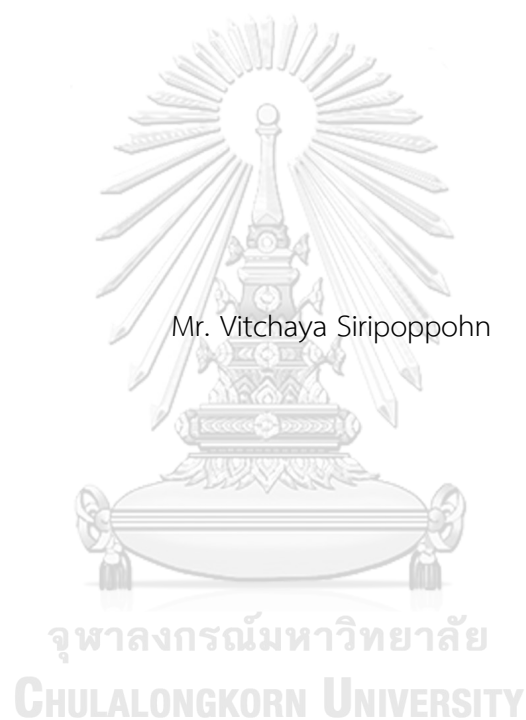


การจำแนกรูปภาพเชิงความหมายของภาพตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิตปกติในกระเพาะอาหาร  
ตามเวลาจริงโดยใช้กระบวนการการเรียนรู้เชิงลึก



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2563  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Real-Time Gastric Intestinal Metaplasia Semantic Segmentation Using Deep Learning  
Approach



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science in Computer Science  
Department of Computer Engineering  
FACULTY OF ENGINEERING  
Chulalongkorn University  
Academic Year 2020  
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

โดย

นายวิษยะ ศิริพพร

สาขาวิชา

วิทยาศาสตร์คอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง  
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

..... คณบดีคณะวิศวกรรมศาสตร์  
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

..... ประธานกรรมการ  
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกุล)

..... กรรมการ  
(อาจารย์ ดร.ดวงดาว วิชิตกุล)

..... กรรมการภายนอกมหาวิทยาลัย  
(อาจารย์ ดร.ธนภัทร ช้างคะจิตร)

วิษยะ ศิริพพบพร : การจำแนกรูปภาพเชิงความหมายของภาพตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหารตามเวลาจริงโดยใช้กระบวนการการเรียนรู้เชิงลึก. (Real-Time Gastric Intestinal Metaplasia Semantic Segmentation Using Deep Learning Approach) อ.ที่ปรึกษาหลัก : ผศ. ดร.พีรพล เวทีกุล

ตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหารจัดอยู่ในประเภทรอยโรคชนิดหนึ่ง เนื่องจากรอยโรคชนิดนี้ตรวจพบได้ยาก ทำให้บ่อยครั้งที่แพทย์มักจะตรวจไม่พบ และมีโอกาสสูงที่จะพัฒนากลายเป็นมะเร็งกระเพาะอาหาร ในปัจจุบัน กระบวนการการเรียนรู้เชิงลึกนั้นไม่สามารถตรวจจับบริเวณที่เป็นตามเวลาจริงได้ ทำให้งานวิจัยส่วนใหญ่จะตรวจหลังจากทำหัตถการทางผู้จัดทำ จึงเสนอแนวทางในการทำโมเดลใหม่ โดยเน้นไปที่การใช้งานตามเวลาจริง โดยนำภาพถ่ายรอยโรคความละเอียดสูง 802 ภาพ จากศูนย์ส่องกล้องโรงพยาบาลจุฬาลงกรณ์ มาทำการปรับปรุงโมเดล BiSeNet จากงานแข่งขัน โดยเพิ่มเทคนิคเพื่อช่วยเพิ่มความแม่นยำของโมเดล โดยการใช้การเรียนรู้แบบโอนถ่ายจากภาพการส่องกล้องทางเดินอาหารส่วนล่าง ใช้การปรับภาพเคลซเพื่อช่วยเพิ่มรายละเอียดของภาพ และใช้การเพิ่มข้อมูลเพื่อให้โมเดลมีความแม่นยำโดยที่มีภาพจำนวนน้อย โดยโมเดลที่ถูกปรับปรุงของผู้จัดทำนั้น สามารถรองรับการใช้งานจริงได้ โดยมีการประมวลผลอยู่ที่ 31.53 เฟรมต่อวินาที และสามารถทำนายภาพที่มีรอยโรคได้แม่นยำถึงร้อยละ 93 ดังนั้น โมเดลของผู้จัดทำ จึงสามารถใช้งานได้ระหว่างการทำการหัตถการ และสามารถทำนายรอยโรคได้แม่นยำใกล้เคียงกับโมเดลยอดนิยม ในตลาดปัจจุบัน

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์  
ปีการศึกษา 2563

ลายมือชื่อนิสิต .....  
ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

# # 6270261321 : MAJOR COMPUTER SCIENCE

KEYWORD: Artificial intelligent, Deep learning model,  
Esophagogastroduodenoscopy

Vitchaya Siripoppohn : Real-Time Gastric Intestinal Metaplasia Semantic Segmentation Using Deep Learning Approach. Advisor: Asst. Prof. PEERAPON VATEEKUL

Gastric intestinal metaplasia (GIM) is a premalignant lesion that is difficult to detect and has a high chance to evolve to gastric cancer diseases. Now, deep learning approach failed to detect GIM lesion in real-time due to slow inference speed. Then, most of the paper in GIM is focus on post-surgery. We proposed the new model adapted from real-time competition "BiSeNet" trained by 802 GIM images and its label from Chulalongkorn Hospital. With 3 techniques, transfer learning from lower gastrointestinal tract image, CLAHE pre-processing, and data augmentation, the model can perform real-time environment with 31.53 frames per second and can predict with 93% highest sensitivity. Thus, our BiSeNet model can perform in real-time with high accuracy equivalent to the baseline model in the market.



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

Field of Study: Computer Science

Student's Signature .....

Academic Year: 2020

Advisor's Signature .....

## กิตติกรรมประกาศ

งานวิจัยและวิทยานิพนธ์ฉบับนี้อาจไม่สำเร็จผลได้ด้วยตัวผู้วิจัยเพียงคนเดียว หากปราศจากผู้ครอบงำที่คอยชี้แนะ และคอยสนับสนุนจนจบโครงการ

ขอขอบคุณอาจารย์ ผศ.ดร. พีรพล เวทีกุล อาจารย์ที่ปรึกษาของวิทยานิพนธ์ฉบับนี้ ที่คอยสั่งสอน คอยชี้แนะทุกกรณีที่มีเรื่องผิดพลาดขึ้น รวมถึงให้โอกาสหาหนทางใหม่ ๆ ในการทำวิจัย ถ้าไม่มีอาจารย์คอยสนับสนุน ตัวผู้วิจัยอาจเดินทางมาถึงจุดเริ่มต้นโครงการ และจุดสิ้นสุดโครงการไม่ได้

ขอขอบคุณอาจารย์ รศ.พญ. รภัส พิทยานนท์ และ พญ. เกศินี เขียรกานนท์ ผู้ร่วมวิจัยในโครงการ ที่คอยให้ความช่วยเหลือในฝั่งแพทย์ และคอยแก้ไขอุปสรรคงานวิจัย ตลอดจนจบเป็นผลงานวิจัย และวิทยานิพนธ์ชิ้นนี้

ขอขอบคุณอาจารย์ นพ. นที พิภพาค, ผศ.นพ. อนพัช สรรพาวัตร, รศ.พญ. นฤมล คล้ายแก้ว และแพทย์ทุก ๆ ท่าน ที่อยู่เบื้องหลังความสำเร็จให้ผู้วิจัยได้มีข้อมูลในการนำไปทำวิจัยจนจบโครงการ

ขอขอบคุณอาจารย์ ศ.นพ.รังสรรค์ ฤกษ์นิมิตร ที่คอยอำนวยความสะดวก และมอบโอกาสให้ผู้วิจัยได้ทำการวิจัยโครงการนี้จนสำเร็จผล

ขอขอบคุณอาจารย์ ศ.ดร. บุญเสริม กิจสิริกุล และคณะกรรมการสอบจบงานวิจัยทุกท่าน ที่คอยตรวจสอบคุณภาพของผลงานตลอดโครงการ

ขอขอบคุณน้องแทน นาย ภาณุกร สุนทรเวชพงษ์ ผู้ร่วมวิจัยในอีกโครงการที่ทำร่วมกันกับศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์ ที่ช่วยผลักดัน สานต่องานวิจัยชิ้นนี้ให้สำเร็จผล

ขอขอบคุณพี่แก้ว ดร. ธีรพงศ์ ปานบุญยืน และพี่ ๆ น้อง ๆ ในแลป Datamind ทุกท่าน ที่คอยให้ความช่วยเหลือในทุก ๆ เรื่อง ตั้งแต่เข้ามาหาวิทยาลัยเริ่มศึกษาศาสตร์วิชา จนถึงสิ้นสุดโครงการการวิจัย

ขอขอบคุณศูนย์ความเป็นเลิศทางการแพทย์ด้านการส่องกล้องระบบทางเดินอาหาร และ ฝ่ายพยาธิวิทยา โรงพยาบาลจุฬาลงกรณ์ ที่คอยอำนวยความสะดวกในการยื่นขอรหัสธรรม และอนุญาตให้ผู้วิจัยสามารถทำวิจัยได้อย่างสมบูรณ์

ขอขอบคุณคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่คอยช่วยเหลือ คอยอำนวยความสะดวกเรื่องเอกสาร รวมไปถึงเรื่องการติดต่อในระยะของโครงการ

ขอขอบคุณเพื่อนแจม เพื่อนโพน เพื่อนพี่ซุง และเพื่อนร่วมรุ่นทุกคน ชีวิตในรั้วมหาลัยจะไม่มีสีสันเลยถ้าขาดเพื่อน ๆ เหล่านี้

ขอขอบคุณน้ากล้า นาย อรรถพล เสนาะสุทธิพันธุ์ และตระกูล ที่สนับสนุนทุนทรัพย์ของผู้วิจัยตลอดตั้งแต่เข้าเรียนมหาวิทยาลัย

ขอขอบคุณตระกูล ศิริพพร ครอบครัวของผู้วิจัย ที่อยู่เบื้องหลังทุกความสำเร็จ ไม่ว่าจะเป็นเรื่องใด ๆ ก็ตาม

และสุดท้าย ขอขอบคุณคุณแม่ ความหมายของทั้งชีวิต

วิษยะ ศิริพพร



## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฎ
บทที่ 1 .....	1
1.1 ที่มาและความสำคัญของปัญหา .....	1
1.2 วัตถุประสงค์ของงานวิจัย .....	3
1.3 ขอบเขตของงานวิจัย.....	3
1.4 ประโยชน์ที่ได้รับ.....	3
1.5 วิธีดำเนินวิจัย.....	4
บทที่ 2 .....	3
2.1 โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network).....	3
2.2 การจำแนกรูปภาพเชิงความหมาย (Semantic Segmentation).....	5
2.3 รูปลาร์จี้ปี (RGB Image).....	6
2.4 การเพิ่มข้อมูล (Data Augmentation).....	7
2.5 การเรียนรู้แบบโอนถ่าย (Transfer learning).....	7
2.6 การเรียนรู้ของเครื่อง (Machine learning) .....	8
2.7 การประเมินประสิทธิภาพ .....	8
2.8 การปรับค่าปกติ (Normalization) .....	10



2.9 การเพิ่มประสิทธิภาพโมเดลในแง่ของความเร็ว.....	11
บทที่ 3 .....	13
3.1 โมเดลสถาปัตยกรรมรูปตัวยู (U-Net architecture) .....	13
3.2 โมเดลสถาปัตยกรรมตีปแลป (DeepLab) .....	15
3.2.1 สถาปัตยกรรมตีปแลปวีหนึ่ง (DeeplabV1).....	16
3.2.2 สถาปัตยกรรมตีปแลปวีสอง (DeeplabV2).....	16
3.2.3 สถาปัตยกรรมตีปแลปวีสามพลัส (DeeplabV3+).....	17
3.2 โมเดลการจำแนกรูปภาพเชิงความหมายตามเวลาจริง ที่ได้รับความนิยมในข้อมูลภาพถ่ายท้อง ถนนแบบเปิด.....	18
3.2.1 โมเดลสถาปัตยกรรมการจำแนกรูปภาพเชิงความหมายแบบเครือข่ายทวิภาคี (Bilateral Segmentation Network Architecture, BiSeNet) [18].....	18
3.3 การปรับภาพแคลช (CLAHE).....	21
บทที่ 4 .....	25
4.1 การเตรียมข้อมูล (Data Preparation) .....	25
4.1.1 คลังข้อมูลคควาเซอร์ (Kvasir Dataset).....	25
4.1.2 คลังข้อมูลซีวีซี (CVC-Colon Dataset).....	25
4.1.3 คลังข้อมูลศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย .....	26
4.1.4 การกำหนดชุดตัวอย่างสอน .....	28
4.2 การประมวลผลก่อน (Data Preprocessing).....	28
4.3 การปรับปรุงโมเดล (Model Implementation).....	28
4.4 การวัดประสิทธิภาพโมเดล (Model Evaluation) .....	29
4.5 การทดสอบโมเดล (Model Deployment).....	29
บทที่ 5 .....	30
บทที่ 6 .....	37

