

การจำแนกและประมาณแคลอรีของอาหาร
จากรูปภาพโดยใช้การเรียนรู้ของเครื่อง
(Classification and Calorie Estimation from
Food Image by Machine Learning)

พิศลย์ เรือนอินทร์

590510525

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
ปีการศึกษา 2562

การจำแนกและประมาณแคลอรีของอาหาร
จากรูปภาพโดยใช้การเรียนรู้ของเครื่อง
(Classification and Calorie Estimation from
Food Image by Machine Learning)

พิศลย์ เรือนอินทร์

590510525

การค้นคว้าอิสระนี้ได้รับการพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
ปีการศึกษา 2562

คณะกรรมการสอบการค้นคว้าอิสระ

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.จักริน ขวชาติ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.จักรเมธ บุตรกระจำง)

วันที่.....เดือน.....พ.ศ.....

กิตติกรรมประกาศ

การค้นคว้าอิสระเล่มนี้สำเร็จลุล่วงได้โดยความอนุเคราะห์จากบุคคลหลายท่าน ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.จักริน ขวชาติ อาจารย์ที่ปรึกษา ที่กรุณาให้ความรู้ แนวทาง และคำแนะนำ รวมทั้งเสียสละเวลาอันมีค่าในการตรวจสอบแก้ไขข้อบกพร่องการระบบการทำงาน และเนื้อในการค้นคว้าอิสระ จนการค้นคว้านี้เสร็จสมบูรณ์

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.จักรเมธ บุตรกระจำง ที่ได้ให้คำปรึกษาคำแนะนำที่ดีเกี่ยวกับการเรียนรู้ของเครื่อง และมอบความรู้วิชาอันมีค่า รวมทั้งมีความกรุณารับเป็นกรรมการสอบการค้นคว้าอิสระนี้ ซึ่งให้คำแนะนำที่ดีมาโดยตลอดและให้ความช่วยเหลือในด้านต่างเป็นอย่างดี

นอกจากนี้ ขอขอบคุณทุกความช่วยเหลือ ทั้งทางตรงและทางอ้อมที่มีได้กล่าวไว้ ณ ที่นี้ ซึ่งท่านเหล่านั้นได้กรุณาให้คำแนะนำในตการทำการค้นคว้าอิสระให้สำเร็จลุล่วงไปด้วยดี

พิศลย์ เรือนอินทร์

590510525

หัวข้อการค้นคว้าอิสระ	การจำแนกและประมาณอาหารจากรูปภาพโดยใช้การเรียนรู้ของเครื่อง
ชื่อเจ้าของโครงการ	นายพิศลย์ เรือนอินทร์ รหัสประจำตัว 590510525
วิทยาศาสตร์บัณฑิต	สาขาวิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. จักริน ขวชาติ

บทคัดย่อ

การขาดแคลนสารอาหารในผู้ป่วยถือเป็นปัญหาที่สำคัญ โดยเฉพาะอย่างยิ่งผู้ป่วยในโรงพยาบาล โดยทั่วไปแล้วนักโภชนาการจะเป็นที่คอยคำนวณสารอาหารที่ผู้ป่วยได้รับ ซึ่งในปัจจุบันปัญหาในโรงพยาบาลคือการขาดแคลนนักโภชนาการ และการคำนวณปริมาณอาหารที่ได้รับใช้เวลานาน การวิจัยนี้จึงจะนำเสนอการประมาณปริมาณอาหารที่ผู้ป่วยในโรงพยาบาลได้รับ โดยการใช้รูปภาพของอาหารในถาดหลุมก่อนรับประทานอาหารและหลังรับประทานอาหาร

ระบบแบ่งออกเป็นสองส่วน คือ การตรวจจับอาหารและจำแนกอาหาร และการประมาณน้ำหนักของอาหาร ในส่วนแรกได้ใช้เทคนิคของ Faster R-CNN และเลือกใช้ ResNet-50 เป็น pre-trained model ซึ่งแบบจำลองการทำนายถูกเรียนรู้ด้วยรูปภาพอาหาร 16,067 รูป โดยแบ่งออกเป็นอาหาร 39 รายการที่แตกต่างกัน แล้วได้ทำการทดสอบการรูปภาพจำนวน 4017 รูป ซึ่งได้ค่า mAP เท่ากับ 73.354 และในส่วนที่สองได้ใช้เทคนิคของ CNN และเลือกใช้ InceptionResNetV2 เป็น pre-trained model แล้วได้ทำการทดสอบกับรูปภาพจำนวน 4017 รูป ซึ่งได้ค่า MAPE เท่ากับ 16.9729 ซึ่งถือว่าเป็นผลลัพธ์ที่ดี เมื่อได้น้ำหนักของอาหารแล้วจึงนำไปเปรียบเทียบบัญญัติไตรยางค์กับข้อมูลแคลอรีต่อน้ำหนักของอาหารจากตารางอ้างอิง

Independent Study Title	Classification and Calorie Estimation from Food Image by Machine Learning
Author	Mr. Pisol Ruenin Student ID 590510525
Bachelor of Science	Computer Science
Supervisor	Asst. Prof. Dr. Jakarin Chawachat

Abstract

Malnutrition in the patient is an important issue, especially for the patient in hospitals. In general, nutritionists calculate the number of supplements with the patient. Currently, the problem in hospitals is the lack of nutritionists, and the calculation of the amount of food takes a long time. In this research, we present a system to estimate the amount of food that the patient in the hospital consume by taking pictures in divided food tray before and after consumption.

Our system has two main parts: food detection and classification, and food weight estimation. In the first part, we use the Faster R-CNN technique and select ResNet-50 as a pre-trained model. Our prediction model was trained on a dataset of 16,067 food images from 39 different classes. The result with 4017 food images showed that $mAP = 73.354$. In the second part, we use the CNN technique, which uses a pre-trained model as InceptionResNetV2. From the experiment with the 4017 food images found that $MAPE = 16.9729$, which is considered a good prediction. After that, the weight of food has used a rule of three to compare a reference table that compares calories by weight.

สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
Abstract	ค
สารบัญตาราง	ฉ
สารบัญภาพ	ช
บทที่ 1 บทนำ	1
1.1 หลักการและเหตุผล	1
1.2 วัตถุประสงค์ของโครงการ	1
1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงประยุกต์และเชิงทฤษฎี	1
1.4 ขอบเขตของโครงการ	2
1.5 แผนการดำเนินงานและระยะเวลาการดำเนินงาน	3
บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง	4
2.1 การเพิ่มข้อมูลตัวอย่าง	4
2.2 โครงข่ายประสาทเทียม	5
2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน	6
2.4 Faster Region with Convolutional Neural Network	6
2.5 ถ่ายโอนการเรียนรู้	7
2.6 Keras	8
2.7 งานวิจัยที่เกี่ยวข้อง	8
บทที่ 3 การวิเคราะห์และออกแบบระบบ	12
3.1 การเตรียมชุดข้อมูลรูปภาพอาหาร	12
3.2 การจำแนกอาหารจากรูปภาพ	14
3.3 การประมาณน้ำหนักของอาหารจากรูปภาพ	14
บทที่ 4 การทดสอบระบบ	16
4.1 การทดสอบระบบการจำแนกอาหารจากรูปภาพ	16
4.2 การทดสอบระบบการประมาณน้ำหนักของอาหารจากรูปภาพ	19

สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
Abstract	ค
สารบัญตาราง	ฉ
สารบัญภาพ	ช
บทที่ 1 บทนำ	1
1.1 หลักการและเหตุผล	1
1.2 วัตถุประสงค์ของโครงการ	1
1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงประยุกต์และเชิงทฤษฎี	1
1.4 ขอบเขตของโครงการ	2
1.5 แผนการดำเนินงานและระยะเวลาการดำเนินงาน	3
บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง	4
2.1 การเพิ่มข้อมูลตัวอย่าง	4
2.2 โครงข่ายประสาทเทียม	5
2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน	6
2.4 Faster Region with Convolutional Neural Network	6
2.5 ถ่ายโอนการเรียนรู้	7
2.6 Keras	8
2.7 งานวิจัยที่เกี่ยวข้อง	8
บทที่ 3 การวิเคราะห์และออกแบบระบบ	12
3.1 การเตรียมชุดข้อมูลรูปภาพอาหาร	12
3.2 การจำแนกอาหารจากรูปภาพ	14
3.3 การประมาณน้ำหนักของอาหารจากรูปภาพ	14
บทที่ 4 การทดสอบระบบ	16
4.1 การทดสอบระบบการจำแนกอาหารจากรูปภาพ	16
4.2 การทดสอบระบบการประมาณน้ำหนักของอาหารจากรูปภาพ	19

สารบัญ (ต่อ)

	หน้า
บทที่ 5 บทสรุป	22
5.1 สรุปผลการดำเนินการ	22
5.2 ข้อเสนอแนะสำหรับการพัฒนาในอนาคต	22
เอกสารอ้างอิง	23

สารบัญตาราง

ตารางที่		หน้า
1.1	ขั้นตอนการดำเนินงานและระยะเวลาดำเนินงาน	3
2.1	ผลการทดลองของ Gozde Ozsert Yigit และ B. Melis Ozyildirim	10
3.1	การเปรียบเทียบแคลอรีกับน้ำหนักของอาหาร	15
4.1	ค่า Mean Absolute Percentage Error (MPAE) ของอาหารแต่ละรายการ	20

