity estimation**
- Serving size
  - Gram, cup
- Number of items

**Calories**
**Nutrition fact**
**Health trend**

**Food Composition Table**

**DietCam** takes 3 pictures

**FoodLog, Japan**
- Incremental learning
- Personalization

**DietLens, NExT++**
- Deep learning
Current status in food recognition

AlexNet – Kung Pao Chicken
VGG – Kung Pao Chicken

Multi-task VGG – Kung Pao Chicken
[chicken, chili, peanut]

Region-based Multi-task VGG
chicken: dice, stir-fry
chili: dry
peanut: roasted
Multi-task Deep Learning

Pooling across scales and regions

Multi-scale

feature map

1 2 3 ... n

softmax

softmax

softmax

Ingredient Cooking Cutting

Region 1

Region n
### Key result

#### Recognition

<table>
<thead>
<tr>
<th>Category</th>
<th>Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>&gt; 95%</td>
<td>Top-5 accuracy</td>
</tr>
<tr>
<td>Ingredient (1,276)</td>
<td>≈ 40%</td>
<td>Recall</td>
</tr>
<tr>
<td>Cutting (8)</td>
<td>62%</td>
<td>Top-1 accuracy</td>
</tr>
<tr>
<td>Cooking (36)</td>
<td>46%</td>
<td>Top-1 accuracy</td>
</tr>
</tbody>
</table>

#### Recipe retrieval

<table>
<thead>
<tr>
<th>Category</th>
<th>Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>90%</td>
<td>63%</td>
</tr>
<tr>
<td>Ingredient (Unknown)</td>
<td>-</td>
<td>27%</td>
</tr>
</tbody>
</table>

Hit rate @ 10
nutrition information = food quantity \times \text{food density}

**Food quantity**: volume, weight or number of items (need measurement)

**Food density**: amount of nutrition per unit (assume given)
WEIGHT ESTIMATION

• Weight-area correlation

• **Procedure**
  • Collect dataset with pre-measured area and weight
  • Identify correlation between area (number of pixels) and weight
  • Online recognition
    ➢ Scale identification
    ➢ Image segmentation
    ➢ Food region identification & pixel counting
VOLUME ESTIMATION

Approach-1: Model fitting
Project 3D model to 2D image for scale & volume estimation

<table>
<thead>
<tr>
<th>Shape</th>
<th>Example Food Type</th>
<th>Dimension Parameters</th>
<th>Locator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cylinder</td>
<td>Orange juice, Milk</td>
<td>Radius, Height</td>
<td><img src="image" alt="Cylinder" /></td>
</tr>
<tr>
<td>Sphere</td>
<td>Apple, Orange</td>
<td>Radius</td>
<td><img src="image" alt="Sphere" /></td>
</tr>
<tr>
<td>Square Box</td>
<td>Chocoalte Cake, Brownie</td>
<td>Width, Length, Height, Rotation Angle</td>
<td><img src="image" alt="Square Box" /></td>
</tr>
<tr>
<td>Slice of Cone/</td>
<td>Spaghetti, Ice Cream</td>
<td>Top Radius, Bottom Radius, Height</td>
<td><img src="image" alt="Slice of Cone" /></td>
</tr>
<tr>
<td>Slice of Sphere</td>
<td></td>
<td></td>
<td><img src="image" alt="Slice of Sphere" /></td>
</tr>
<tr>
<td>Prism</td>
<td>Bread, Scrambled Eggs</td>
<td>Area, Height</td>
<td><img src="image" alt="Prism" /></td>
</tr>
<tr>
<td>Irregular Shape</td>
<td>Banana, Pear</td>
<td>Scale X, Scale Y, Scale Z, (Rotation Angle)</td>
<td><img src="image" alt="Irregular Shape" /></td>
</tr>
</tbody>
</table>
VOLUME ESTIMATION

Approach-2: Multi-view shape recovery
Reconstruct 3D model from video sequence

Step:
- Camera pose estimation
- 3D volumetric model reconstruction
- Voxel counting
VOLUME ESTIMATION

Approach-3:
Volume = Area x Depth

Step:
- Area estimation of bowl
- Depth value of the food (in a bowl) using depth camera
LIMITATIONS

- Limited number of food types
- Hard to scale up
- May not be user friendly
- How about deep-learning based approach?
  - Training data ..
NUMBER OF FOOD ITEMS

• A new (and simple) problem?

• Application

✓ Tell me *which* and *how much* should I eat today?
EXAMPLE

Nutrition information
- Calories: $3 \times 402$ cal
- Sodium: $3 \times 182$ mg
- Sugar: $3 \times 25$ g
- Protein: $3 \times 3$ g

Nutrition Facts

Costa - Caramel Shortbread

<table>
<thead>
<tr>
<th>Serving</th>
<th>1 slice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calories</td>
<td>402</td>
</tr>
<tr>
<td>Total Fat</td>
<td>26 g</td>
</tr>
<tr>
<td>Saturated</td>
<td>13 g</td>
</tr>
<tr>
<td>Polyunsaturated</td>
<td>0 g</td>
</tr>
<tr>
<td>Monounsaturated</td>
<td>0 g</td>
</tr>
<tr>
<td>Trans</td>
<td>0 g</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0 mg</td>
</tr>
<tr>
<td>Sodium</td>
<td>182 mg</td>
</tr>
<tr>
<td>Potassium</td>
<td>460 mg</td>
</tr>
<tr>
<td>Total Carbs</td>
<td>38 g</td>
</tr>
<tr>
<td>Dietary Fiber</td>
<td>1 g</td>
</tr>
<tr>
<td>Sugars</td>
<td>25 g</td>
</tr>
<tr>
<td>Protein</td>
<td>3 g</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>0%</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>0%</td>
</tr>
<tr>
<td>Calcium</td>
<td>0%</td>
</tr>
<tr>
<td>Iron</td>
<td>0%</td>
</tr>
</tbody>
</table>
FOOD COUNTING

• Motivation: How good can DCNN perform this task?

A toy cookie dataset
100 types of cookies
6,776 images
COOKIE DATASET
Pre-trained DCNN model

172 food
Single label learning

353 ingredients
Multi label learning

Convolution + ReLU
max pooling

Fully connected + ReLU
Softmax
Sigmoid
Food Counting

Convolution + ReLU max pooling

Fully connected + ReLU Softmax Sigmoid

Food counting
Single label learning
FEATURE MAP

Input images

Features map
Max pooling-5
# Confusion Matrix

<table>
<thead>
<tr>
<th>Correct Answer</th>
<th>1 item</th>
<th>2 items</th>
<th>3 items</th>
<th>4 items</th>
<th>5-6 items</th>
<th>above-6 items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 item</td>
<td>191</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2 items</td>
<td>13</td>
<td>157</td>
<td>19</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3 items</td>
<td>4</td>
<td>16</td>
<td>139</td>
<td>28</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>4 items</td>
<td>0</td>
<td>4</td>
<td>28</td>
<td>106</td>
<td>28</td>
<td>3</td>
</tr>
<tr>
<td>5-6 items</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>155</td>
<td>22</td>
</tr>
<tr>
<td>above-6 items</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>182</td>
</tr>
</tbody>
</table>
INCOMPLETE PREDICTION
SURPRISING RESULT
Ingredients (how to scale up?)
Nonvisible ingredients (??)

Cooking method (in progress)
Cutting method (in progress)

Volume estimation (some lights on counting)
FUTURE WORK

**Fundamental problems**
multi-dish segmentation + recognition + counting

**Recommendation**
Which and how much should I eat?