บรรณานุกรม..... 39

ประวัติผู้เขียน..... 45



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## สารบัญตาราง

	หน้า
ตารางที่ 1 ผลการทดลองในแต่ละชุดการทดลอง.....	33
ตารางที่ 2 ค่าเฉลี่ยของผลการทดลองของทั้งสามชุดการทดลอง .....	34
ตารางที่ 3 ค่าความมั่นใจ (Confidence Interval) ที่ร้อยละ 0.95 ของผลการทดลองทั้งสามชุดการทดลอง .....	34



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

## สารบัญภาพ

	หน้า
ภาพที่ 1 ภาพจากการทำหัตถการการส่องกล้องทางเดินอาหารส่วนบนโดยใช้แสงขาว .....	2
ภาพที่ 2 ภาพรวมของโครงข่ายประสาทแบบคอนโวลูชัน (อ้างอิงจาก fig2, [23]).....	3
ภาพที่ 3 ตัวอย่างการขยายความ (อ้างอิงจาก Cousera course, [24]).....	4
ภาพที่ 4 ตัวอย่างการทำคอนโวลูชัน (อ้างอิงจาก Cousera course, [24]).....	5
ภาพที่ 5 ภาพรวมของโครงข่ายประสาทแบบคอนโวลูชัน ที่ทำหน้าที่ในการจำแนกรูปภาพเชิง ความหมาย.....	6
ภาพที่ 6 การรวมสัญญาณภาพอาร์จีบี .....	6
ภาพที่ 7 ตัวอย่างการเพิ่มข้อมูลภาพรอยโรค.....	7
ภาพที่ 8 ตัวอย่างกระบวนการ การเรียนรู้แบบโอนถ่าย.....	8
ภาพที่ 9 ตัวอย่างการคำนวณ Intersect over Union .....	10
ภาพที่ 10 ขั้นตอนการทำงานของโอนิกซ์ (อ้างอิงจาก Nvidia.com/tensorrt).....	11
ภาพที่ 11 ขั้นตอนการทำงานของเทนเซอร์อาที (อ้างอิงจาก Nvidia.com/tensorrt).....	12
ภาพที่ 12 สถาปัตยกรรมรูปตัวยู (อ้างอิงจาก fig 1, [12]) .....	14
ภาพที่ 13 การทำคอนโวลูชันแบบหลุม (อ้างอิงจาก fig 6, [30]).....	15
ภาพที่ 14 ขั้นตอนการทำงานของตีปแลปวีหนึ่ง (อ้างอิงจาก fig 1, [31]).....	16
ภาพที่ 15 การการรวมเชิงพื้นที่รูปกรวยเหลี่ยมหลุม (อ้างอิงจาก fig 4, [31]).....	17
ภาพที่ 16 โมเดลไบร์เน็ต (อ้างอิงจาก fig 2, [18]).....	19
ภาพที่ 17 ผลลัพธ์จากการทดลองโมเดลไบร์เน็ตบนข้อมูลท้องถนน (อ้างอิงจาก fig 3, [18]).....	20
ภาพที่ 18 ตัวอย่างการคำนวณหาค่าสมดุลฮิสโทแกรมของภาพขนาด 4x4 จุดสี .....	22
ภาพที่ 19 ผลลัพธ์ที่ได้จากการทำสมดุลฮิสโทแกรมของรูปตัวอย่าง .....	22

ภาพที่ 20 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการสมมูลฮิสโทแกรม.....	22
ภาพที่ 21 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการสมมูลฮิสโทแกรมแบบยืดหยุ่น .....	23
ภาพที่ 22 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการแคลช .....	23
ภาพที่ 23 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกตัดแปลงไปเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพที่ผ่านการทำสมมูลฮิสโทแกรมแบบยืดหยุ่น ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุงโดยวิธีการแคลช .....	23
ภาพที่ 24 กระบวนการในการทำแคลชบนภาพสี โดยทำผ่านช่องสัญญาณ RGB .....	24
ภาพที่ 25 ขั้นตอนการดำเนินงาน.....	25
ภาพที่ 26 ตัวอย่างภาพจากคลังข้อมูลคหะวาเซอร์ (ก) ภาพต้นฉบับ (ข) ผลเฉลี่ย โดยพื้นที่สีขาวจะแทนบริเวณที่เป็นติ่งเนื้อ (ค) ผลเฉลี่ย โดยบริเวณที่ถูกตีกรอบจะเป็นบริเวณที่เป็นติ่งเนื้อ (อ้างอิงจาก [21]).....	26
ภาพที่ 27 ตัวอย่างภาพจากคลังข้อมูลศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย (ก) ภาพต้นฉบับ (ข) ผลเฉลี่ยที่ได้จากการใช้เครื่องมือเลเบลมี.....	26
ภาพที่ 28 ขั้นตอนการเตรียมข้อมูลผ่านเครื่องมือ LabelMe.....	27
ภาพที่ 29 ขั้นตอนการครอบตัดบริเวณที่ไม่สนใจ (ก) ภาพต้นฉบับและผลเฉลี่ย (ข) ภาพต้นฉบับและผลเฉลี่ย หลังจากถูกครอบตัดแล้ว .....	27
ภาพที่ 30 ภาพรวมกระบวนการของโมเดล โดยตัวเลขจะแทนเทคนิคต่าง ๆ ที่ผู้จัดทำใช้ 1) การเรียนรู้แบบโอนถ่าย 2) การปรับภาพโดยแคลช 3) การเพิ่มข้อมูล และ 4) โมเดล BiSeNet.....	29
ภาพที่ 31 ตัวอย่างผลลัพธ์จากการทดสอบตัวอย่างภาพที่มีรอยโรค บนคลังทดสอบ (ก) ภาพต้นฉบับ (ข) ผลเฉลี่ย (ค) ผลลัพธ์จากโมเดล U-Net (ง) ผลลัพธ์จากโมเดล DeepLabV3 (จ) ผลลัพธ์จากโมเดล BiSeNet (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (ช) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE (ซ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning +	

CLAHE + Augmentation (ฅ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation + Hard negative sample ..... 35

ภาพที่ 32 ตัวอย่างผลลัพธ์จากการทดสอบตัวอย่างภาพที่ไม่มีรอยโรค บนคลังทดสอบ (ก) ภาพต้นฉบับ (ข) ผลเฉลย (ค) ผลลัพธ์จากโมเดล U-Net (ง) ผลลัพธ์จากโมเดล DeepLabV3 (จ) ผลลัพธ์จากโมเดล BiSeNet (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (ช) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE (ซ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation (ฅ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation + Hard negative sample ..... 36



## บทที่ 1

### บทนำ

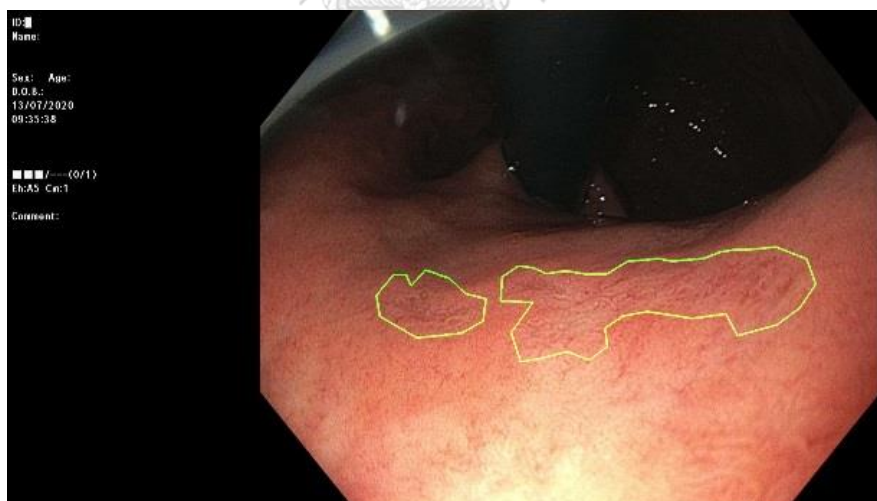
#### 1.1 ที่มาและความสำคัญของปัญหา

ติ่งเนื้อชนิดเซลล์แปรตัวแบบผิดปกติในกระเพาะอาหาร (Gastric intestinal metaplasia, GIM) คือการเปลี่ยนแปลงลักษณะของเยื่อบุกระเพาะอาหาร ไปเป็นลักษณะของเยื่อบุลำไส้เล็ก ถือว่าเป็นรอยโรคก่อนมะเร็งชนิดหนึ่ง (Premalignant lesion) ที่มีความเสี่ยงในการเกิดมะเร็งในกระเพาะอาหาร [1] ในการทำหัตถการส่องกล้อง การตรวจจ็บรอยโรคชนิดนี้จะแตกต่างจากการตรวจจ็บติ่งเนื้อในลำไส้ เนื่องจากรอยโรคจะมีความไม่ชัดเจนของขอบเขตที่เป็น ทำให้หลุดรอดจากการตรวจจ็บได้ง่ายกว่าเมื่อเปรียบเทียบกับก้อนติ่งเนื้อที่มีรูปร่างชัดเจน การทำหัตถการการโดยใช้กล้องแสงขาว (White-light endoscopy, WLE) เพียงอย่างเดียวอาจทำให้แพทย์ฝึกหัดนั้นพลาดในการตรวจจ็บบริเวณที่เกิดโรคได้ [2-4] ในส่วนของเทคนิคด้านการส่องกล้อง ได้มีการประยุกต์วิธีการต่าง ๆ เพื่อเพิ่มความสามารถในการตรวจจ็บรอยโรค เช่น โพรโตคอลการตรวจชิ้นเนื้อแบบสุ่มหลายตัว (Multiple random biopsy protocol, Sydney protocol) [5] และการปรับปรุงภาพส่องกล้องวิธีต่าง ๆ (Image-enhanced endoscopy, IEE) ตัวอย่างเช่น โหมดแสงมืด (Narrow-band image, NBI), การปรับปรุงสีของภาพตามสเปกตรัมยืดหยุ่น (Flexible spectral imaging color enhancement, FICE), I-SCAN และการถ่ายภาพด้วยคอนโฟคอลเลเซอร์ (Confocal laser imaging) [6-8] โดยทั่วไป การทำโพรโตคอลการตรวจชิ้นเนื้อแบบสุ่มหลายตัวนั้นจะมีค่าใช้จ่ายที่สูงมาก การปรับปรุงภาพส่องกล้องจึงเป็นที่นิยมกว่า เนื่องจากใช้เวลาที่น้อยกว่าในการฝึกสอนนักส่องกล้องให้มีความชำนาญในการตรวจจ็บบริเวณที่ผิดปกติในกระเพาะอาหาร

ในทางวิศวกรรมศาสตร์ โมเดลการเรียนรู้เชิงลึก (Deep learning model, DLM) ได้เข้ามามีบทบาทในงานด้านการแพทย์เป็นอย่างมาก โมเดลที่ได้รับความนิยมในการส่องกล้องทางเดินอาหารส่วนบนนั้น จะเป็นโมเดลที่ทำหน้าที่ตรวจจ็บ (Detection) และการจำแนกรูปภาพเชิงความหมาย (Segmentation) บริเวณที่เป็นรอยโรคมะเร็งกระเพาะอาหารเป็นหลัก [9, 10] ในบรรดาโมเดลทั้งหมด ณ. ปัจจุบัน โมเดล DeepLabV3+ [11] และ U-Net [12] ถูกจัดให้เป็นโมเดล State-of-the-art สำหรับโมเดลการจำแนกรูปภาพในงานทางการแพทย์ อย่างไรก็ตาม โมเดลเหล่านี้ไม่สามารถทำงานได้ตามเวลาจริงที่ต้องการการประมวลผลของโมเดล มากกว่า 25 เฟรมต่อวินาที (Frame per second, FPS) [13] เนื่องจากการทำงานของโมเดลการจำแนกรูปภาพเชิงความหมายนั้น จำเป็น

จะต้องใช้การคำนวณจำนวนมากในภาพถ่ายความละเอียดสูง เป็นเหตุให้ภาพจากกล้องส่องเกิดความเหลื่อมล้ำระหว่างเฟรมของภาพ นำมาซึ่งอาการกระตุกในขณะที่ทำการส่องกล้อง