สารบัญภาพ

ภาพที่		หน้า
2.1	ตัวอย่างรูปภาพก่อนและหลังทำ Data Augmentation	4
2.2	โครงสร้างของโครงข่ายประสาทเทียม	5
2.3	โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน	6
2.4	โครงสร้างของ Faster Region with Convolutional Neural Network	7
2.5	3D Reconstructed point of gram	8
2.6	วิธีการวัดพื้นที่ของอาหาร	11
3.1	ผังการทำงานของการทำงานจำแนกและประมาณแคลอรีจากรูปภาพ	12
3.2	ตัวอย่างรูปภาพอาหาร	13
4.1	การเปรียบเทียบค่าของ Loss และ Epochs ของการจำแนกวัตถุใน RPN	16
4.2	การเปรียบเทียบค่าของ Loss และ Epochs ของการหาขอบเขตของวัตถุใน RPN	17
4.3	การเปรียบเทียบค่าของ Loss และ Epochs ของการจำแนกวัตถุใน Classifier	17
4.4	การเปรียบเทียบค่าของ Loss และ Epochs ของการหาขอบเขตของวัตถุใน Classifier	17
4.5	ค่า AP ของอาหารแต่ละรายการ	19
4.6	การเปรียบเทียบค่าของ Loss และ Epochs ในการประมาณน้ำหนัก	19
4.7	ตัวอย่างการแสดงผลการทำงานของการทำงานจำแนกอาหารและประมาณแคลอรีบนแอปพลิเคชัน	21

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

ในแต่ละวันแต่ละคนต้องการพลังงานจากอาหารหรือที่เรียกว่า แคลอรี นั้นไม่เท่ากัน ซึ่งมีปัจจัยหลายอย่างเป็นตัวกำหนดปริมาณแคลอรีที่ควรจะได้รับในแต่ละคน เช่น อายุ ลักษณะของกิจกรรมในแต่ละวัน การเจ็บป่วย เป็นต้น หากได้รับแคลอรีมากหรือน้อยเกินไปในหนึ่งวันจะเกิดผลกระทบต่อร่างกายได้ โดยเฉพาะอย่างยิ่งในผู้ป่วย ปริมาณแคลอรีและสารอาหารในแต่ละวันนั้นจะถูกกำหนดโดยแพทย์หรือนักโภชนาการ เพื่อช่วยให้ผู้ป่วยได้รับพลังงานอย่างเพียงพอในแต่ละวัน

ในปัจจุบันวิธีการคำนวณปริมาณแคลอรีที่ผู้ป่วยต้องการคือการให้นักโภชนาการคำนวณโดยตรง เริ่มต้นนักโภชนาการนั้นจะทราบพลังงานที่ผู้ป่วยต้องการในแต่ละวัน จากนั้นจะกำหนดประเภทและปริมาณของอาหารในแต่ละมื้อ ในการวัดปริมาณแคลอรีที่ผู้ป่วยได้รับในแต่ละมือนักโภชนาการจะนำอาหารไปชั่งน้ำหนักก่อนรับประทานอาหารและหลังรับประทานอาหาร จากนั้นจึงคำนวณปริมาณแคลอรีที่หายไปจากน้ำหนักก่อนและหลังเทียบกับปริมาณแคลอรีเฉลี่ย วิธีการดังกล่าวยังมีความคลาดเคลื่อนจากปริมาณของอาหารก่อนรับประทาน เพราะ ปริมาณอาหาร 1 หน่วยที่ให้แกผู้ป่วยแต่ละคนอาจจะมีปริมาณไม่เท่ากัน ยิ่งไปกว่านั้นยังต้องเสียเวลาไปกับการนำอาหารไปแยกส่วนและชั่งน้ำหนักในแต่ละครั้งอีกด้วย

ด้วยปัญหาเหล่านี้ ผู้จัดทำจึงได้เสนอแบบจำลองที่สามารถจำแนกและประมาณแคลอรีจากรูปภาพ โดยนำหลักการเรียนรู้ของเครื่องมาประยุกต์ใช้ เพื่อระบุปริมาณแคลอรีก่อนรับประทานอาหาร ปริมาณแคลอรีหลังรับประทานอาหาร และปริมาณแคลอรีที่หายไป เพื่อลดความคลาดเคลื่อนของปริมาณอาหารก่อนรับประทานอาหาร และสามารถช่วยลดเวลาในการประเมินแคลอรีอาหารของผู้ป่วยในแต่ละคนได้

1.2 วัตถุประสงค์ของโครงการ

- 1) เพื่อจำแนกอาหารจากรูปภาพโดยใช้การเรียนรู้ของเครื่อง
- 2) เพื่อประมาณแคลอรีอาหารที่หายไปจากรูปโดยใช้การเรียนรู้ของเครื่อง

1.3 ประโยชน์ที่จะได้รับการศึกษาเชิงประยุกต์และเชิงทฤษฎี

- 1) สามารถจำแนกอาหารจากรูปภาพได้
- 2) สามารถประมาณแคลอรีอาหารที่หายไปจากรูปภาพได้

1.4 ขอบเขตของโครงการ/วิธีการวิจัย

1.4.1 ขอบเขตทางสถาปัตยกรรม

1) ฮาร์ดแวร์ (Hardware)

Processor: AMD Ryzen 7 1700 Eight-Core Processor

Memory: 16 GB

GPU: NVIDIA GeForce GTX 1080

VRAM: 8 GB

2) ซอฟต์แวร์ (Software)

- Windows 10 Education 64-bit
- Python 3.7.3
- Jupyter Notebook 6.0.1
- Keras 2.3.1

1.4.2 ขอบเขตของระบบงาน

การจำแนกและประมาณแคลอรีของอาหารจากรูปภาพ โดยใช้การเรียนรู้ของเครื่อง ได้แบ่งการทำงานออกเป็น 2 ส่วน คือ ส่วนของการจำแนกอาหารจากรูปภาพ และส่วนของการประมาณแคลอรีอาหารจากรูปภาพ โดยแต่ละส่วนมีขอบเขตการทำงานดังนี้

1) การจำแนกอาหารจากรูปภาพ

แบบจำลองสามารถจำแนกอาหารได้ โดยใช้วิธีการของ Faster Region Convolutional Neural Network Faster R-CNN [4] ซึ่งจะสามารถแบ่งคุณสมบัติต่างๆ ของรูปภาพ เช่น เส้นขอบ สี จุดตัด เป็นต้น จากการผ่านตัวกรองที่ถูกปรับรูปแบบให้เหมาะสม เพื่อนำมาเป็นข้อมูลในมิติต่างๆ นำไปหาตำแหน่งและพื้นที่ที่สนใจของอาหาร ก่อนที่จะผ่านส่วนของการจำแนกกว่าเป็นอาหารชนิดใด

2) การประมาณแคลอรีอาหารที่หายไปจากรูปภาพ

แบบจำลองสามารถประมาณแคลอรีอาหาร โดยใช้ Convolutional Neural Network [1] นำเข้าผลลัพธ์จากขั้นตอนของการจำแนกอาหารจากรูปภาพเพื่อนำข้อมูลคุณลักษณะต่าง ๆ ที่ได้จากขั้นตอนดังกล่าว เป็นข้อมูลช่วยระบุน้ำหนักของอาหารก่อนที่จะนำไปเปรียบเทียบกับข้อมูลที่ได้จากแคลอรีต่อน้ำหนักอาหารจากตารางอ้างอิง

1.4.3 ขอบเขตของข้อมูล

1) ข้อมูลนำเข้า

รูปภาพอาหารในถาดก่อนรับประทานอาหารและหลังรับประทานอาหาร จากอาหารผู้ป่วยในโรงพยาบาล ซึ่งมีอาหารทั้งหมด 39 ชุด แต่ละรูป หรือชุดมีอาหารไม่เกิน 3 ประเภท

2) ข้อมูลส่งออก

ชื่อของอาหาร และปริมาณแคลอรีของอาหารที่หายไป

1.5 แผนการดำเนินงานและระยะเวลาดำเนินงาน

การศึกษานี้เริ่มดำเนินงานตั้งแต่เดือนสิงหาคม พ.ศ. 2562 ถึง เดือนมีนาคม พ.ศ. 2563 แสดงรายละเอียดการดำเนินงานดังตารางที่ 1.1 ซึ่งมีขั้นตอนการดำเนินงานดังนี้

1. ศึกษาทฤษฎีและหลักการที่เกี่ยวข้อง
2. วิเคราะห์ข้อมูล
3. ออกแบบการทดลอง
4. ทดสอบผลการทดลอง
5. วิเคราะห์ผลลัพธ์ และสรุปผลการทดลอง
6. จัดทำเอกสาร

ตารางที่ 1.1 ขั้นตอนการดำเนินงานและระยะเวลาดำเนินงาน

ระยะเวลา ขั้นตอนการดำเนินงาน	ปี พ.ศ. 2562					ปี พ.ศ. 2563		
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
1. ศึกษาทฤษฎีและหลักการที่เกี่ยวข้อง								
2. วิเคราะห์ข้อมูล								
3. ออกแบบการทดลอง								
4. ทดสอบผลการทดลอง								
5. วิเคราะห์ผลลัพธ์ และสรุปผลการทดลอง								
6. จัดทำเอกสาร								