ในปัจจุบัน การนำโมเดลมาใช้งานจริง ยังเกิดปัญหาการหน่วงของภาพอยู่มาก โดยผลวิจัยของ Rodriguez Diaz ที่ประยุกต์ DeepLabV3+ ในการจำแนกตั้งเนื้อในลำไส้ ได้ผลลัพธ์การประมวลผลของโมเดลเพียงแค่ 10 เฟรมต่อวินาที [14], เช่นเดียวกับกับในงานของ Wang ที่ประยุกต์ DeepLabV3+ กับภาพรอยโรคในกระเพาะอาหาร ได้ผลลัพธ์การประมวลผลของโมเดลเพียง 12 เฟรมต่อวินาที [15], ส่วนในงานของ Sun ที่ประยุกต์ U-Net ได้ผลลัพธ์การประมวลผลของภาพทดสอบความละเอียด 384x288 จุดสี ของโมเดลเพียง 22 เฟรมต่อวินาที [16] และอ้างอิงจากการทดลองเบื้องต้นของโมเดล U-Net ของผู้วิจัย เมื่อทดสอบบนชุดทดสอบของโรงพยาบาลจุฬา ที่มีขนาด 1920x1080 จุดสีนั้น จะได้ผลลัพธ์การประมวลผลของโมเดลเพียงแค่ 3 เฟรมต่อวินาที ซึ่งไม่เพียงพอต่อการใช้งานตามเวลาจริง และปัญหาที่เกิดจากจำนวนภาพที่มีน้อย ทำให้โมเดลการจำแนกรูปภาพเชิงความหมายบริเวณที่เป็นรอยโรคได้ไม่ดีมาก ดังนั้น การเพิ่มเทคนิคต่าง ๆ เพื่อเพิ่มความแม่นยำนอกเหนือจากการปรับปรุงโมเดล เช่น การเรียนรู้แบบโอนถ่าย [17], การปรับปรุงภาพ และการเพิ่มข้อมูล จึงมีความสำคัญ และช่วยเพิ่มความแม่นยำของโมเดล



ภาพที่ 1 ภาพจากการทำหัตถการการส่องกล้องทางเดินอาหารส่วนบนโดยใช้แสงขาว

วิทยานิพนธ์นี้มุ่งหวังที่จะปรับปรุงโมเดลการเรียนรู้เชิงลึกในชื่อ Bilateral Segmentation Network (BiSeNet) [18] ที่มีประสิทธิภาพการทำงานตามเวลาจริง อ้างอิงจากงานแข่งขันทางภาพถ่ายท้องถนน CityScapes [19] พร้อมทั้งเพิ่มเทคนิคใหม่ โดยจะเสนอทั้งหมดสามเทคนิค (1) การเรียนรู้แบบโอนถ่าย โดยใช้ public dataset ตั้งเนื้อจากลำไส้เป็นส่วนประกอบ (CVC-ClinicDB



[20], Kvasir-SEG [21]) (2) ใช้เทคนิค Contrast Limited Adaptive Histogram Equalized (CLAHE) [22] ในการปรับปรุงภาพก่อนเข้าโมเดลเพื่อเพิ่มประสิทธิภาพ (3) ใช้เทคนิคการเพิ่มปริมาณข้อมูลในชุดฝึกสอน (Augmentation) โดยทั้งสามเทคนิคนี้สามารถปรับปรุงความแม่นยำในการตรวจจับรอยโรคให้ดียิ่งขึ้น สามารถใช้งานได้จริงในการส่องกล้องทางเดินอาหารส่วนบน

## 1.2 วัตถุประสงค์ของงานวิจัย

เพื่อพัฒนาโมเดลการจำแนกรูปภาพเชิงความหมาย (Semantic segmentation) ให้สามารถตรวจจับตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหาร (Gastric intestinal Metaplasia) ได้อย่างมีประสิทธิภาพ โดยทดสอบบนวิดีโอที่บันทึกการทำหัตถการส่องกล้องระบบทางเดินอาหารส่วนบน (Upper gastroscopy video) และชุดทดสอบ ของศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย โดยมีเป้าหมายหลักคือรองรับการทำงานตามเวลาจริง (Real-time system) หรือได้ความเร็วในการประมวลผลของภาพต่อวินาที (Frame rate per second) มากกว่า 25 ภาพ ต่อวินาที และมีความแม่นยำที่มากกว่าโมเดลบรรทัดฐาน (Baseline model)

## 1.3 ขอบเขตของงานวิจัย

- งานวิจัยชิ้นนี้มุ่งเน้นไปที่การตรวจจับ บริเวณที่เป็นตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหารเพียงอย่างเดียว
- งานวิจัยชิ้นนี้มุ่งเน้นไปที่การพัฒนาโมเดลที่รองรับการทำงานตามเวลาจริง (ความเร็วในการประมวลผลของภาพต่อวินาที มากกว่า 25 ภาพ ต่อวินาที) โดยจะเลือกโมเดลการจำแนกรูปภาพเชิงความหมาย BiSeNet มาปรับปรุงแก้ไข และเพิ่มประสิทธิภาพให้โมเดลสามารถทำงานได้ดีในภาพถ่ายการทำหัตถการส่องกล้องระบบทางเดินอาหารส่วนบน
- งานวิจัยชิ้นนี้จัดทำและทดสอบ บนข้อมูลภาพถ่ายกระเพาะอาหารความละเอียดสูง ของศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย โดยทางผู้วิจัยได้จัดเตรียมข้อมูลจริงจากผู้ป่วย 136 คน ร่วมกันกับทางทีมแพทย์

## 1.4 ประโยชน์ที่ได้รับ

ได้โมเดลที่มีประสิทธิภาพในการจำแนกรูปภาพเชิงความหมายที่เป็นตั้งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหารได้อย่างแม่นยำ และรองรับการทำงานตามเวลาจริง โดยสามารถพัฒนาต่อยอด นำโมเดลนี้ไปใช้งานจริงได้ในอนาคต

### 1.5 วิธีดำเนินงานวิจัย

- 1) ศึกษางานวิจัยที่เกี่ยวข้องกับหัวข้อที่จะทำ ทั้งงานวิจัยทางการแพทย์ และงานวิจัยทางวิศวกรรมศาสตร์
- 2) ศึกษาคลังข้อมูล ทั้งคลังข้อมูลแบบเปิด (Open dataset) และคลังข้อมูลจากศูนย์ส่งกล้องโรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย เพื่อทำความเข้าใจในตัวข้อมูล และหาแนวทางในการทำวิจัยให้เหมาะสม
- 3) จำลองการทำงานของโมเดล Baseline ที่ได้รับการรองรับ เพื่อใช้เป็นมาตรฐานเปรียบเทียบ บนชุดข้อมูลของผู้จัดทำ อีกทั้งเป็นการประเมินประสิทธิภาพของโมเดลของวิทยานิพนธ์ชิ้นนี้
- 4) ประยุกต์ ปรับปรุง และเพิ่มเติมวิธีการใหม่ ๆ จากการศึกษางานวิจัยที่เกี่ยวข้อง เพื่อให้เกิดองค์ความรู้ และเกิดการต่อยอด
- 5) วิเคราะห์ผลการทดลองจากงานวิจัยที่ได้รับการรองรับ ข้อที่ 3) และจากผู้วิจัย ข้อที่ 4) ว่ามีข้อดี และข้อเสียอย่างไร มีประสิทธิภาพต่างกันอย่างไรบ้าง
- 6) ตีพิมพ์ และจัดเสนอผลงานทางวิชาการ
- 7) จัดทำวิทยานิพนธ์



### 1.7 จริยธรรมการวิจัย

คณะกรรมการจริยธรรมการวิจัยในคน คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ดำเนินการให้การรับรองโครงการวิจัยตามแนวทางหลักจริยธรรมการวิจัยในคนที่เป็นมาตรฐานสากล ได้แก่ Declaration of Helsinki, The Belmont Report, CIOMS Guideline และ International Conference on Harmonization in Good Clinical Practice หรือ ICH-GCP, COA No. 1549/2020, IRB No. 726/62.

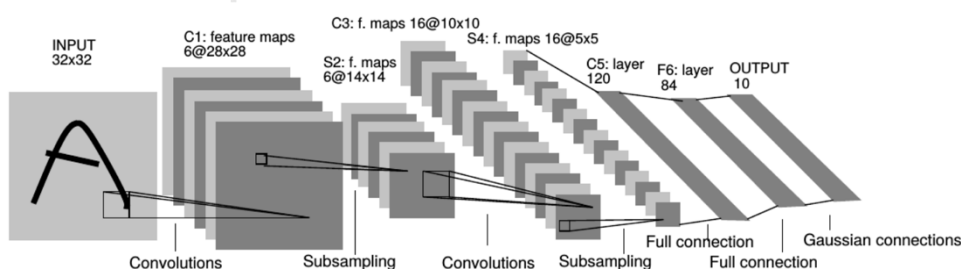


## บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องกับวิทยานิพนธ์ชิ้นนี้ประกอบไปด้วย โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network), การจำแนกรูปภาพเชิงความหมาย (Semantic segmentation), รูปอาร์จีบี (RGB image), การเพิ่มข้อมูล (Data augmentation), การเรียนรู้แบบโอนถ่าย (Transfer learning), การเรียนรู้แบบมีผู้สอน (Supervised learning), การเรียนรู้แบบกึ่งมีผู้สอน (Semi-supervised learning), การประเมินประสิทธิภาพ, การปรับค่าปกติ (Normalization), และการเพิ่มประสิทธิภาพโมเดลในแง่ของความเร็ว

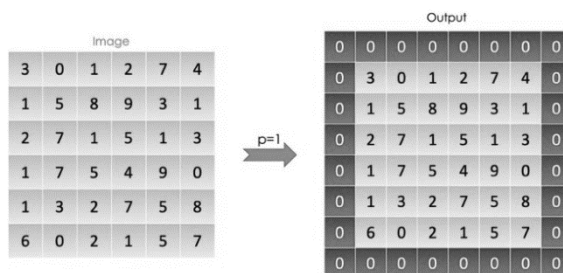
### 2.1 โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional neural network)

การเรียนรู้เชิงลึกที่เกี่ยวข้องกับรูปภาพ จะนำรูปนำเข้า มาสกัดผ่านตัวแยกคุณสมบัติ (Feature Extractor) จนได้แผนที่คุณสมบัติ (Feature Map) ที่อุดมไปด้วยข้อมูลเชิงบริบท (Context Information) จำนวนมาก ข้อมูลเชิงบริบทเหล่านี้ จะผ่านชั้นผลรวม (Pooling Layer) เพื่อให้ได้แผนที่คุณสมบัติเข้มข้น (Rich Context Information) จุดสว่างที่เกิดขึ้นบนแผนที่นั้น จะเป็นจุดที่โมเดลเกิดความสนใจ (Attention) ซึ่งแผนที่คุณสมบัติเหล่านี้ สามารถแปลงความหมายได้ผ่านชั้นเชื่อมต่อสมบูรณ์ (Fully Connected Layer, Dense Layer) โดยสามารถออกแบบผลลัพธ์ที่ต้องการจากโมเดลได้ การต่อกันของชั้นเหล่านี้ จะถูกเรียกว่า โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network) [23]



ภาพที่ 2 ภาพรวมของโครงข่ายประสาทแบบคอนโวลูชัน (อ้างอิงจาก fig2, [23])

ในบางกรณี ตัวแยกคุณสมบัติจะไม่สามารถสกัดข้อมูลจากขอบภาพรับเข้าได้ เนื่องจากเกิดปัญหาขนาดของตัวแยกคุณสมบัติ ที่มีขนาดใหญ่เกินกว่าพื้นที่ที่เหลือจากการสกัดภาพ ดังนั้น การเพิ่มพื้นที่ขยายความเข้าไปที่ขอบของภาพรับเข้า จะช่วยให้ตัวแยกคุณสมบัติ สามารถสกัดข้อมูลของภาพได้



ภาพที่ 3 ตัวอย่างการขยายความ (อ้างอิงจาก Coursera course, [24])

การก้าวข้าม (Stride) เป็นการกำหนดว่าจะเลื่อนตัวแยกคุณสมบัติออกไปเป็นปริมาณเท่าไร จะส่งผลต่อขนาดของผลลัพธ์ที่เกิดจากการผ่านชั้นคอนโวลูชัน โดยการก้าวข้ามที่มาก จะทำให้ขนาดของแผนที่คุณสมบัติมีขนาดเล็กลง โดยจะเติมค่าศูนย์เข้าไปบริเวณขอบภาพ โดยที่ตัวแปร  $p$  จะแสดงถึงจำนวนแถวและหลักที่จะขยายความออกไปทั้ง 4 ด้านของภาพ

หลังจากที่กำหนดตัวแปรขยายความและการก้าวข้ามแล้ว จึงจะสามารถกำหนดผลลัพธ์จากการทำคอนโวลูชันระหว่างเทนเซอร์ (tensor) และตัวแยกคุณสมบัติ ในงานประเภทรูปนั้น จะแทนตัวเทนเซอร์ และตัวแยกคุณสมบัติ ด้วยสมการคณิตศาสตร์ดังต่อไปนี้

$$\dim(\text{image}) = (n_H, n_W, n_C) \quad (1)$$

$$\dim(\text{filter}) = (f, f, n_C) \quad (2)$$

กำหนดให้

$n_H$  แทนความสูงของรูปภาพรับเข้า

$n_W$  แทนความกว้างของรูปภาพรับเข้า

$n_C$  แทนจำนวนช่องสัญญาณของรูปภาพรับเข้า

$f$  แทนขนาดของฟิลเตอร์จัตุรัสที่มีความกว้างและความยาวเท่ากัน

ในกรณีของรูปอาร์จีบี จะมีจำนวนช่องสัญญาณเท่ากับสามช่อง ประกอบไปด้วยช่องสัญญาณสีแดง, ช่องสัญญาณสีเขียว, และช่องสัญญาณสีน้ำเงิน ในเชิงปฏิบัติ ขนาดของตัวแยกคุณสมบัติจะมีลักษณะเป็นกล่องสี่เหลี่ยมที่มีความลึกเท่ากับจำนวนช่องสัญญาณ เพื่อที่จะสามารถสกัดคุณสมบัติจากเทนเซอร์ได้ โดยคาดหวังให้ขนาดของช่องสัญญาณ ลดลงเหลือเพียงช่องเดียว หลังจากผ่านการคำนวณ

$$\text{conv}(I, K)_{x,y} = \sum_{i=1}^{n_H} \left( \sum_{j=1}^{n_W} \left( \sum_{k=1}^{n_C} (K_{i,j,k} I_{x_{i-1}, y_{j-1}, k}) \right) \right) \quad (3)$$