บทที่ 2

หลักการและทฤษฎีที่เกี่ยวข้อง

ในการจำแนกและประมาณแคลอรีของอาหารจากรูปภาพโดยใช้การเรียนรู้ของเครื่องจำเป็นต้องศึกษาและเข้าใจเกี่ยวกับการจำแนกและการวิเคราะห์การถดถอยของข้อมูลโดยใช้โครงข่ายประสาทเทียม และการประมาณปริมาณของอาหารจากรูปภาพ รวมไปถึงการประมวลผลข้อมูลเพื่อให้ข้อมูลถูกนำมาใช้งานได้อย่างมีประสิทธิภาพ จากที่กล่าวมาจึงได้สรุปหลักการและทฤษฎีที่เกี่ยวข้องดังนี้

2.1 การเพิ่มตัวอย่างข้อมูล (Data Augmentation)

การเพิ่มตัวอย่างข้อมูลเป็นการสร้างข้อมูลใหม่จากการปรับเปลี่ยนลักษณะของข้อมูลเดิม เช่น หมุนรูปภาพ, ปรับความเข้มสีของรูปภาพ, ขยายหรือย่อรูปภาพ เป็นต้น โดยนอกจากจะทำให้ข้อมูลมีจำนวนมากขึ้นแล้ว ยังสามารถหลีกเลี่ยงการ Overfitting ของโมเดล จากการทดลองของ Jason Wang และ Luis Perez [3] ที่ได้ทดลองกับการจำแนกรูปภาพด้วยการเรียนรู้เชิงลึก จะเห็นได้ว่าข้อมูลที่มีการเพิ่มตัวอย่างข้อมูลมีผลลัพธ์ที่ดีกว่าข้อมูลที่ไม่ได้เพิ่มตัวอย่างข้อมูล



ภาพที่ 2.1 ตัวอย่างรูปภาพก่อนและหลังการทำ Data Augmentation

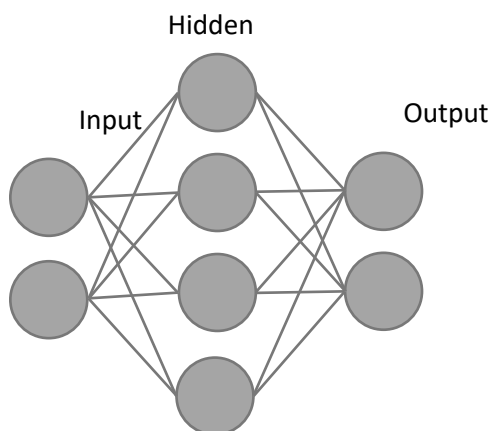
2.2 โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม [2] เป็นการจำลองมาจากสมองของมนุษย์ โดยมีหน่วยประมวลผลหลายๆ ส่วนเชื่อมโยงกันด้วยโครงข่ายประสาทมากมาย โดยหน่วยที่เล็กที่สุดคือ Neural ซึ่งแต่ละ Neural จะมี Activation Function เช่น Sigmoid, Rectified Linear Unit, Softmax เป็นต้น ทำหน้าที่รับข้อมูลนำเข้ามาประมวลผลแล้วนำส่งออกต่ออย่างชั้นถัดไป ซึ่งโครงข่ายประสาทเทียมที่จะใช้ในการจำแนกหรือการวิเคราะห์การถดถอยขึ้นอยู่กับ Activation Function ที่ใช้ โดยโครงข่ายประสาทเทียมจะประกอบไปด้วย 3 ส่วน ดังนี้

2.2.1 Input Layer ชั้นรับข้อมูลโดยมีจำนวนของ Neural ตามจำนวนข้อมูลที่เกี่ยวข้องกับข้อมูลส่งออก

2.2.2 Hidden Layer ชั้นที่อยู่ระหว่างกลางซึ่งมีผลต่อการทำงานและประสิทธิภาพของโมเดล ซึ่งจำนวนชั้นและจำนวนของ Neural ขึ้นอยู่กับลักษณะการกระจายตัวของข้อมูล

2.2.3 Output Layer ชั้นส่งออกข้อมูลโดยมีจำนวนของ Neural ขึ้นอยู่กับรูปแบบของผลลัพธ์

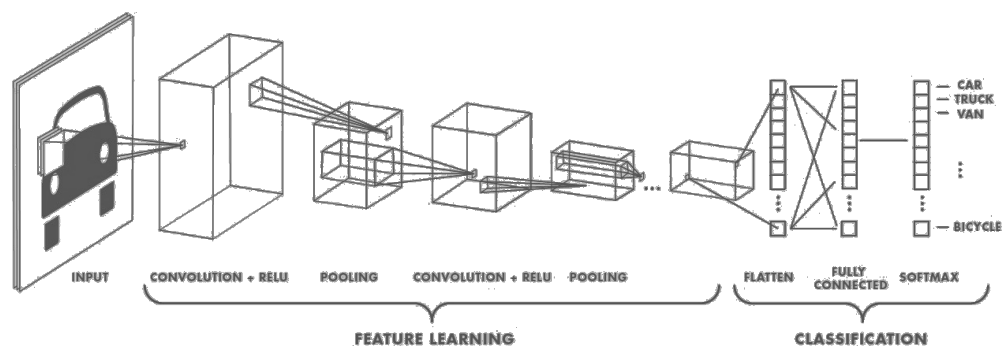


ภาพที่ 2.2 โครงสร้างของโครงข่ายประสาทเทียม

2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network)

โครงข่ายประสาทเทียมแบบคอนโวลูชัน [1] เป็นการเรียนรู้แบบเชิงลึกที่สามารถแบ่งคุณสมบัติย่อยต่างๆ เช่น เส้นขอบ, สี, ขอบเขตของวัตถุที่สนใจ เป็นต้น ออกมาเป็นข้อมูลที่ใช้สำหรับจำแนกหรือวิเคราะห์การถดถอยของข้อมูลซึ่งปกติแล้วข้อมูลเหล่านั้นมนุษย์ต้องหาวิธีแบ่งคุณสมบัติต่างๆ ด้วยเหตุนี้โครงข่ายประสาทเทียมแบบคอนโวลูชันจึงแก้ปัญหาดังกล่าวได้ โดยแบ่งส่วนประกอบออกเป็น 3 ส่วน ดังนี้

- 2.3.1 Convolution Layer** มีหน้าที่สร้างหน้าภาพกรงรูปภาพเพื่อดึงคุณลักษณะของรูปภาพ โดยที่ลักษณะของหน้าภาพที่ใช้กรงเหล่านี้นจะถูกเรียนรู้จากโมเดลว่าควรใช้หน้าภาพลักษณะใดจะได้คุณลักษณะของข้อมูลที่เหมาะสม
- 2.3.2 Pooling Layer** มีหน้าที่เพิ่มประสิทธิภาพของการประมวลผลข้อมูลโดยการลดมิติของข้อมูลจากการดึงข้อมูลที่แสดงถึงพื้นที่นั้นๆ ออกมาให้มีขนาดลดลง โดยมี 2 ประเภทหลักที่นิยมคือ Max Pooling เป็นการดึงค่าสูงสุดในบริเวณที่กำหนด และ Average Pooling เป็นการหาค่าเฉลี่ยของบริเวณที่กำหนด
- 2.3.3 Fully Connected Layer** มีหน้าที่รับเวกเตอร์ที่ถูกแปลงจากเมทริกซ์ของการส่งออกจากกรงคุณลักษณะของข้อมูล เพื่อนำข้อมูลเหล่านั้นมาประมวลผลตามลักษณะของโครงข่ายประสาทเทียมในรูปแบบของผลลัพธ์ที่ต้องการ

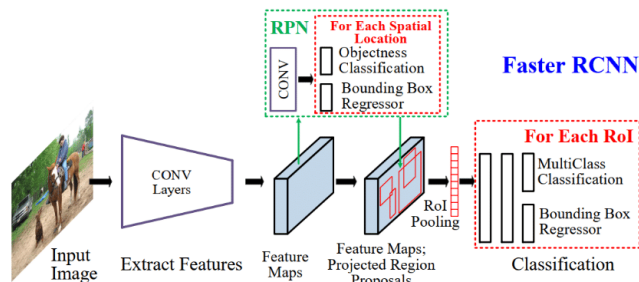


ภาพที่ 2.3 โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

(แหล่งที่มา: <https://bit.ly/2L7mBKK>)

2.4 Faster Region with Convolutional Neural Network (Faster R-CNN)

Faster R-CNN พัฒนามาจาก R-CNN และ Fast R-CNN ซึ่งเป็น Two-Stage Object Detection โดยจะทำการหาพื้นที่ที่สนใจบนรูปภาพด้วยความใกล้เคียงกับพื้นที่วัตถุในเฉลยผ่านค่า Intersection over Union (IoU) แล้วถึงจะทำการจำแนกพื้นที่ดังกล่าว ซึ่ง Shaoqing Ren และคณะ [4] ได้นำวิธีการตัดพื้นที่ที่น่าสนใจด้วย Selective Search ใน CPU ออกไป แล้วใช้วิธี Region Proposal Network (RPN) มาแทน ซึ่งทำงานบน GPU ทั้งหมด ทำให้ลดระยะในการส่งข้อมูลระหว่าง CPU และ GPU ความเร็วในการประมวลผลจึงมากกว่า R-CNN และ Faster R-CNN โดย RPN จะระบุวัตถุ และขอบเขตวัตถุที่สนใจ ก่อนจะระบุประเภทของรูปภาพ และพิกัดตำแหน่งที่อยู่ โดยชั้นของ Classifier



ภาพที่ 2.4 โครงสร้าง Faster Region with Convolutional Neural Network

(แหล่งที่มา: https://www.researchgate.net/figure/High-level-diagram-of-Faster-RCNN-16-for-generic-object-detection-2-Inception-v2-The_fig3_334987612)