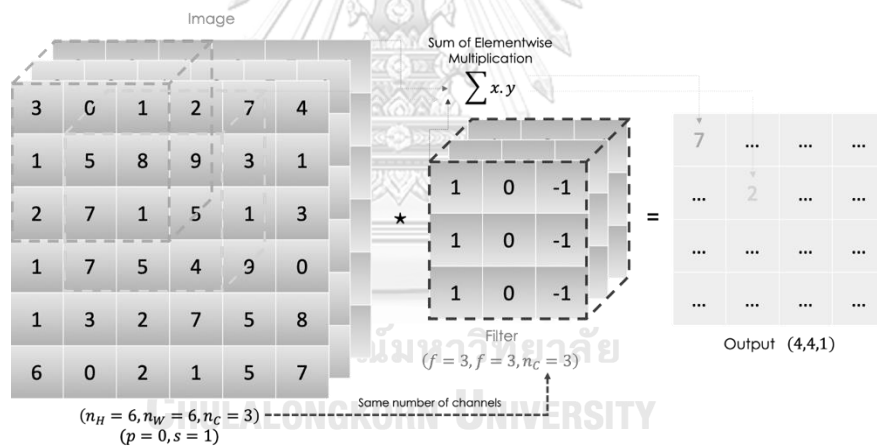
$$\text{dim}(\text{conv}(I, K)) = \left( \frac{n_H + 2p - f}{s} + 1, \frac{n_W + 2p - f}{s} + 1 \right); s > 0 \quad (4)$$

$$\text{dim}(\text{conv}(I, K)) = (n_H + 2p - f, n_W + 2p - f); s = 0 \quad (5)$$

กำหนดให้

$p$  แทนจำนวนการขยายความ

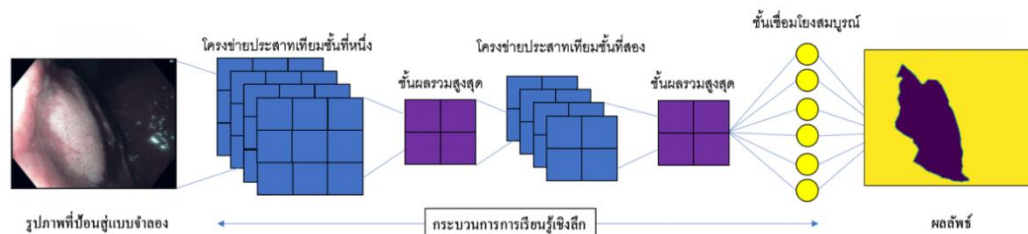
$s$  แทนจำนวนการก้าวข้าม



ภาพที่ 4 ตัวอย่างการทำคอนโวลูชัน (อ้างอิงจาก Coursera course, [24])

## 2.2 การจำแนกรูปภาพเชิงความหมาย (Semantic Segmentation)

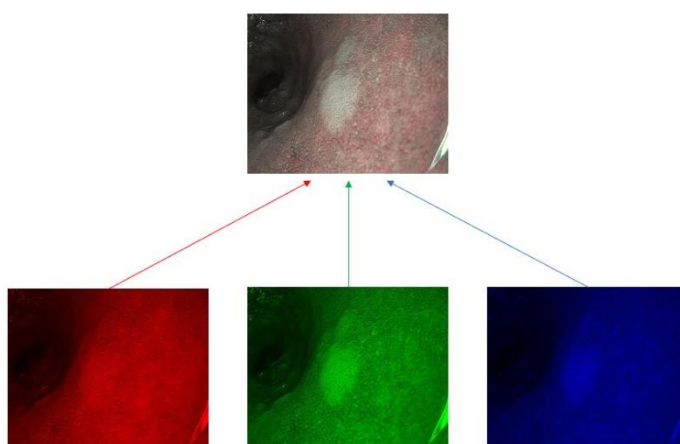
เป้าหมายในการจำแนกรูปภาพเชิงความหมาย คือการทำนายทุกจุดสีบนรูปโดยอ้างอิงหมวดหมู่ที่สนใจ โดยผลลัพธ์จะได้ภาพที่มีรายละเอียดสูง โดยโมเดลส่วนใหญ่จะให้ผลลัพธ์เป็นภาพที่ถูกจำแนกเชิงความหมาย ที่มีขนาดเท่ากับภาพรับเข้า ในงานที่มีผลเฉลยเพียงแค่นิดเดียว ผลลัพธ์ของโมเดลการจำแนกรูปภาพเชิงความหมาย จะให้ภาพที่มีเพียงสองสีเท่านั้น โดยจะแทนถึงบริเวณที่สนใจ และบริเวณที่ไม่สนใจ (Background) โดยภาพผลลัพธ์ที่ได้ในรูปตัวอย่างที่ 5 สีม่วงแทนบริเวณที่โมเดลทำนายแล้วเกิด ความมั่นใจมากกว่าร้อยละ 50



ภาพที่ 5 ภาพรวมของโครงข่ายประสาทแบบคอนโวลูชัน ที่ทำหน้าที่ในการจำแนกรูปภาพเชิง  
ความหมาย

### 2.3 รูปอาร์จีบี (RGB Image)

รูปอาร์จีบีจะประกอบไปด้วยช่องสัญญาณ 3 ช่อง คือช่องสัญญาณสีแดง, ช่องสัญญาณสีเขียว, และช่องสัญญาณสีน้ำเงิน โดยรูปอาร์จีบีจะนำแนวคิดจากการมองเห็นของมนุษย์มาปรับใช้ ด้วยการผสมสีระหว่างช่องสัญญาณทั้งสามให้เกิดเป็นสีที่แตกต่างกัน ดังนั้นแล้ว ช่องสัญญาณแต่ละช่อง จะมีมูลค่าเสมือนเป็น 1 รูปที่ประกอบไปด้วยชุดตัวเลขในรูปสองมิติ ระหว่าง 0 ถึง 255 โดยค่าน้อย ยิ่งน้อย ค่าสว่างที่ตำแหน่งนั้นจะใกล้เคียงกับสีดำ ในทางกลับกัน ค่ายิ่งมาก ค่าสว่างที่ตำแหน่งดังกล่าวจะใกล้เคียงกับสีขาว พอได้ค่ามาแล้วจึงจะทำการคูณเข้ากับตัวแปลงสัญญาณตามแต่ละสีเพื่อปรับจากภาพขาวดำ ให้เป็นค่าสเกลของแต่ละสี ในขั้นตอนสุดท้าย จะนำค่าจากทั้งสามช่องสัญญาณมารวมกันให้ได้เป็นรูปสี

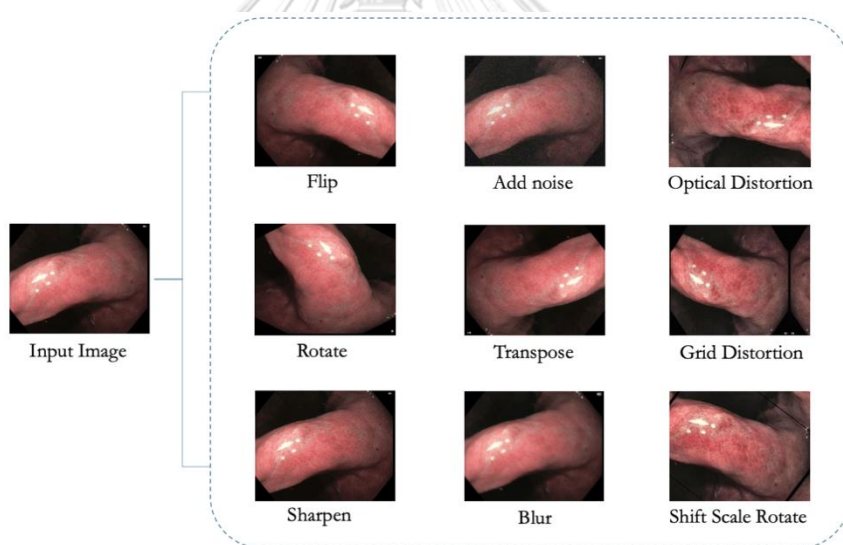


ภาพที่ 6 การรวมสัญญาณภาพอาร์จีบี



## 2.4 การเพิ่มข้อมูล (Data Augmentation)

เป้าหมายหลักในการทำกระบวนการนี้ คือการเพิ่มเติมคลังข้อมูลให้มีปริมาณเพื่อแก้ไขปัญหาปริมาณข้อมูลไม่เพียงพอต่อการฝึกสอนโมเดล โดยหลักการของการเพิ่มข้อมูล จะนำภาพดั้งเดิม มาทำกระบวนการเพิ่มข้อมูลแบบอ่อน (weak augmentation) เช่น การหมุนภาพ, ปรับความสว่าง, การซูมภาพ และกระบวนการเพิ่มข้อมูลแบบเข้ม (strong augmentation) เช่น การปรับความยืดหยุ่น, การปรับโครงสร้างภาพ โดยการเพิ่มข้อมูลนั้น ผู้จัดทำจะดูถึงความสอดคล้องและความสมเหตุสมผลของวิธีการนั้นด้วย ยกตัวอย่างเช่น ภาพถ่ายหัตถการการส่องกล้องทางเดินอาหาร ส่วนบนนั้น กระเพาะอาหารของผู้ป่วยทุกคนจะมีโทนสีเป็นสีส้ม ดังนั้นการปรับสีของภาพให้เป็นสีเขียว หรือสีน้ำเงิน จะไม่สมเหตุสมผลเพราะโอกาสที่จะเจอผู้ป่วยที่มีสีของกระเพาะอาหารเป็นสีอื่นที่ผิดปกติเป็นไปได้ได้น้อยมาก ดังนั้นการเพิ่มข้อมูล จำเป็นต้องผ่านการพิจารณาจากทั้งผู้วิจัย และทีมแพทย์ เพื่อฝึกสอนโมเดลให้มีความเที่ยงตรงมากที่สุด โดยในงานวิจัยชิ้นนี้จะใช้โปรแกรม อัลบูเมนเตชัน [25] ในการสร้างภาพข้อมูล

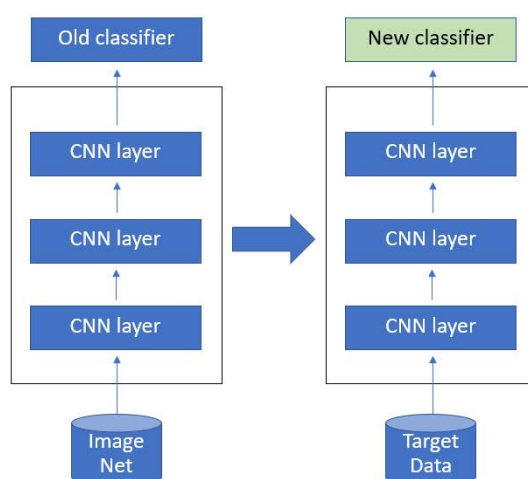


ภาพที่ 7 ตัวอย่างการเพิ่มข้อมูลภาพรอยโรค

## 2.5 การเรียนรู้แบบโอนถ่าย (Transfer learning)

การเรียนรู้แบบโอนถ่าย จะนำความรู้ที่ได้จากการสอนโมเดลบนข้อมูลที่ใกล้เคียงกัน มาทำการสอนโมเดลเป้าหมาย [17] ในงานวิจัยชิ้นนี้จะนำความรู้ที่ได้จากการสอนโมเดล โดยใช้คลังข้อมูลเปิดควาเซอร์ ที่มีความใกล้เคียงกันกับคลังข้อมูล ศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย เป้าหมายหลักของการเรียนรู้แบบโอนถ่าย เพื่อที่จะบรรเทาปัญหาข้อมูลจากทางโรงพยาบาลจุฬา ที่มีปริมาณน้อย ส่งผลต่อความแม่นยำของโมเดลที่ได้ โดยจะนำจะนำขั้นโครงข่าย

คอนโวลูชันที่ถูกสอนโดยข้อมูลฝึกสอนจำนวนมาก (e.g., ImageNet [26], COCO [27] และ Pascal VOC [28]) โอนถ่ายมาฝึกสอนข้อมูลเป้าหมาย โดยจะเรียกชั้นเหล่านี้ว่า ค่าน้ำหนักที่ถูกสอนล่วงหน้า (Pre-Trained Weight)



ภาพที่ 8 ตัวอย่างกระบวนการ การเรียนรู้แบบโอนถ่าย

## 2.6 การเรียนรู้ของเครื่อง (Machine learning)

การเรียนรู้ของเครื่องที่ถูกหยิบยกมาใช้ในงานวิจัยชิ้นนี้ จะประกอบไปด้วย การเรียนรู้แบบมีผู้สอน (Supervised learning) และ การเรียนรู้แบบกึ่งมีผู้สอน (Semi-supervised learning)

การเรียนรู้แบบมีผู้สอน เป็นการเรียนรู้ของเครื่องที่จะนำข้อมูลป้อนเข้า จับคู่กับผลเฉลยจากผู้สอน เข้าไปฝึกสอนโมเดล โดยข้อมูลนำเข้าและผลเฉลยนั้นจะทำการปรับจูนโมเดลให้สามารถทำนายผลลัพธ์ที่ต้องการ ส่วนการเรียนรู้แบบกึ่งมีผู้สอนนั้น จะเป็นการใช้ข้อมูลที่ไม่มีผลเฉลยแต่แรก แต่จะสร้างผลเฉลยขึ้นมาโดยใช้โมเดลในการสร้างผลเฉลยแบบเสมือน (Pseudo label) โดยผลเฉลยจากการทำนายของโมเดลแบบมีผู้สอน จะถูกนำมาประกบคู่กับข้อมูลถูกทำนายที่ไม่มีผลเฉลย แล้วจึงนำไปฝึกสอนร่วมกับข้อมูลที่มีผู้สอนในตอนแรก เพื่อเป็นการเพิ่มข้อมูลในโมเดล และเพิ่มความแม่นยำมากขึ้น

## 2.7 การประเมินประสิทธิภาพ

ในการวัดประสิทธิภาพของโมเดล ผู้จัดทำได้ใช้มาตรวัดในการประเมินประสิทธิภาพของโมเดลสองรูปแบบ 1) มาตรวัดประสิทธิภาพในการตัดแยกของโมเดล (Classification performance criteria) จะเป็นการนับจำนวนที่ภาพโมเดลสามารถคัดแยกได้อย่างถูกต้อง 2) มาตรวัดประสิทธิภาพ

การจำแนกของโมเดล (Segmentation performance criteria) จะประเมินโดยการเทียบผลลัพธ์จากการจำแนกของโมเดลต่อภาพในชุดทดสอบ เทียบกับผลเฉลยจากทีมแพทย์ โดยมาตรวัดที่ใช้ในงานวิจัยชิ้นนี้ จะประกอบไปด้วย ค่า Sensitivity, Specificity, Positive predictive value (PPV), Negative predictive value (NPV), Accuracy, และ F1-Score สำหรับการวัดประสิทธิภาพการตัดแยกของโมเดล และ Intersect over Union, และ Error สำหรับการวัดประสิทธิภาพการจำแนกของโมเดล

การวัดประสิทธิภาพการตัดแยกของโมเดลนั้น จะนำผลลัพธ์ที่ได้จากการจำแนกภาพรอยโรคและภาพปกติ มาคำนวณหาเมตริกสับสน (Confusion matrix) โดยการจำแนกภาพรอยโรคที่มีค่า Intersect over Union (IoU) มากกว่าร้อยละ 0.3 จะถูกนับว่าเป็นผลบวกจริง (True positive) และการจำแนกภาพที่ไม่มีรอยโรค แล้วได้ค่า Error น้อยกว่าร้อยละ 0.01 จะถูกนับว่าเป็นผลลบจริง (True negative) [29] โดยการหาค่า Sensitivity, Specificity, Accuracy, และ F1 จะเป็นไปดังสมการ

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (7)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$F1 = \frac{2 \cdot (\text{Sensitivity} \cdot \text{Specificity})}{\text{Sensitivity} + \text{Specificity}} \quad (9)$$

กำหนดให้

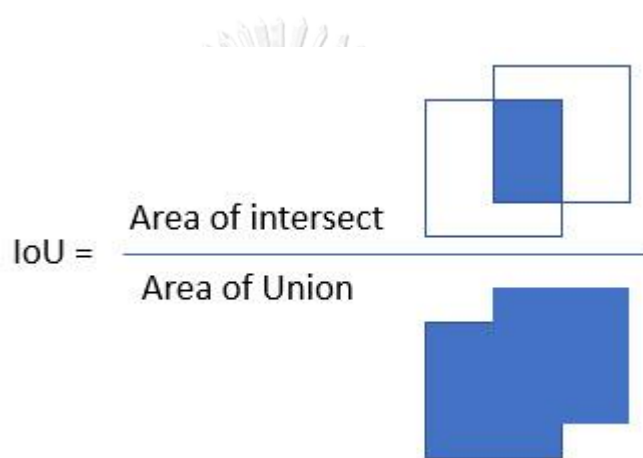
TP, True Positive, ผลบวกจริง: คือ จำนวนภาพที่โมเดลทายว่ามีรอยโรค บนข้อมูลที่มีผลเฉลยที่มีรอยโรคปรากฏ

FP, False Positives, ผลบวกเท็จ: คือ จำนวนภาพที่โมเดลทายว่ามีรอยโรค แต่บนข้อมูลที่มีผลเฉลย ไม่มีรอยโรคปรากฏ

TN, True Negative, ผลลบจริง: คือ จำนวนภาพที่โมเดลทายว่าไม่มีรอยโรค บนข้อมูลที่มีผลเฉลยที่ไม่มีรอยโรคปรากฏ

FN, False Negatives: ผลลบเท็จ: คือ จำนวนภาพที่โมเดลทายว่าไม่มีรอยโรค แต่บนข้อมูลที่มีผลเฉลยมีรอยโรคปรากฏ

ส่วนการวัดประสิทธิภาพการจำแนกของโมเดลนั้น มาตรฐานดังกล่าวจะสนใจพื้นที่ที่ถูกทับซ้อนระหว่างผลเฉลย กับผลการทำนาย ค่าความแม่นยำจากการใช้มาตรวัดจะอยู่ระหว่าง 0 ถึง 1 โดยค่ายิ่งมาก จะหมายถึงบริเวณที่โมเดลทำนายนั้น ทับซ้อนกับผลเฉลยมาก โดยมาตรวัดนี้จะถูกเรียกว่า Intersect over Union (IoU) ส่วนค่าความคลาดเคลื่อนนั้น (Error) จะถูกคำนวณจากพื้นที่ที่โมเดลทำนายเป็นผลลัพธ์บวกจากภาพที่ไม่มีรอยโรค หาดด้วยพื้นที่ทั้งหมดของภาพ โดยมีการคำนวณดังนี้



$$\text{IoU} = \frac{\text{Area of intersect}}{\text{Area of Union}}$$

ภาพที่ 9 ตัวอย่างการคำนวณ Intersect over Union

ในส่วนมาตรวัดด้านความเร็วประมวลผล จะบ่งบอกถึงจำนวนภาพผลลัพธ์ที่ได้จากการทำนายโมเดลในหนึ่งวินาที โดยการที่จะทำโมเดลให้สามารถใช้งานได้ตามเวลาจริงนั้น ขั้นต่ำจะต้องการ 25 ภาพต่อวินาที

## 2.8 การปรับค่าปกติ (Normalization)

การปรับค่าปกติ จำเป็นอย่างยิ่งสำหรับการคำนวณที่เกี่ยวกับภาพ เนื่องจากจุดสีของภาพรับเข้าแต่ละช่องสัญญาณ จะมีค่าอยู่ที่ระหว่าง 0 ถึง 255 การปรับค่าปกติจะทำให้ค่าของจุดสีแต่ละจุดในรูป อยู่ทีระหว่าง 0 ถึง 1 ทำให้การคำนวณไม่เกิดปัญหาค่าบานออก (Divergent) จนส่งผลให้โมเดลรับการคำนวณที่มากเกินไปจนเกิดข้อขัดข้องขึ้น

$$\text{Normalization} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (10)$$

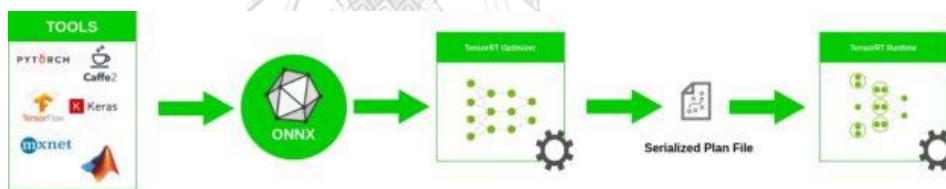
กำหนดให้

$x$  คือ ค่าจุดสีของภาพ จะมีค่าต่ำสุด และสูงสุด อยู่ระหว่าง 0 ถึง 255

## 2.9 การเพิ่มประสิทธิภาพโมเดลในแง่ของความเร็ว

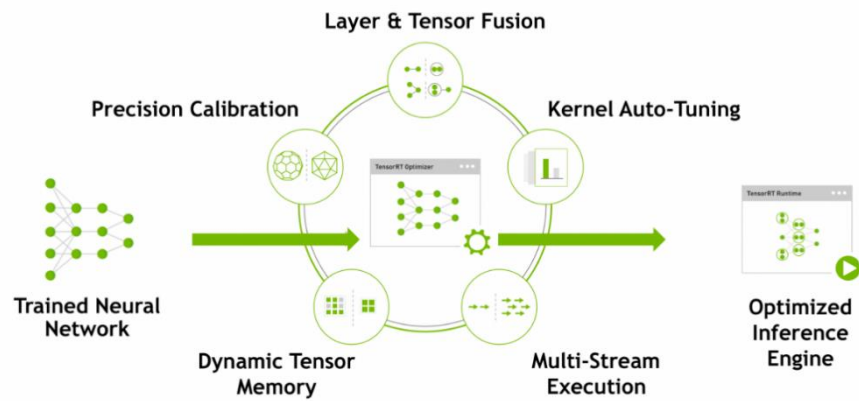
ในการเพิ่มประสิทธิภาพโมเดลในแง่ของความเร็ว จะประกอบไปด้วยสองโปรแกรม โอนิกซ์ (ONNX, Open Neural Network Exchange) และ เทนเซอร์อาที (TensorRT)

โอนิกซ์เป็นรูปแบบชนิดหนึ่งสำหรับการทำงานของเครื่อง และการเรียนรู้เชิงลึก ที่สามารถเปลี่ยนถ่ายโมเดลจากแพลตฟอร์มพื้นฐานที่แตกต่างกัน (เช่น TensorFlow, PyTorch, MATLAB, Keras) ให้อยู่ในโมเดลรูปแบบโอนิกซ์ หรือให้อยู่ในรูปแบบที่มีการคำนวณที่เรียบง่ายขึ้น มีการตัดบางการคำนวณที่ไม่จำเป็นออก และลดความซับซ้อนของโมเดลลง โดยโอนิกซ์จะเปลี่ยนถ่ายโมเดลตั้งต้นให้อยู่ในรูปแบบที่สามารถใช้งานหน่วยประมวลผลกราฟิกของ NVIDIA ได้อย่างมีประสิทธิภาพโดยทั่วไปแล้ว โอนิกซ์จะช่วยให้โมเดลมีขนาดเล็กลงโดยประมาณ 2 เท่า ทำให้ใช้เวลาในการคำนวณน้อยลง และเพิ่มประสิทธิภาพในแง่ของความเร็วทั้งตอนฝึกสอน และตอนนำไปใช้งาน



ภาพที่ 10 ขั้นตอนการทำงานของโอนิกซ์ (อ้างอิงจาก [Nvidia.com/tensorrt](https://nvidia.com/tensorrt))

เทนเซอร์อาทีเป็นฟังก์ชันติดตั้งบนโมเดลเสมือนของ NVIDIA (NVIDIA's parallel programming model, CUDA) โดยจะช่วยเพิ่มประสิทธิภาพในแง่ของความเร็วของโมเดลการเรียนรู้เชิงลึก โดยจะมอบการปรับปรุงที่เหมาะสมในรูปแบบ 8 บิต และ 16 บิต เช่น การไหลของวิดีโอ (Video streaming) หรืองานทางด้านภาษาธรรมชาติ โดยเป้าหมายหลักจะลดการอนุมานลง (Reduced precision inference) ทำให้มีผลช่วยลดความหน่วงของโมเดลเวลานำไปใช้งานจริง



ภาพที่ 11 ขั้นตอนการทำงานของเทนเซอร์อาที (อ้างอิงจาก [Nvidia.com/tensorrt](https://nvidia.com/tensorrt))



### บทที่ 3

#### งานวิจัยที่เกี่ยวข้อง

#### 3.1 โมเดลสถาปัตยกรรมรูปตัวยู (U-Net architecture)

โมเดลสถาปัตยกรรมรูปตัวยู เป็นสถาปัตยกรรมที่ถูกออกแบบมาโดยใช้โครงข่ายประสาทแบบคอนโวลูชันสมบูรณ์ และดัดแปลงให้จำแนกรูปภาพเชิงความหมายได้ดีในภาพถ่ายทางการแพทย์ [12] สถาปัตยกรรมรูปตัวยูจะมีจุดแตกต่างจากสถาปัตยกรรมการเรียนรู้เชิงลึกตัวอื่นตรงที่

(1) ชั้นโครงข่ายถูกออกแบบมาให้มีความสมมาตรทั้งฝั่งเข้ารหัส (Encoder) และฝั่งถอดรหัส (decoder) คล้ายรูปตัวยู