2.5 โอนการเรียนรู้ (Transfer Learning)

โอนการเรียนรู้ [5] เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยเป็นการนำความรู้เดิมที่ถูกเรียนรู้จากปัญหาชุดหนึ่ง นำไปเรียนรู้ให้ปัญหาอื่นๆ ที่มีลักษณะของปัญหาใกล้เคียงกันได้ จากการทำงานดังกล่าวจึงช่วยให้แก้ปัญหาดังกล่าวได้เร็วขึ้นและมีข้อจำกัดมากมายได้ แต่หากต้นแบบของการโอนความรู้ไม่สอดคล้องกับปัญหาที่กำหนดก็จะทำให้เกิดความปัญหาที่เรียกว่า Negative Transfer ได้ การโอนความรู้สามารถแบ่งได้เป็น 3 ประเภท ดังนี้

2.5.1 Inductive Transfer Learning คือ ต้นแบบและเป้าหมายอยู่ในโดเมนเดียวกัน แต่ผลลัพธ์ของงานต้นแบบและเป้าหมายแตกต่างกัน เช่น ต้นแบบเป็นการเรียนรู้ข้อมูลรูปภาพของมนุษย์เพื่อจำแนกเพศ แต่เป้าหมายเป็นการเรียนรู้มนุษย์เพื่อจำแนกอาชีพ

2.5.2 Unsupervised Transfer Learning คือ มีลักษณะคล้ายกับ Inductive Transfer Learning แต่เป็นในเรื่องงานของ Unsupervised เป็นโดเมนหลักของปัญหา เช่น ต้นแบบต้องการจัดกลุ่มของรูปภาพตัวเลขอารบิก แต่เป้าหมายต้องการจัดกลุ่มรูปภาพของตัวเลขไทย เป็นต้น

2.5.3 Transductive Transfer Learning คือ ต้นแบบและเป้าหมายอยู่ต่างโดเมนกัน แต่ต้องการผลลัพธ์งานเดียวกัน เช่น ต้นแบบและเป้าหมายต้องการจำแนกเลข 8 โดยที่ต้นแบบเป็นจำนวนของวงกลม แต่เป้าหมายเป็นรูปภาพของตัวเลขอารบิก

งานวิจัยนี้ได้นำใช้หลักการโอนการเรียนรู้ด้วย Pre-trained model ในการจำแนกรายการอาหารและประมาณน้ำหนักของอาหารจากรูปภาพโดยใช้ ResNet50 [19] ที่มีความลึก 50 ชั้น และ InceptionResNetV2 [18] ที่มีความลึก 164 ชั้น ตามลำดับ ซึ่งทั้งสองโครงสร้างเป็น Convolutional Neural Network ที่ถูกเรียนรู้ด้วยรูปภาพมากกว่าหนึ่งล้านรูป ที่สามารถแบ่งรูปภาพออกได้เป็น 1000 ประเภท จาก ImageNet

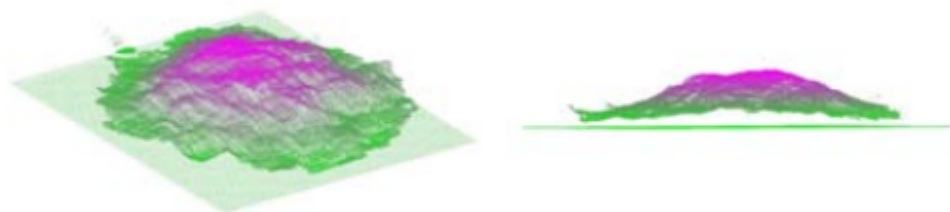
2.6 Keras

Keras [6] เป็น API สำหรับโครงข่ายประสาทเทียมระดับสูง ที่ถูกพัฒนาด้วยภาษา Python ประโยชน์ คือ ช่วยลดภาระการเรียนรู้โมเดลสำหรับโครงสร้างประสาทเทียมระดับต่ำที่มีความซับซ้อน สามารถเข้าใจได้ง่ายและไวต่อการปรับเป็นมิตรต่อผู้ใช้งาน ซึ่งทำงานอยู่บนโมเดล TensorFlow, CNTK และ Theano นอกจากนี้ยังรองรับการเรียนรู้เชิงลึกทั้ง Convolutional Neural Network และ Recurrent Neural Network สามารถประมวลผลผ่านได้ทั้ง CPU และ GPU

2.7 งานวิจัยที่เกี่ยวข้อง

2.7.1 การประมาณปริมาตรอาหารจากรูปภาพ

จากการทดลองที่เกี่ยวข้องกับการประมาณปริมาตรของอาหารจากรูปภาพโดย Manika และคณะ [7] ได้มีการทดลองหาปริมาตรของอาหารถูกประมาณด้วยการเริ่มคำนวณมุมกล้องของแต่ละเฟรมรูปภาพด้วย RANSAC หลังจากนั้นจึงประมวลผลด้วย Dense Stereo Reconstruction เพื่อจำลองจากสองมิติให้กลายเป็นสามมิติ แล้วปรับสเกลให้ภาพสามมิติหลังจากนั้นจึงใช้วิธี Delaunay triangulation เทียบลงบนอาหาร แล้วคำนวณปริมาตรจากปริมาตรของ Delaunay triangles ทั้งหมด ซึ่งมีความคลาดเคลื่อนอยู่ในช่วง 1.5% ถึง 10.9%



ภาพที่ 2.5 3D reconstructed point cloud of gram

(แหล่งที่มา : Recognition and Volume Estimation
of Food Intake using a Mobile Device [7])

จากการทดลองของ Joachim และคณะ [8] ได้ทำการสร้างภาพสามมิติใหม่จากรูปภาพอาหารสองรูปเพื่อประมาณปริมาตรของอาหารโดยใช้ Stereo Vision ซึ่งเป็นการจำลองการมองตาซ้ายและตาขวาของมนุษย์ทำให้ทราบได้ว่าวัตถุที่อยู่ภาพมีลึกเท่าใด จำเป็นต้องมีแผนที่เป็นการอ้างอิงขนาดของอาหารในรูปภาพทั้งสองรูป ในการสร้างรูปภาพสามมิตินั้นจะต้องภาพสามขั้นตอน คือ Rectification, Stereo Matching และ Point Cloud Generation เมื่อสร้างรูปสามมิติจึงแบ่งพื้นผิวของอาหารออกเป็นสามเหลี่ยมโดย

Delaunay Triangulation แล้วจะหาปริมาตรได้จากผลรวมของพื้นผิวสามเหลี่ยมทั้งหมด คูณกับค่าเฉลี่ยความสูงของมุม ซึ่งได้ผลการทดลองกับชุดข้อมูลรูปภาพสามชุดมีค่า Mean Absolute Percentage Error เท่ากับ 8.2%, 9.8%, 7.4%

นอกจากนี้ได้มีการทดลองโดย Chang และคณะ [9] ประมาณปริมาตรจากรูปภาพหนึ่งรูปสามารถประมาณได้จากการสังเกตรูปทรงของอาหารซึ่งได้มีการทดลอง เพื่อเปรียบเทียบกับรูปทางเรขาคณิตกับอาหาร แล้วนำข้อมูลที่เกี่ยวข้องกับรูปทรงของอาหารที่มีอยู่คำนวณโดยใช้สูตรการคำนวณของรูปทรงเรขาคณิตที่มีลักษณะใกล้เคียง ซึ่งมีความคลาดเคลื่อนอยู่ในช่วง 6.9% ถึง 10.7% และได้มีการทดลองจากรูปภาพหลายๆมุมของอาหาร แล้วหารูปทรงของอาหารโดยใช้วิธี Shape from Silhouettes โดยวิธีนี้จะสร้างแบบจำลองสามมิติจากชุดของ contour ที่มีโครงมาจากการฉายภาพของวัตถุลงบนลำดับของระนาบรูปภาพสองมิติ ได้ผลการทดลองที่มีความคลาดเคลื่อนอยู่ในช่วง 5% ถึง 18.2%

2.7.2 การจำแนกอาหารจากรูปภาพ

การจำแนกอาหารจากรูปภาพได้มีการทดลองโดย Duc และคณะ [10] ด้วยการใช้ประโยชน์จาก Local Appearance ด้วยวิธี Non-redundant Local Binary และ Global Structural Information ด้วยวิธีการหาความสัมพันธ์เชิงพื้นที่ระหว่างจุดที่สนใจของรูปภาพอาหาร นอกจากนี้ยังนำวิธีดังกล่าวประยุกต์ใช้รวมกัน ซึ่งผลลัพธ์จะเห็นได้ว่าการนำทั้งสองวิธีรวมกันสามารถช่วยเพิ่มประสิทธิภาพของการจำแนกได้ ซึ่งได้แบ่งการทดลองออกเป็นสองวิธี โดยวิธีแรกใช้รูปแบบของ Bag-of-Features แต่วิธีที่สองใช้การรวมกันของ Local Texture และ Global Structure ซึ่งได้ผลความแม่นยำเฉลี่ย 68% และ 69% ตามลำดับ