(2) ช่วงข้ามการเชื่อมต่อ (Skip connection) ของฝั่งเข้ารหัส จะนำมาเชื่อมต่อกัน (Concatenation Operation) กับ ฝั่งถอดรหัส แทนที่จะนำมารวมกันแบบปกติ ทำให้ชั้นที่เกิดช่วงข้ามการเชื่อมต่อ จะมีรายละเอียดคุณสมบัติต่ำ (Low level feature) จากฝั่งถอดรหัสติดมาด้วย ทำให้ช่วงการเพิ่มรายละเอียด (Upsampling) จะมีข้อมูลเพิ่มมากขึ้น และช่วยให้รายละเอียดคุณสมบัติต่ำ ไม่หล่นหายระหว่างทาง และด้วยคุณสมบัติของสถาปัตยกรรมที่ถูกออกแบบมาให้มีความสมมาตร และมีช่วงข้ามการเชื่อมต่อระหว่างทั้งสองฝั่ง ทำให้เกิดแผนที่คุณลักษณะ (Feature map) จำนวนมาก ในช่วงการเพิ่มรายละเอียด

โดยสถาปัตยกรรมรูปตัวยู จะประกอบไปด้วยส่วนหลัก 3 ส่วน

#### (1) ช่วงการลดรายละเอียด (Downsampling)

ในช่วงการลดรายละเอียด จะประกอบไปด้วยโครงข่ายประสาทแบบคอนโวลูชัน 4 กลุ่ม แต่ละกลุ่มจะประกอบไปด้วย

- ชั้นโครงข่ายประสาทแบบคอนโวลูชัน ขนาด  $3 \times 3$  และ ฟังก์ชันกระตุ้นพร้อมกับการให้ชุดข้อมูลเป็นมาตรฐาน (Activation function + Batch normalization)
- ชั้นโครงข่ายประสาทแบบคอนโวลูชัน ขนาด  $3 \times 3$  และ ฟังก์ชันกระตุ้นพร้อมกับการให้ชุดข้อมูลเป็นมาตรฐาน (Activation function + Batch normalization)
- ชั้นผลรวมสูงสุด (Max pooling layer) ขนาด  $2 \times 2$  ตามลำดับ

จุดประสงค์หลักของการลดรายละเอียด เพื่อที่จะตรวจจับข้อมูลบริบท (Contexture information) ของภาพที่ป้อนเข้าโมเดล ทำให้สามารถจำแนกตำแหน่งที่สนใจได้ โดยที่ข้อมูลบริบทที่ได้จะถูกถ่ายโอนไปยังช่วงการเพิ่มรายละเอียดโดยอาศัยผ่านทางช่วงการเชื่อมต่อ

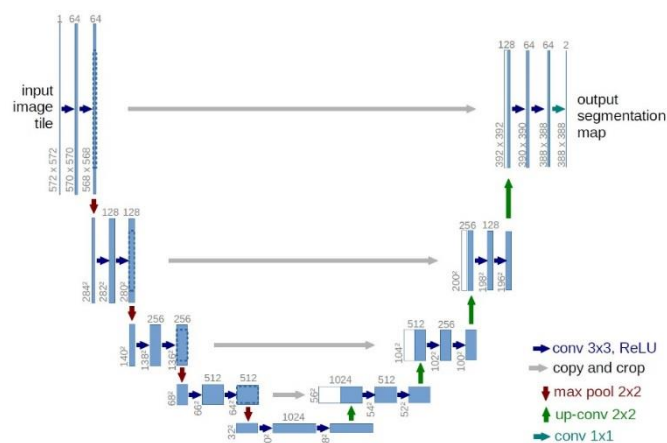
(2) ช่วงคอขวด (Bottleneck)

ช่วงคอขวด เป็นช่วงรอยต่อระหว่างช่วงการลดรายละเอียด และช่วงการเพิ่มรายละเอียด ประกอบไปด้วยโครงข่ายประสาทแบบคอนโวลูชันอย่างง่ายสองชั้น และการเจือจาง (Dropout)

(3) ช่วงการเพิ่มรายละเอียด

ในช่วงการเพิ่มรายละเอียด จะประกอบไปด้วยโครงข่ายประสาทแบบคอนโวลูชัน 4 กลุ่ม แต่ละกลุ่มจะประกอบไปด้วย

- ชั้นโครงข่ายประสาทคอนโวลูชันแบบย้อนกลับ (Deconvolution) ที่มีการก้าว่าง 2 ระดับ (Stride)
- การเชื่อมต่อแผนที่คุณลักษณะที่ถูกครอบตัดจากฝั่งเข้ารหัส สู่มฝั่งถอดรหัส ผ่านทางช่วงข้ามการเชื่อมต่อ
- ชั้นโครงข่ายประสาทแบบคอนโวลูชัน ขนาด  $3 \times 3$  และ ฟังก์ชันกระตุ้นพร้อมกันทำให้ชุดข้อมูลเป็นมาตรฐาน
- ชั้นโครงข่ายประสาทแบบคอนโวลูชัน ขนาด  $3 \times 3$  และ ฟังก์ชันกระตุ้นพร้อมกันทำให้ชุดข้อมูลเป็นมาตรฐาน ตามลำดับ



ภาพที่ 12 สถาปัตยกรรมรูปตัวยู (อ้างอิงจาก fig 1, [12])

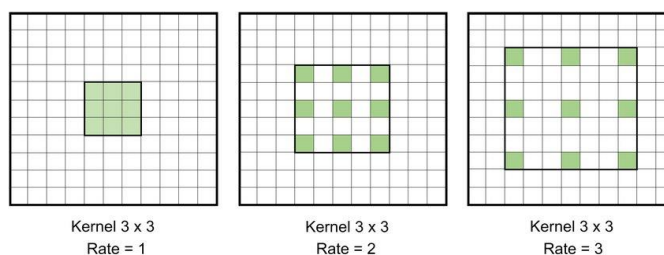


สรุปประโยชน์ของสถาปัตยกรรมโดยสังเขป

- สถาปัตยกรรมรูปตัวยูจะรวมข้อมูลตำแหน่งจากช่วงการลดรายละเอียด เข้ากับข้อมูลบริบทจากช่วงการเพิ่มรายละเอียดเข้าด้วยกัน ทำให้ได้รับข้อมูลโดยสังเขปที่มีทั้งตำแหน่งและบริบท ซึ่งจำเป็นอย่างยิ่งในการทำนายพื้นที่ที่จะจำแนกรูปภาพเชิงความหมาย
- สถาปัตยกรรมรูปตัวยู จะไม่มีชั้นเชื่อมโยงสมบูรณ์ (Fully connected layer, Dense Layer) ดังนั้นขนาดของรูปที่ต่างกันในคลังข้อมูล จึงไม่เป็นประเด็นที่ทำให้ตัวโมเดลทำงานผิดพลาด
- สถาปัตยกรรมรูปตัวยู จะมีช่วงข้ามการเชื่อมต่อจากช่วงลดรายละเอียด ไปยังช่วงเพิ่มรายละเอียด ทำให้เกิดแผนที่คุณลักษณะจำนวนมาก ขึ้นระหว่างดำเนินการ ซึ่งจะเป็ประโยชน์อย่างยิ่งต่องานที่เกี่ยวกับภาพถ่ายทางการแพทย์ ที่มีปริมาณข้อมูล อยู่อย่างจำกัด

### 3.2 โมเดลสถาปัตยกรรมตีปแลป (DeepLab)

สถาปัตยกรรมตีปแลป ถูกพัฒนาโดยทีมนักวิจัยจากกูเกิล (Google) สถาปัตยกรรมตัวนี้ได้้นำการทำคอนโวลูชันแบบหลุม (Atrous convolution) ในช่วงการเพิ่มรายละเอียด การทำคอนโวลูชันแบบหลุม จะปรับความกว้างของน้ำหนึ่กของตัวกรองให้สามารถนำข้อมูลที่อยู่ไกล มาพิจารณาในการคำนวณแผนที่คุณลักษณะได้ เป็นแนวคิดเพื่อใช้แก้ไขปัญหาการสูญเสียคุณลักษณะของแผนที่คุณลักษณะ ที่เกิดมาจากการคำนวณ ระหว่างการหาผลรวมสูงสุด และการก้าวข้าม ของชั้นเชื่อมโยงสมบูรณ์ในโครงข่ายประสาท เป็นจำนวนหลาย ๆ ครั้ง ทำให้โมเดลหลงลืมคุณลักษณะต่ำไปเมื่อถูกคำนวณซ้ำหลาย ๆ รอบ โดยหนึ่งในวิธีที่ทางทีมนักวิจัยใช้ คือการลดรายละเอียด ของชั้นโครงข่ายประสาท ที่อยู่ในช่วงการเพิ่มรายละเอียด โดยการทำคอนโวลูชันแบบหลุมนี้จะเพิ่มประสิทธิภาพในการค้นหาตำแหน่งที่กว้างขึ้นโดยที่ไม่เพิ่มจำนวนลักษณะพิเศษ (Parameter) อีกทั้งยังไม่เพิ่มภาระในการคำนวณ ของโมเดลอีกด้วย



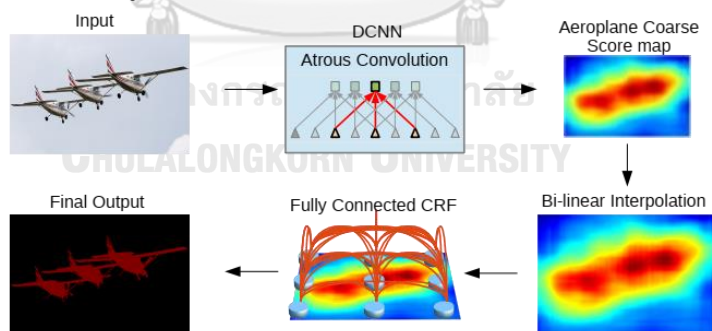
ภาพที่ 13 การทำคอนโวลูชันแบบหลุม (อ้างอิงจาก fig 6, [30])

### 3.2.1 สถาปัตยกรรมตีปแลปวีหนึ่ง (DeeplabV1)

สถาปัตยกรรมตัวนี้ได้นำแนวคิดการทำคอนโวลูชันแบบหลุมมาปรับใช้ [31] โดยการถอดการคำนวณการลดรายละเอียดในช่วงลดรายละเอียดออก แล้วจึงเพิ่มแนวคิดดังกล่าวลงไปในโครงข่ายประสาท ส่งผลให้การคำนวณนั้น เกิดแผนที่คุณลักษณะเพิ่มขึ้นมากกว่าแบบปกติ สามารถแก้ไขปัญหาการสูญเสียคุณลักษณะของแผนที่คุณลักษณะได้อย่างมีประสิทธิภาพ

อีกปัญหาหนึ่งของโมเดลจำแนกรูปภาพเชิงความหมาย คือการเก็บรายละเอียดของคุณลักษณะได้ไม่ครบ อันเนื่องมาจากการผ่านชั้นโครงข่ายประสาท จึงมีการนำสนามสุ่มแบบมีเงื่อนไข (Conditional Random field) เข้ามาแก้ปัญหา โดยวิธีดังกล่าวจะหาจุดสีที่มีความคล้ายกัน แล้วสร้างความสัมพันธ์ของวัตถุที่สนใจผ่านกลุ่มของจุดสีเพื่อกำหนดขอบเขตของวัตถุดังกล่าว ทำให้โมเดลสามารถจับรายละเอียดได้ราบเรียบในแต่ละวัตถุประเภท

สถาปัตยกรรมตีปแลปวีหนึ่ง จะนำภาพรับเข้า ผ่านชั้นโครงข่ายประสาทที่มีการทำคอนโวลูชันแบบหลุม การผ่านชั้นดังกล่าวจะได้คะแนนความสนใจแบบหยาบ (Coarse score map) โดยคะแนนความสนใจแบบหยาบนี้ คือภาพป้อนเข้าที่ผ่านการเพิ่มรายละเอียดและอยู่ในรูปแบบของการประมาณค่าในช่วงสองเส้น (Bi-Linear Interpolation) และผ่านสนามสุ่มแบบมีเงื่อนไข เพื่อเพิ่มประสิทธิภาพของการจำแนกรูปภาพเชิงความหมาย

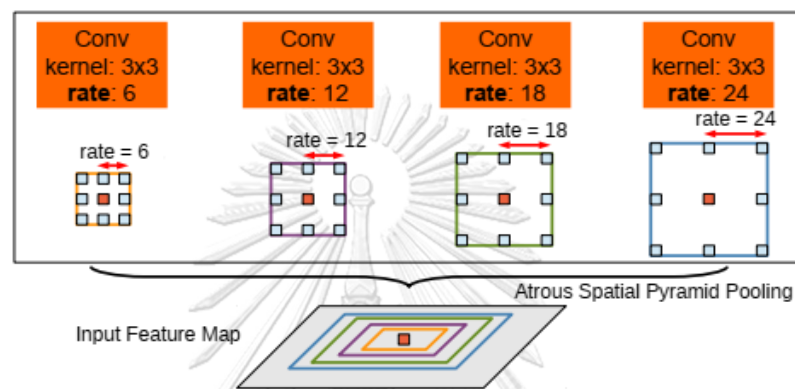


ภาพที่ 14 ขั้นตอนการทำงานของตีปแลปวีหนึ่ง (อ้างอิงจาก fig 1, [31])