ปัจจุบันการจำแนกอาหารจากรูปภาพนิยมใช้วิธีการเรียนรู้เชิงลึกด้วย โครงข่ายประสาทเทียมแบบคอนโวลูชัน มาช่วยในการจำแนกมากขึ้น ซึ่งมีงานวิจัยที่ได้เปรียบเทียบประสิทธิภาพของโครงสร้างแต่ละแบบของโครงข่ายประสาทเทียมโดย Gozde Ozsert Yugit และ B. Melis Ozyildirim [11] จะเห็นได้ว่ามีความแม่นยำในระดับที่ดีโดยได้เปรียบเทียบโครงสร้างของโครงข่ายประสาทเทียม 5 โครงสร้าง ซึ่งมีทั้งโครงสร้างที่เป็นที่นิยมและโครงสร้างที่สร้างขึ้นเองโดยมีโครงสร้างโครงข่ายประสาทเทียม คือ CaffeNet, AlexNet, โครงสร้างหมายเลข 1, โครงสร้างหมายเลข 2 และโครงสร้างหมายเลข 3 โดยได้ผลดังตารางที่ 2.1

ตารางที่ 2.1 ผลการทดลองของ Gozde Ozsert Yigit และ B. Melis Ozyildirim

Model	Accuracy (%) (SGD)	Accuracy (%) (Nesterov's accelerated gradient)	Accuracy (%) (Adam)
CaffeNet	80.51	81.55	83.7
AlexNet	82.07	83.01	86.92
Structure 1	70.12	71.55	73,8
Structure 2	38.85	47.63	71.7
Structure 3	53.78	68.55	69.92

นอกจากนี้ได้มีการทดลองโดย Keiji Yanai และ Yoshiyuki Kawano [12] ได้ใช้ Deep Convolutional Neural Network โดยใช้เทคนิคของ Pre-training และ Fine-tuning ซึ่ง DCNN เป็นโครงสร้างที่ชนะเลิศใน ImageNet Large Scale Visual Recognition Challenge 2012 ผู้ทดลองได้เปรียบเทียบกับ Random Forest โดยได้แบ่ง DCNN ออกเป็น 3 ชุด คือ DCNN ที่ถูกไม่ได้ Fine-tuned, Fine-tuned DCNN ที่ถูก Pre-trained ด้วย ImageNet1000 และ Fine-tuned DCNN ที่ถูก Pre-trained ด้วย ImageNet ที่รวมกับรูปภาพที่เกี่ยวข้องกับอาหาร ซึ่งได้ผลการทดลองของ RF-based, DCNN, DCNN(fine-tuned), FOOD-DCNN (fine-tuned) ดังนี้ 50.76%, 56.40%, 68.44% และ 70.4 ตามลำดับ

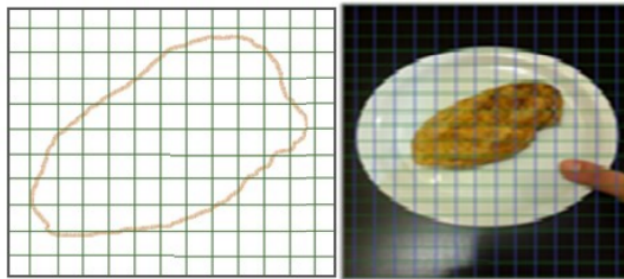
2.7.3 การประมาณแคลอรีจากรูปภาพ

การประมาณแคลอรีจากรูปภาพอาหารได้มีการทดลองโดย Koichi Okamoto และ Keiji Yanai [13] ใช้วิธีการหาขอบเขตของอาหารญี่ปุ่นด้วย K-means แล้วตัดพื้นที่โดยรอบของอาหารด้วย GrabCut ก่อนที่จะนำไปจำแนกว่าเป็นอาหารชนิดใด หลังจากนั้นจึงนำมาเปรียบเทียบขนาดกับข้อมูลขนาดอาหารจริง ซึ่งได้ผลการทดลองได้ average absolute errors เท่ากับ 52.321 kcal และ relative average errors เท่ากับ 21.3%

จากการทดลองของ Yanchao Liang และ Jianhua Li [14] ได้ทำการประมาณแคลอรีจากรูปภาพอาหารมุด้านบน และมุด้านข้าง โดยนำการตรวจจับวัตถุเพื่อแยกอาหารแต่ละชนิดในรูปภาพด้วยวิธี Faster Region with Convolution Neural Network (Faster R-CNN) ซึ่งมีคุณสมบัติในการค้นหาวัตถุย่อยในรูปภาพได้ แล้วทำการหา

ส่วนปริมาตรของอาหารเปรียบกับเหรียญหนึ่งหยวนที่ทราบจำนวนพิกเซลแล้ว เมื่อทราบปริมาตรแล้วจึงหาความสัมพันธ์กับข้อมูลแคลอรีของอาหารที่มีอยู่โดยได้ผลลัพธ์ค่าความคลาดเคลื่อนเฉลี่ย 20%

จากการทดลองของ Parisa และคณะ [15] ประเมินแคลอรีจากรูปภาพอาหารได้ โดยต้องใช้รูปภาพจากทั้งมุมด้านบน และมุมด้านข้าง ประกอบด้วยนิ้วหัวแม่มือซึ่งระบบทราบขนาดของนิ้วหัวแม่มืออยู่ก่อนแล้ว โดยจะสามารถทราบได้ทั้งพื้นที่และความลึกที่แท้จริงของอาหาร การเปรียบเทียบกับนิ้วหัวแม่มือทำได้โดยการสร้างตารางสี่เหลี่ยมบนรูปภาพ เพื่อคำนวณพื้นที่ในแต่ละช่อง แล้วจึงนำพื้นที่ที่ได้เปรียบเทียบกับข้อมูลแคลอรีของอาหารที่กำหนดซึ่งได้ผลการทดลองที่มีค่าแม่นยำเฉลี่ยประมาณ 86%



ภาพที่ 2.6 Methodology for food portion area measurement
(แหล่งที่มา : Measuring Calorie and Nutrition from Food Image [15])

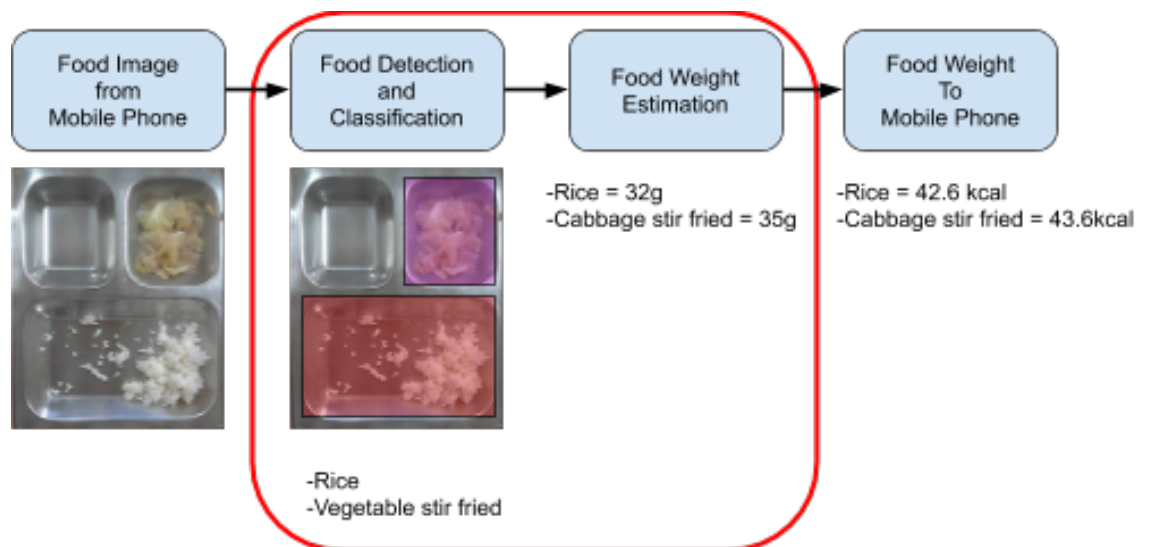
นอกจากนี้การทดลองของ Takumi Ege และ Keiji Yanai [16] ยังได้ทำการนำการเรียนรู้เชิงลึกมาจำแนกหมวดหมู่ของอาหารและประมาณแคลอรีด้วย Multi-task Convolutional Neural Network โดยใช้การ Fine-tuning ด้วย VGG16 ซึ่งเป็น Pre-trained model ที่แข่งขันใน Large Scale Visual Recognition Challenge 2014 ซึ่งแบ่งของการทดลองประมาณแคลอรีออกเป็นวิธี Single-task และ Multi-task ได้ผลความคลาดเคลื่อนของแคลอรีเฉลี่ย คือ 105.73 kcal และ 94.14 kcal

บทที่ 3

วิเคราะห์และออกแบบระบบ

การจำแนกและประมาณแคลอรีของอาหารจากรูปภาพโดยใช้การเรียนรู้ของเครื่อง จำเป็นต้องมีการวิเคราะห์และออกแบบการทำงานของส่วนประกอบดังต่อไปนี้

- 1) การเตรียมชุดข้อมูลรูปภาพอาหาร
- 2) การจำแนกอาหารด้วยรูปภาพ
- 3) การประมาณน้ำหนักของอาหารจากรูปภาพ



ภาพที่ 3.1 ผังการทำงานของการทำงานการจำแนกและประมาณแคลอรีจากรูปภาพ

3.1 การเตรียมชุดข้อมูลรูปภาพอาหาร

รูปภาพของอาหารที่ได้นำมาทดลองนั้นเป็นรูปภาพอาหารผู้ป่วยของโรงพยาบาลมหาราช เชียงใหม่ ซึ่งถ่ายจากกล้องของโทรศัพท์มือถือ โดยถาดมีอัตราส่วนพอดีหรือใกล้เคียงกับขนาดของรูปภาพ อาหารจะอยู่ในถาดอลูมิเนียมที่มีสามหลุมสำหรับใส่อาหาร แต่ละถาดจะมีอาหารไม่เกินสามรายการ และในถาดเดียวกันจะมีอาหารไม่ซ้ำกัน รูปภาพทั้งหมดมี 16067 รูปภาพ โดยแบ่งรายการอาหารออกเป็น 39 รายการ ดังนี้