### 3.2.2 สถาปัตยกรรมตีปแลปวีสอง (DeeplabV2)

สถาปัตยกรรมตีปแลปวีสองได้รับการพัฒนาต่อเนื่องมาจากสถาปัตยกรรมรุ่นก่อนหน้า โดยจะเพิ่มการรองรับการมีอยู่ของวัตถุหลากหลายสเกล (Existence of object at multiple scale) ในการรองรับการมีอยู่ของวัตถุหลากหลายสเกล วิธีการมาตรฐานที่จะจัดการกับปัญหานี้ คือการเปลี่ยนแปลงสเกลของรูปต้นฉบับ และรวมกลุ่ม (Aggregat) แผนที่คุณลักษณะ

วิธีการเปลี่ยนแปลงสเกลของรูปต้นฉบับ ที่งานวิจัยชิ้นนี้จัดทำขึ้น คือการใช้ การรวมเชิงพื้นที่ รูปกรวยเหลี่ยมหลุม (Atrous spatial pyramid pooling) โดยแนวคิดดังกล่าวจะนำ การทำคอนโวลูชันแบบหลุมหลากหลายขนาด ที่มีค่าการสุ่มหลายค่า (Sampling rate) บนแผนที่คุณลักษณะ ก่อนที่จะนำมาหลอมรวม (Fuse) เข้าด้วยกัน ทำให้โมเดล สามารถระบุวัตถุชนิดเดียวกัน ที่มีขนาดแตกต่างกันได้ ส่งผลให้สถาปัตยกรรมรุ่นนี้มีความแม่นยำเพิ่มสูงขึ้น



ภาพที่ 15 การรวมเชิงพื้นที่รูปกรวยเหลี่ยมหลุม (อ้างอิงจาก fig 4, [31])

### 3.2.3 สถาปัตยกรรมดีปแลปวีสามพลัส (DeeplabV3+)

สถาปัตยกรรมดีปแลปวีสามพลัสจะเป็นการต่อยอดจากสถาปัตยกรรมรุ่นก่อนหน้า ที่มีความสามารถในการแก้ไขปัญหาการสูญเสียคุณลักษณะเมื่อผ่านการคำนวณในชั้นลึก, การเก็บรายละเอียดของคุณลักษณะได้ครบครัน, และการรองรับการมีอยู่ของวัตถุหลากหลายสเกล โดยสถาปัตยกรรมตัวนี้จะเพิ่มความสามารถในการตรวจจับเหลี่ยมมุมได้คมชัดขึ้น โดยอาศัยการกู้คืนข้อมูลเชิงพื้นที่แบบค่อยเป็นค่อยไป

สถาปัตยกรรมดีปแลปวีสามพลัส จะนำตัวเข้ารหัส และตัวถอดรหัส ที่ผ่านการดัดแปรโดยเพิ่มการทำคอนโวลูชันแบบหลุมมาใช้ในการแก้ไขปัญหาที่โมเดลไม่สามารถเก็บเหลี่ยมมุมได้ชัดเจน [32]

โดยทั่วไปนั้น โครงข่ายการเข้ารหัส และถอดรหัส จะประกอบไปด้วย

- โมเดลการเข้ารหัส ที่จะลดจำนวนแผนที่คุณลักษณะและตรวจจับข้อมูลการจำแนกรูปภาพเชิงความหมายแบบค่อยเป็นค่อยไป

- โมเดลการถอดรหัสที่ฟื้นฟูข้อมูลเชิงพื้นที่แบบค่อยเป็นค่อยไป

โครงข่ายการเข้ารหัสของสถาปัตยกรรมตัวนี้ จะนำแนวคิดโครงข่ายความฉลาดทางลึกแบบแยกกันได้ (Depth-wise separable convolution) [33] เพื่อเพิ่มประสิทธิภาพการคำนวณให้มากขึ้น

3.2 โมเดลการจำแนกรูปภาพเชิงความหมายตามเวลาจริง ที่ได้รับความนิยมในข้อมูลภาพถ่ายท้องถนนแบบเปิด

3.2.1 โมเดลสถาปัตยกรรมการจำแนกรูปภาพเชิงความหมายแบบเครือข่ายทวิภาคี (Bilateral Segmentation Network Architecture, BiSeNet) [18]

โดยทั่วไปแล้ว การจำแนกรูปภาพเชิงความหมายจะสกัดคุณลักษณะจากภาพ หรือวิดีโอที่นำเข้า โดยใช้องค์ประกอบสองอย่าง ประกอบไปด้วย

- ข้อมูลเชิงพื้นที่ที่มั่งคั่ง (Rich spatial information)
- เขตข้อมูลเปิดกว้างขนาดใหญ่ (Sizable receptive field)

อย่างไรก็ตาม วิธีการสมัยใหม่จะพยายามทำสององค์ประกอบดังกล่าว ให้รองรับการใช้งานตามเวลาจริง ทำให้ทั้งประสิทธิภาพของโมเดล และความแม่นยำ ตกลงเป็นอย่างมาก

ในงานวิจัยที่ผ่านมาในเชิงของสถาปัตยกรรมการจำแนกรูปภาพ แสดงให้เห็นถึงวิธีการที่จะช่วยเร่งความเร็วของการอนุมาน (Accelerate a Model Inference Speed) ดังต่อไปนี้

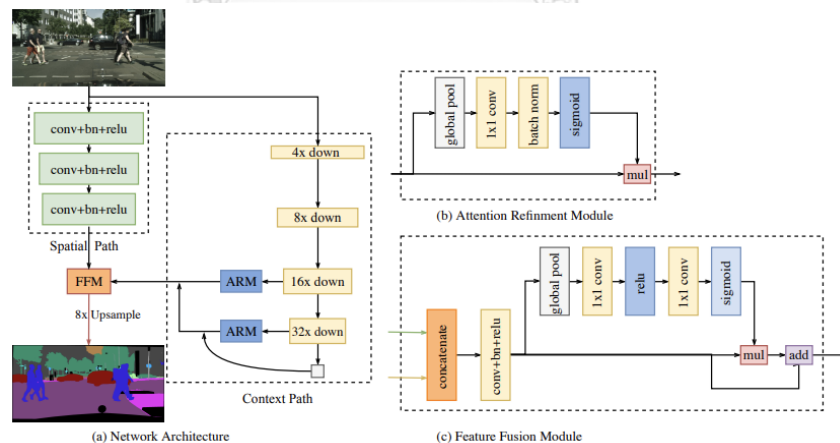
- จำกัดขนาดของภาพนำเข้า (Restrict image size) [34] กำหนดความสูง และความยาวของภาพนำเข้า ภาพใดก็ตามที่มีขนาดเกินกว่าค่าที่กำหนด จะถูกปรับขนาด หรือครอบตัด
- การตัดแต่งช่องสัญญาณ (Pruning channels) [35] เป็นวิธีการที่จะลดจำนวนช่องสัญญาณลงโดยตรง ทำให้โครงข่ายมีขนาดลดลง ทำให้การทำงาน ทั้งบนหน่วยประมวลผลกลาง และหน่วยประมวลผลด้านกราฟิก มีประสิทธิภาพเพิ่มขึ้นอย่างมาก

- การลดรายละเอียด (Downsampling) ในอีกความหมายหนึ่ง คือการลดจำนวนจุดสีของภาพนำเข้า ทำให้ภาพมีขนาดเล็กลง เมื่อผ่านการคำนวณในช่วงลดรายละเอียด

วิธีการข้างต้น ถึงแม้ว่าจะช่วยเพิ่มประสิทธิภาพในแง่ของความเร็วก็จริง แต่ก็ต้องแลกมาด้วยความแม่นยำที่ตกลงไป ด้วยเหตุผลดังกล่าว ทีมนักวิจัยจึงได้พยายามประยุกต์วิธีการที่จะช่วยรักษาข้อมูลเชิงพื้นที่จากข้อเสียข้างต้น โดยปรับปรุงโครงสร้างรูปตัวยู [12] ลงไปในแกนหลัก โครงสร้างดังกล่าวจะช่วยเพิ่มปริมาณข้อมูลเชิงพื้นที่และช่วยเพิ่มเติมข้อมูลที่สูญหาย อย่างไรก็ตาม การประยุกต์โครงสร้างรูปตัวยูลงไปจะมีข้อเสียสองประการ

- โครงสร้างรูปตัวยูแบบเต็มจะลดประสิทธิภาพด้านความเร็วลง เนื่องจากการสกัดข้อมูลต้องการการคำนวณเพิ่มเติมจากปกติ
- ข้อมูลเชิงพื้นที่ ที่สูญเสียจากการตัดแต่งช่องสัญญาณ ไม่สามารถกู้คืนได้โดยง่าย

ดังนั้น โมเดลสถาปัตยกรรมการจำแนกรูปภาพเชิงความหมายแบบเครือข่ายทวิภาคี จึงได้ปรับปรุงโมเดล โดยแยกออกมาเป็นสองกิ่ง ประกอบไปด้วยกิ่งเชิงพื้นที่ (Spatial path) และกิ่งบริบท (Context path)



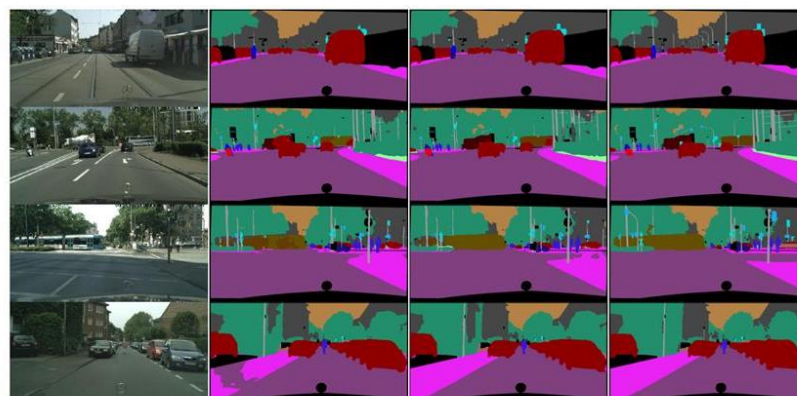
ภาพที่ 16 โมเดลไบร์เน็ต (อ้างอิงจาก fig 2, [18])

กิ่งเชิงพื้นที่ จะเป็นวิธีการที่จะรักษาไว้ ซึ่งข้อมูลเชิงบริบทของภาพนำเข้า กิ่งเชิงพื้นที่จะประกอบไปด้วย โครงข่ายคอนโวลูชัน สามชั้น แต่ละชั้นของโครงข่ายคอนโวลูชันจะมีอัตราการก้าวข้าม เป็น 2 ตามด้วยการทำชุดข้อมูลให้เป็นมาตรฐาน และหน่วยแก้ไขเส้นตรง (Rectified Linear

Unit, ReLU) โดยกึ่งเชิงพื้นที่จะให้ผลลัพธ์ออกมาเป็นแผนที่คุณลักษณะที่มีขนาดพื้นที่ น้อยกว่าภาพป้อนเข้า 8 เท่า โดยแผนที่ดังกล่าว จะอุดมไปด้วยข้อมูลเชิงบริบท หรือข้อมูลลักษณะเฉพาะชั้นต่ำ (low level feature) จำนวนมาก

กึ่งบริบท มีวัตถุประสงค์หลักเพื่อทำงานคู่ขนานกันไปกับกึ่งเชิงพื้นที่ ด้วยคุณสมบัติที่ถูกทำมาเพื่อให้เป็นโมเดลขนาดเบา (Lightweight) และมีการทำผลรวมค่าเฉลี่ยทั้งหมดในชั้นโครงข่ายคอนโวลูชันในชั้นสุดท้าย (Global average pooling) ทำให้ได้รับข้อมูลบริบทขนาดใหญ่ ยกตัวอย่างโมเดลที่ประสบความสำเร็จในงานวิจัย เช่นโมเดล อินเซ็ปชัน (Inception) [36] ในสถาปัตยกรรมดีปแลปวิสามพลัส (DeeplabV3+) [11] ที่ใช้เทคนิคการทำโครงข่ายความฉลาดทางลึกแบบแยกกันได้ โดยกึ่งบริบทจะมีหน้าที่มอบข้อมูลเชิงพื้นที่ หรือข้อมูลลักษณะเฉพาะชั้นสูง (high level feature) อย่างเพียงพอ ซึ่งข้อมูลดังกล่าวนี้จะส่งผลต่อความแม่นยำของโมเดลอย่างมีนัยสำคัญด้วยหลักการลดรายละเอียดอย่างรวดเร็ว

ในแง่ของการเพิ่มความแม่นยำยิ่งขึ้นนั้น นักวิจัยชาวจีน Changqian ได้เพิ่มโมเดลการรวมกันของรายละเอียด (Feature Fusion Module, FFM) เพื่อปรับปรุงประสิทธิภาพของรายละเอียดเพิ่มขึ้น และเพิ่มโมเดลการปรับแต่งจุดสนใจ (Attention Refinement Module, ARM) เพื่อเพิ่มน้ำหนักของรายละเอียดที่สำคัญในแต่ละชั้น โดยโมเดลการปรับแต่งจุดสนใจจะให้รายละเอียดของบริบทจากการทำผลรวมค่าเฉลี่ยทั้งหมดในกึ่งบริบท ซึ่งกระบวนการทั้งหมดจะปราศจากกระบวนการการเพิ่มรายละเอียด จึงไม่ต้องพึ่งทรัพยากรในการคำนวณกระบวนการนี้