1. ข้าวสวย
2. แกงส้มผักบั้งหมู
3. ผัดวุ้นเส้น
4. แกงฮังเลหมู
5. ผัดกะหล่ำปลีเต้าหู้
6. แกงจืดแตงกวาหมู
7. ไก่ผัดพริกสด
8. แกงจืดผักกาดขาวหมู
9. ผัดซีอิ้วหมู
10. จอผักกาด
11. ลาบหมูเมือง
12. แกงเขียวหวานไก่
13. ผัดหมูโบราณ
14. หมูผัดพริกขิง
15. แกงจืดวุ้นเส้น
16. หลนปลาเค็ม
17. แกงฟักเขียวไก่
18. ไข่พะโล้
19. ผัดกะเพราหมู
20. แกงจืดเต้าหู้ไข่
21. ตุ่นฟักเขียวไก่
22. น้ำพริกอ่อน
23. พะแนงหมู
24. แกงขนุนหมู
25. ไก่อบ
26. แกงเผ็ดหมู
27. ผัดผักกะหล่ำปลีฝอยหมู
28. หุ้มไก่
29. ผัดกะเพราไก่บดถั่วฝักยาว
30. แกงจืดหัวไชเท้า
31. ต้มส้มขาหมูเห็ด
32. ผัดผักกาดขาวหมูบด
33. แกงอ่อมหมู
34. แกงผักกาดเขียวไก่
35. โจ๊ก
36. ต้มข่าไก่
37. แกงจืดกะหล่ำปลีหมู
38. ข้าวกล้อง
39. ผัดพริกแกงหมู



ภาพที่ 3.2 ตัวอย่างรูปภาพอาหาร

3.2 การจำแนกอาหารจากรูปภาพ

วิธีการจำแนกอาหารจากรูปภาพได้ถูกแบ่งออกเป็นสองวิธีหลักโดยขึ้นอยู่กับวิธีการของการประมาณแคลอรีในขั้นตอนต่อไป รูปภาพจะถูกแบ่งออกเป็นสามชุด คือ Training set และ Test set มีอัตราส่วน 8:2 โดย Training set จะผ่านการทำให้ Data Augmentation เพื่อให้รูปมีความหลากหลายมากยิ่งขึ้นและป้องกันการ Overfitting ของโมเดล โดยวิธีที่ทดลองมีดังต่อไปนี้

การใช้งาน Faster Region with Convolutional Neural Network หรือ Faster R-CNN จะช่วยค้นหวัตถุที่อยู่ในรูปภาพ โดยใช้ ResNet-50 เป็น pre-trained model ทำให้สามารถระบุได้ว่ารูปภาพดังกล่าวมีอาหารรายการใดบ้าง และขอบเขตของอาหารแต่ละรายการอยู่บริเวณใด จึงสามารถนำประโยชน์ของขอบเขตของอาหารที่ระบุอย่างชัดเจนประยุกต์ใช้กับการประมาณแคลอรีของอาหารในขั้นตอนถัดไปได้

3.3 การประมาณแคลอรีของอาหารจากรูปภาพ

จากการที่ได้พบทวนวรรณกรรมทำให้ได้แนวคิดการประมาณแคลอรีของอาหารจากรูปภาพ เนื่องจากวิธีการใช้ปริมาตรหากใช้เพียงแค่นี้รูปจำเป็นต้องระบุคุณสมบัติพื้นฐานของอาหารก่อน ซึ่งอาหารก่อนรับประทานอาจจะสามารถระบุได้ แต่หลังรับประทานอาหารไม่สามารถระบุได้ จึงแบ่งออกเป็นสองวิธีหลัก คือ การคำนวณพื้นที่ของอาหาร และการใช้การเรียนรู้เชิงลึกโดย Convolutional Neural Network หาน้ำหนักของอาหารก่อนรับประทาน และหลังรับประทานเพื่อเปรียบเทียบหาปริมาณแคลอรีจากการเทียบกับแคลอรีอาหารต่อน้ำหนักแต่ละรายการที่ถูกกำหนดไว้ โดยรูปภาพจะถูกแบ่งออกเป็นสามชุด คือ Training set และ Test set มีอัตราส่วน 8:2 มีการทดลองดังต่อไปนี้

การใช้ Convolutional Neural Network จะทำการแบ่งคุณสมบัติต่างๆของรูปภาพอาหารแต่ละประเภทก่อนที่จะระบุว่าอาหารในรูปภาพนั้นมีปริมาณเท่าใด โดย Test set จะผ่านกระบวนการ Data Augmentation เพื่อป้องกันการ Overfitting หลังจากนั้นจะใช้หลักการของ Transfer Learning โดยมี Pre-trained model คือ InceptionResnetV2 [18] ก่อนมาเชื่อมกับ Fully Connected Neural Network แล้วทำการ Fine-tuning ให้เหมาะสมกับรูปภาพอาหาร โดยกำหนดรูปภาพขนาดมีความกว้าง 300 พิกเซล และความยาว 300 พิกเซล

หลังจากได้น้ำหนักของอาหารจากการประมาณแล้ว นำน้ำหนักของอาหารทำนายเทียบบัญญัติไตรยางค์กับแคลอรีต่อน้ำหนักที่ได้ถูกกำหนดไว้แล้วดังตารางที่ 3.1

ตารางที่ 3.1 การเปรียบเทียบแคลอรีกับน้ำหนักของอาหาร

Food Type	Weight (g)	Energy (kcal)	Food Type	Weight (g)	Energy (kcal)	Food Type	Weight (g)	Energy (kcal)	Food Type	Weight (g)	Energy (kcal)
food01	128	134.7	food11	110	210.3	food21	103	150.5	food31	95	180.6
food02	132	497.6	food12	167	115.1	food22	105	102.4	food32	230	478
food03	110	101.1	food13	70	162	food23	105	214.7	food33	120	475.6
food04	100	171.7	food14	115	252.7	food24	103	249.3	food34	130	288.4
food05	135	109.4	food15	160	326.1	food25	103	190.9	food35	118	292
food06	130	114	food16	125	197.8	food26	105	224.6	food36	105	324.9
food07	120	142.3	food17	120	159.8	food27	255	572.5	food37	-	-
food08	100	91.6	food18	75	178.2	food28	165	204.3	food38	-	-
food09	125	187.4	food19	175	163.1	food29	170	211.9	food39	-	-
food10	132	428.8	food20	110	266.1	food30	170	361.6			

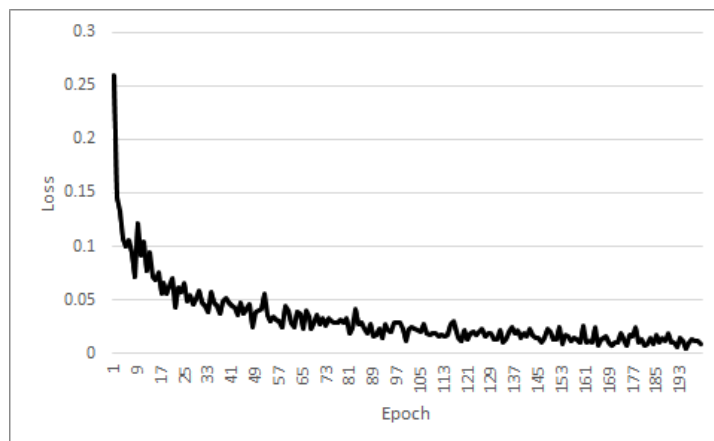
บทที่ 4

การทดสอบระบบ

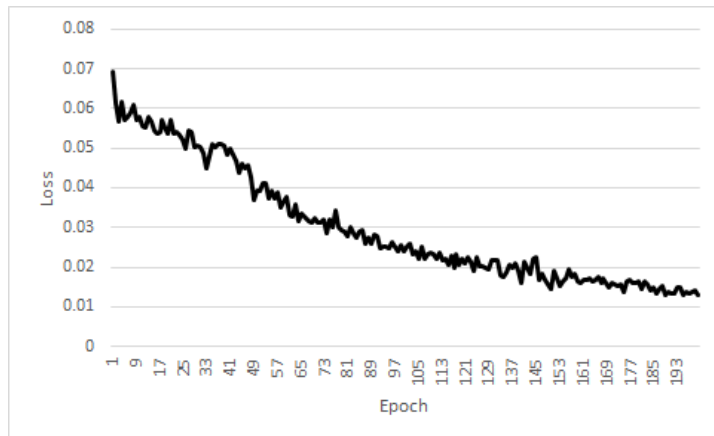
ในการทดสอบระบบผู้วิจัยได้แบ่งการทดสอบออกเป็นสองส่วน คือ การทดสอบการจำแนกอาหารจากรูปภาพ และการประมาณน้ำหนักอาหารจากรูปภาพ ซึ่งจะมีผลการทดสอบตั้งแต่การเรียนรู้ของโมเดลและการวัดประสิทธิภาพของโมเดล โดยจะทดสอบบนชุดรูปภาพอาหารทดสอบจำนวน 4,017 รูปภาพ

4.1 ผลการทดสอบระบบการจำแนกอาหารจากรูปภาพ

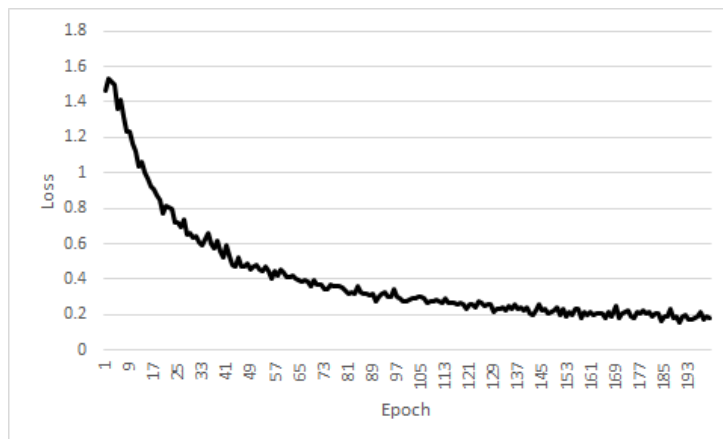
เนื่องจากการใช้ Faster R-CNN จึงต้องมีสอง Loss Function สำหรับ Region Proposal Network (RPN) model และ Classifier Model ซึ่ง RPN model และ Classifier Model มีข้อมูลส่งออกสองค่า คือ การจำแนกวัตถุและการหาขอบเขตของวัตถุ



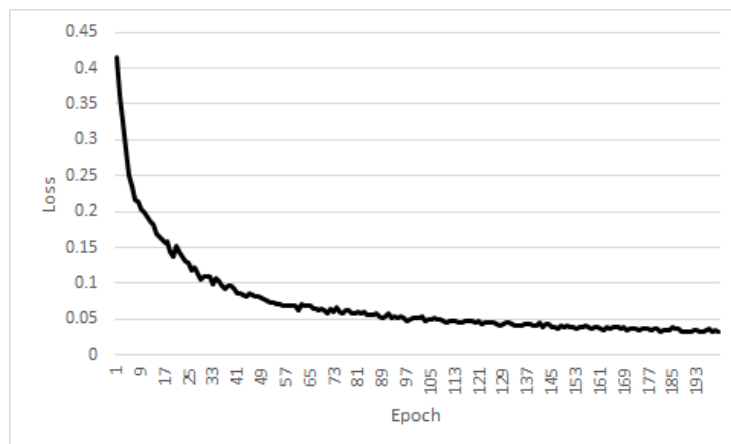
ภาพที่ 4.1 การเปรียบเทียบค่าของ Loss และ Epochs ของการจำแนกวัตถุใน RPN



ภาพที่ 4.2 การเปรียบเทียบค่าของ Loss และ Epochs ของการหาขอบเขตของวัตถุใน RPN



ภาพที่ 4.3 การเปรียบเทียบค่าของ Loss และ Epochs ของการจำแนกวัตถุใน Classifier



ภาพที่ 4.4 การเปรียบเทียบค่าของ Loss และ Epochs
ของการหาขอบเขตของวัตถุใน Classifier

ภาพที่ 4.1 และ 4.2 แสดงให้เห็นว่าค่า Loss ของการจำแนกวัตถุมีค่าสูงในช่วงแรกของการเรียนรู้ ดังนั้นจะเห็นได้ว่าจำเป็นต้องเวลามากขึ้นในการเรียนรู้ โดยได้กำหนดการเรียนรู้ไว้ที่ 200 ครั้ง เช่นเดียวกับภาพที่ 4.3 และ 4.4 ซึ่งทำให้เราสามารถเปรียบเทียบการจำแนกและหาขอบเขตได้อย่างชัดเจน

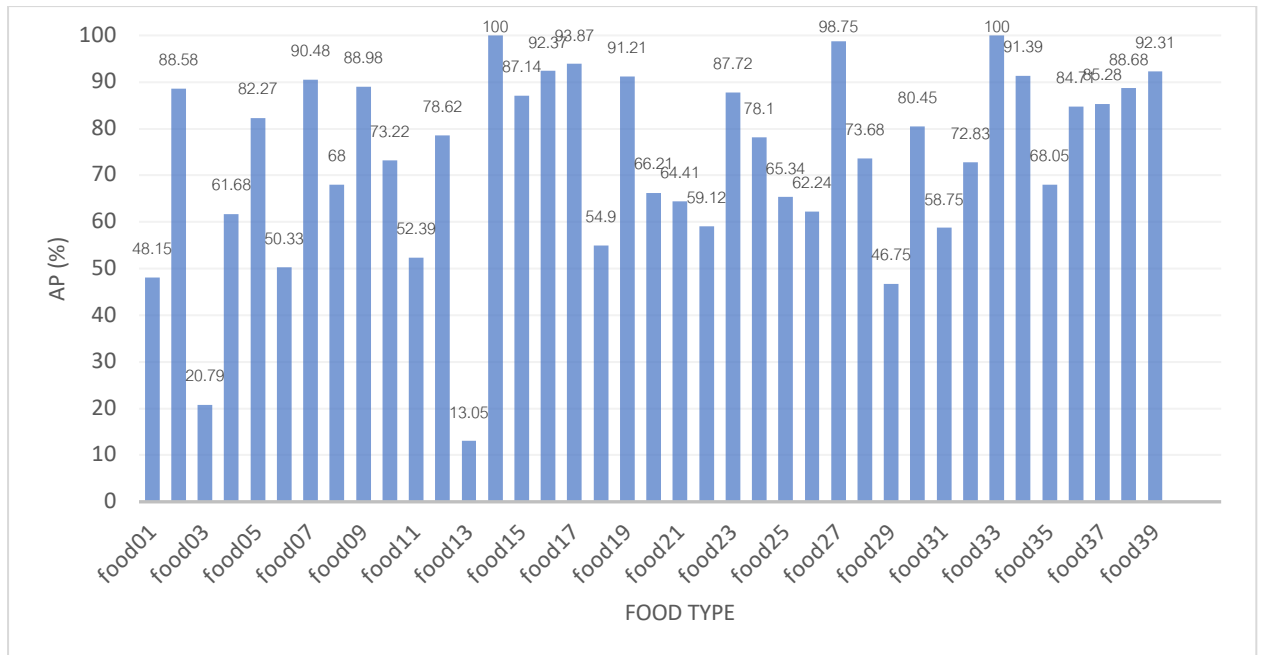
สำหรับการประเมินความมีประสิทธิภาพของการตรวจจับและจำแนก เราได้ทดสอบกับรูปภาพอาหาร 4,017 รูป ได้ใช้ค่า Mean Average Precision (mAP) เป็นเครื่องมือในการวัด เพราะว่าเป็นการวัดผลการทำนายที่ถูกต้องและพื้นที่ที่ทำนายกับพื้นที่ของวัตถุจริง โดยอันดับแรกได้คำนวณค่า Average Precision (AP) ของอาหารแต่ละรายการ หลังจากนั้นจึงนำมาคำนวณหาค่า mAP ซึ่งการคำนวณหา AP จำเป็นต้องทราบค่าของ Intersect over Union (IoU) ก่อน

$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$

โดยกำหนด Threshold ให้ไว้ถ้าหาก IoU มีค่ามากกว่า 0.5 และอาหารที่ถูกทำนายตรงกับอาหารจริง จะกำหนดให้เป็น True Positive นอกจากนั้นเป็น False Positive หลังจากนั้นจึงหาพื้นที่ใต้กราฟของ Precision และ Recall เพื่อหาค่า AP หลังจากนั้นได้คำนวณค่าของ Mean Average Precision (mAP) ด้วยสมการดังนี้

$$mAP = \frac{1}{|class|} \sum_{c \in class} AP(c)$$

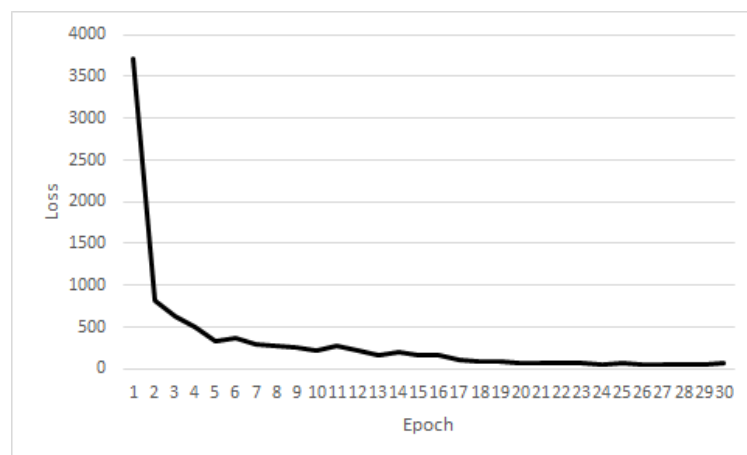
ซึ่งจากการทดลองได้ค่า mAP เท่ากับ 73.354 แม้ว่าผลลัพธ์ mAP จะมีค่าอยู่ระดับปานกลาง แต่พบว่าครึ่งหนึ่งของอาหารมี AP สูงกว่า 75 และ อาหาร 15 รายการ จาก 39 รายการ มี AP สูงกว่า 85 เนื่องจากการวิจัยนี้เป็นระบบต้นแบบ ซึ่งเราคาดหวังว่าจะเพิ่มภาพอาหารในข้อมูลเพิ่มเติมเพื่อให้ mAP มีค่าสูงมากขึ้น



ภาพที่ 4.5 ค่า AP ของอาหารแต่ละรายการ

4.2 ผลการทดสอบระบบการประมาณน้ำหนักอาหารจากรูปภาพ

เราประมาณน้ำหนักของอาหารโดยใช้การทำนายการถดถอยดังนั้นเราจึงใช้ Mean Squared Error เป็น Loss Function ที่จะคำนวณค่าเฉลี่ยกำลังสองระหว่างน้ำหนักที่ทำนายและน้ำหนักจริง



ภาพที่ 4.6 การเปรียบเทียบค่าของ Loss และ Epochs ในการประมาณน้ำหนัก

การประเมินประสิทธิภาพของการประมาณน้ำหนักอาหาร ได้ใช้รูปภาพทดสอบเหมือนกับการจำแนกอาหารโดยมีรูปภาพอาหาร 4,017 รูป และได้ใช้ค่า Mean Absolute Percentage Error (MAPE) ในการประเมินการประมาณน้ำหนักอาหาร ซึ่ง MAPE สามารถคำนวณได้จากสมการดังนี้

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

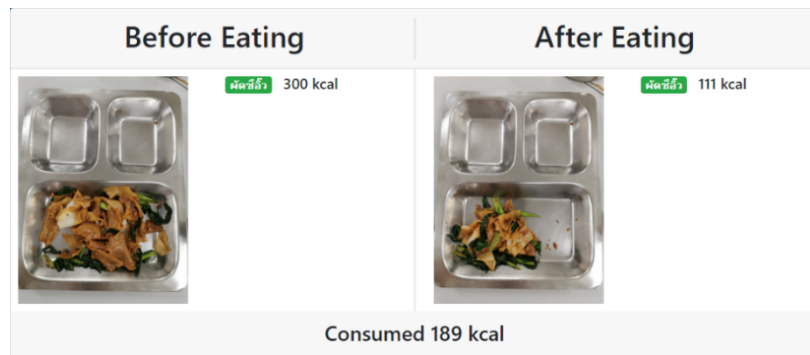
โดย A_t คือ ค่าจริง และ F_t คือ ค่าที่ถูกทำนาย

ซึ่งจากผลการทดลองกับข้อมูลทดสอบ จะเห็นได้ว่าอาหารบางรายการมีค่า MAPE ที่สูง เพราะว่ามีจำนวนรูปน้อย ยกตัวอย่างเช่น FOOD39 มีรูปภาพเพียงแค่ 13 รูปเท่านั้นที่ถูกทดสอบ

ตารางที่ 4.1 ค่า Mean Absolute Percentage Error (MPAE) ของอาหารแต่ละรายการ

Food Type	MAPE	Food Type	MAPE	Food Type	MAPE	Food Type	MAPE
food01	13.5454	food11	21.5203	food21	15.4855	food31	12.5751
food02	16.5453	food12	32.8511	food22	16.4265	food32	12.4081
food03	16.9564	food13	11.1085	food23	15.4655	food33	23.2571
food04	12.3263	food14	20.2444	food24	18.6577	food34	11.2876
food05	15.0029	food15	11.1640	food25	15.6046	food35	14.6511
food06	11.4097	food16	15.6743	food26	25.6277	food36	19.3801
food07	18.0382	food17	16.6256	food27	16.9343	food37	17.3686
food08	6.5155	food18	16.8279	food28	24.6898	food38	16.6468
food09	12.1952	food19	10.1909	food29	14.4526	food39	41.8455
food10	14.1468	food20	13.6431	food30	22.6465		

นอกจากนี้เรายังคาดหวังให้การจำแนกอาหารและการประมาณแคลอรีสามารถใช้งานบนแอปพลิเคชันที่สามารถใช้งานและแสดงผลได้ง่ายดังตัวอย่างภาพที่ 4.7 เพื่อความสะดวกและลดระยะเวลาการทำงานของนักโภชนาการ หรือบุคคลที่เกี่ยวข้อง



ภาพที่ 4.7 ตัวอย่างการแสดงผลลัพธ์การจำแนกอาหารและปริมาณแคลอรีบนแอปพลิเคชัน

บทที่ 5

บทสรุป

5.1 สรุปผลการดำเนินงาน

การประมาณน้ำหนักอาหารเป็นปัญหาที่สำคัญ เนื่องจากการขาดแคลนนักโภชนาการและกระบวนการประมาณอาหารใช้เวลานาน ซึ่งได้มีนักวิจัยหลายคนสนใจที่จะแก้ปัญหานี้ ในงานวิจัยนี้ เราจึงได้นำเสนอระบบการประมาณปริมาณอาหารสำหรับผู้ป่วยในโรงพยาบาล โดยที่การทำงานของระบบของเราคือการถ่ายรูปอาหารก่อนและหลังรับประทานอาหาร ระบบจะนำรูปภาพก่อนและหลังรับประทานอาหารไปประมวลผลเพื่อหาน้ำหนัก หลังจากนั้นเราจะนำเอาน้ำหนักที่ได้มาเทียบกับตารางอ้างอิงเพื่อให้ได้ปริมาณแคลอรีของอาหาร ทำให้เราสามารถหาปริมาณแคลอรีที่ผู้ป่วยรับประทานอาหารในมื้อนั้นๆ ได้

5.2 ข้อเสนอแนะสำหรับการพัฒนาในอนาคต

สำหรับงานในอนาคตเราจะขยายการทำงานสำหรับการประมาณแคลอรี นอกจากนี้ยังวางแผนที่จะขยายรายการอาหารที่จะใช้ในการประมาณ ในช่วงเริ่มต้นจะเพิ่มอาหารที่อยู่ในมหาวิทยาลัย นอกจากนี้เมื่อเราเพิ่มรูปภาพใหม่ จำเป็นจะต้องเรียนรู้ใหม่ทั้งหมดซึ่งใช้เวลานาน เราจึงกำลังหาวิธีใหม่ที่ไม่จำเป็นต้องใช้เวลาในการเรียนรู้ข้อมูลใหม่ทั้งหมด

เอกสารอ้างอิง

- [1] Yuzhen Lu, “Food image recognition by using convolutional neural networks (CNNs)”, arXiv:1612.00983, December 2016.
- [2] Alan Lapedes and Robert Farber, “How Neural Nets Work”, Technical Report LA-UR-88-418, Los Alamos, NM: Los Alamos National Laboratory, 1988.
- [3] Luis Perez and Jason Wang, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”, arXiv:1712.04621, December 2017.
- [4] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, arXiv: 1506.01497, January 2015.
- [5] Sinno Jialin Pan and Qiang Yang, “A Survey on Transfer Learning”, IEEE Transactions on Knowledge and Data Engineering, Volume 22, Issue 10, pp. 1345-1359, October 2009.
- [6] Chollet. François and others, “Keras”, 2015. [Online] . Available: <https://www.keras.io>
- [7] Manika Puri, Zhiwei Zhu, Qian Yu, Ajay Divakaran and Harpreet Sawhney, “Recognition and volume estimation of food intake using a mobile device”, 2009 Workshop on Applications of Computer Vision (WACV), Snowbird, UT, USA, pp. 1-8, December 2009.
- [8] Joachim Dehais, Marios Anthimopoulos, Sergey Sevchik and Stavroula Mougiakakou, “Two-view 3D Reconstruction for Food Volume Estimation”, IEEE Transactions on Multimedia, Volume 19, Issue 5, pp. 1090-1099, May 2017.
- [9] Chang Xu, Ye He, Nitin Khanna, Albert Parra, Carol Boushey and Edward Delp, “Image- Based Food Volume Estimation”, 5th International Workshop on Multimedia for Cooking & Eating Activities, pp. 75-80, October 2013.
- [10] Duc Thanh Nguyen, Zhimin Zong, Philip O. Ogunbona, Yasmine Probst and Wanqing Li, “Food Image Classification using Local Appearance and Global Structural Information”, Neurocomputing, Volume 140, pp. 242-251, September 2014.

- [11] Gozde Ozsert Yigit and B. Melis Ozyildirim, “ Comparison of Convolutional Neural Network Models for Food Image Classification” , Journal of Information and Telecommunication, Volume 2, Issue 3, pp. 347-357, February 2018.
- [12] Keiji Yanai and Yoshiyuki Kawano, “ Food Image Recognition using Deep Convolutional Network with Pre- training and Fine- tuning” , 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, pp. 1-6, July 2015.
- [13] Koichi Okamoto and Keiji Yanai, “ An Automatic Calorie Estimation System of Food Images on a Smartphone” , 2nd International Workshop on Multimedia Assisted Dietary Management, pp. 63-70, October 2016.
- [14] Yanchao Liang and Jianhua Li, “Computer Vision-Based Food Calorie Estimation: Dataset, Method, and Experiment” , arXiv preprint arXiv:1705.07632, May 2017.
- [15] Parisa Pouladzadeh, Sgervin Shirmohammadi and Rana Al-Maghrabi, “Measuring Calorie and Nutrition from Food Image” , IEEE Transactions on Instrumentation and Measurement, Volume 63, Issue 8, pp. 1947-1956, August 2014.
- [16] Takumi Ege and Keiji Yanai, “ Simultaneous Estimation of Food Categories and Calories with Multi-task CNN” , 2018 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, May 2017.
- [17] Carsten Rother, Vladimir Kolmogorov and Andrew Blake, “GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts” , CAN Transactions on Graphics (TOG), Volume 23, Issue 3, pp. 309-314, August 2004.
- [18] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke and Alexander A. Alemi, “ Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning” , 31st AAAI Conference on Artificial Intelligence, San Francisco, California, USA, February 2017.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition” , arXiv preprint arXiv:1512.03385, December 2015.