(a) Image (b) U-Shape (c) BiSeNet (d) GT

ภาพที่ 17 ผลลัพธ์จากการทดลองโมเดลไบนารีเน็ตบนข้อมูลท้องถนน (อ้างอิงจาก fig 3, [18])

กล่าวโดยสรุป โมเดลสถาปัตยกรรมการจำแนกรูปภาพเชิงความหมายแบบเครือข่ายทวิภาคี จะพยายามหาจุดกึ่งกลาง ระหว่างความแม่นยำในการจำแนกรูปภาพเชิงความหมาย และความเร็วในการประมวลผล โดยมีการทดลองบนคลังข้อมูลภาพถ่ายท้องถนนแบบเปิด โดยเป็นคลังข้อมูลที่มีขนาดภาพ 2048 x 1024 จุดสี โมเดลดังกล่าวได้ความแม่นยำในการตรวจจับพื้นที่ (Intersect over Union) อยู่ที่ ร้อยละ 68.4 และมีความเร็วในการประมวลผล อยู่ที่ 105 ภาพต่อวินาที ซึ่งไวกว่าโมเดลที่เข้าแข่งขันทั้งหมดที่มีค่าความแม่นยำใกล้เคียงกัน บนคลังข้อมูลชุดนี้

### 3.3 การปรับภาพแคลช (CLAHE)

การปรับภาพแคลช เป็นอีกเทคนิคหนึ่งทางการมองเห็นของคอมพิวเตอร์ ที่ช่วยเพิ่มรายละเอียดของภาพ หรือวีดิโอผ่านการแยกสีอย่างชัดเจน (Image contrast) โดยจะเน้นไปที่การปรับสีของภาพให้กระจายไปในทุก ๆ ส่วนของขอบเขตสี ทำให้ภาพที่มีความเอนเอียงของย่านสีไปในทางใดทางหนึ่งมากเกินไป เช่น ภาพถ่ายในที่มืด ที่มีความเอนเอียงของสีในภาพค่อนข้างมืด กลับมาชัดเจนได้ โดยหลักการทำงานของแคลชพัฒนามาจาก การสมมูลฮิสโทแกรม (Histogram equalization) [37] ของนักวิจัย... โดยจะนำทุกจุดสีบนรูปมาจัดหาผลรวมฮิสโทแกรมของแต่ละย่านสี เรียงลำดับตามฟังก์ชันการกระจายสะสม (Cumulative distribution function, cdf) และทำการกระจายค่าของฮิสโทแกรมไปยังระนาบจุดสีดังสมการ

$$h(v) = \text{round} \left( \frac{cdf(v) - cdf_{min}}{(H \times W) - cdf_{min}} \times (L - 1) \right) \quad (12)$$

กำหนดให้

$v$  แทนค่าความหนาแน่นของจุดสี

$cdf_{min}$  แทนค่าที่น้อยที่สุดของฟังก์ชันการกระจายสะสมของจุดสี

$H$  แทนค่าความสูงของภาพ

$W$  แทนค่าความกว้างของภาพ

$L$  แทนค่า gray level ที่กำหนด (โดยทั่วไป จะกำหนดค่าอยู่ที่ 256)

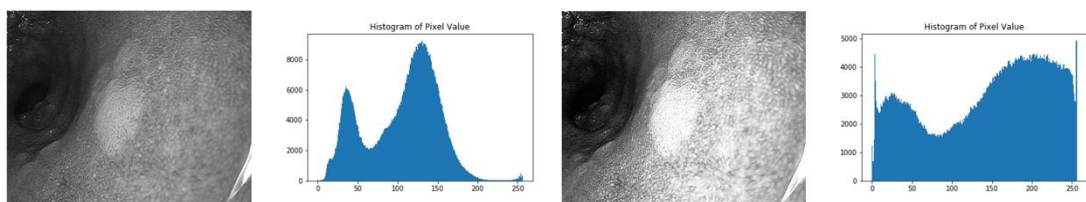


221	28	28	28	$v$	$cdf(v)$	$h(v)$
234	234	28	95	28	4	0
192	192	95	95	95	9	107
221	192	95	95	192	12	170
				221	14	213
				234	16	255

ภาพที่ 18 ตัวอย่างการคำนวณหาค่าสมดุสอิสรกรรมของภาพขนาด 4x4 จุดสี

221	28	28	28	213	0	0	0
234	234	28	95	255	255	0	107
192	192	95	95	170	170	107	107
221	192	95	95	213	170	107	107

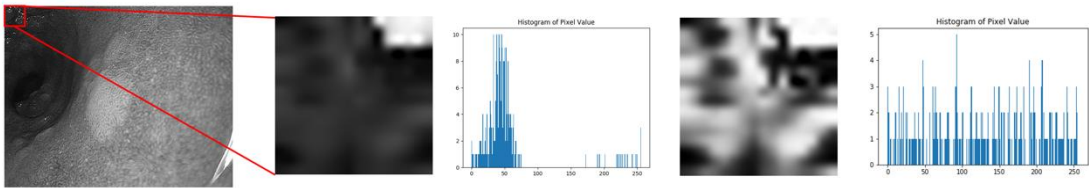
ภาพที่ 19 ผลลัพธ์ที่ได้จากการทำสมดุสอิสรกรรมของรูปตัวอย่าง



ภาพที่ 20 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการสมดุสอิสรกรรม

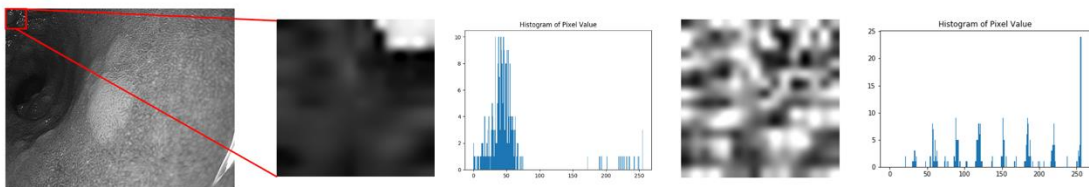
โดยนักวิจัย Stephen M.Pizer ได้นำวิธีการสมดุสอิสรกรรมจากที่เคยทำทั้งภาพ มาปรับปรุงแค่เพียงบางส่วนของรูปภาพ โดยจะไล่ทำที่ส่วนของภาพจนครบ หลักการแยกทำที่ละส่วนนี้ จะถูกเรียกว่า การสมดุสอิสรกรรมแบบยืดหยุ่น (Adaptive Histogram Equalized, AHE) [38]



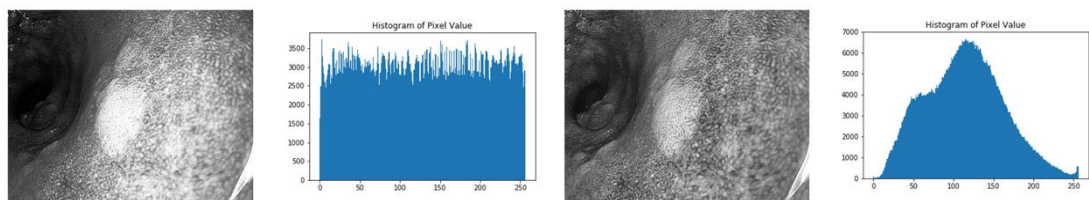


ภาพที่ 21 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการสมมูลฮิสโทแกรมแบบยืดหยุ่น

ในส่วนของการทำแคชชั่น จะเพิ่มตัวแปรเข้าไปในวิธีการการสมมูลฮิสโทแกรมแบบยืดหยุ่น เพื่อกำหนดค่าสูงสุดที่จะปรับปรุงภาพ โดยหลักการทำงานของแคชชั่นจะตัดค่าส่วนเกินออกจากฐานฮิสโทแกรมก่อนที่จำไปคำนวณค่าฟังก์ชันสะสม ทำให้แก้ไขปัญหาสัญญานรบกวน (Noise) ที่เกิดจากการเคลื่อนที่ที่เป็นสัญญาณรบกวนตรงขอบมุมของภาพที่มักจะมีจุดสีสีดำ ไปในการคำนวณฮิสโทแกรม จากการทดลอง รูปภาพที่ผ่านการปรับปรุงแคชชันในงานวิจัยชิ้นนี้ จะแยกย่านสีของบริเวณที่เป็นรอยโรค กับบริเวณปกติได้ชัดเจนขึ้น และเพิ่มค่าความแม่นยำของโมเดลอย่างมีนัยสำคัญ

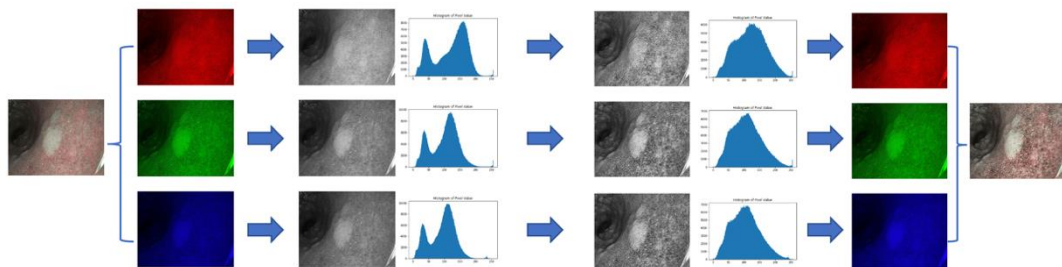


ภาพที่ 22 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกนำไปทำเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพปกติ ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุง โดยวิธีการแคช



ภาพที่ 23 ตัวอย่างภาพรอยโรคก่อนมะเร็งที่ถูกตัดแปลงไปเป็นภาพขาว-ดำ และกราฟฮิสโทแกรมแสดงถึงผลรวมของแต่ละจุดสีในภาพ โดยฝั่งซ้ายแสดงถึงภาพที่ผ่านการทำสมมูลฮิสโทแกรมแบบยืดหยุ่น ฝั่งขวาแสดงถึงภาพปกติหลังถูกปรับปรุงโดยวิธีการแคช

ในงานวิจัยชิ้นนี้ ได้ดัดแปลงการวิธีการทำแคลชมาใช้กับภาพสี โดยผู้จัดทำได้เสนอกระบวนการดังกล่าวผ่านช่องสัญญาณ RGB โดยจะนำภาพ มาแยกเป็นแต่ละช่องสัญญาณ R, G, และ B เพื่อนำไปเข้าสู่กระบวนการเข้ารหัส (Encoding) โดยทั้งสามช่องสัญญาณนั้น จะถูกแปลงให้กลายเป็นภาพขาว-ดำ โดยการตัดช่องสัญญาณที่เก็บค่า 0 ที่ ผลลัพธ์จากการเข้ารหัสจะได้ภาพขาว-ดำ ของแต่ละช่องสัญญาณ หลังจากนั้นจะนำภาพที่ได้ไปทำการถอดรหัส (Decoding) ผ่านกระบวนการการทำแคลชที่กล่าวไปข้างต้น ในขั้นตอนสุดท้ายจะเป็นการรวมแต่ละช่องสัญญาณเข้าด้วยกัน ทำให้ได้ภาพสีที่ผ่านการปรับปรุงแคลช และเห็นตำแหน่งของรอยโรคที่ชัดเจน

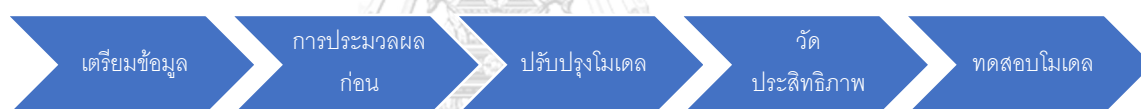


ภาพที่ 24 กระบวนการในการทำแคลชบนภาพสี โดยทำผ่านช่องสัญญาณ RGB

## บทที่ 4

### แนวคิดและวิธีดำเนินงาน

ในบทนี้ ผู้วิจัยต้องการพัฒนาโมเดลที่สามารถรองรับการทำงานตามเวลาจริง และยังคงไว้ซึ่งความแม่นยำในการทำนาย โดยมีขั้นตอนตามลำดับดังต่อไปนี้ (1) การเตรียมข้อมูล ข้อมูลดิบที่ทางผู้วิจัยได้รับนั้น ยังไม่สมบูรณ์และพร้อมใช้งานตั้งแต่วันที่ที่ได้รับ จึงจำเป็นต้องมีกระบวนการเตรียมข้อมูลเพื่อให้พร้อมในการใช้งาน ในที่นี้ จะประกอบไปด้วยการเตรียมข้อมูลของทั้งฝั่งผู้วิจัย และฝั่งแพทย์ (2) การประมวลผลก่อน โดยนำโมเดลจากงานแข่งขันภาพถ่ายท้องถนนแบบเปิด เพื่อใช้ปรับปรุงให้เข้ากับข้อมูลภาพถ่ายหัตถการการส่องกล้อง (3) ปรับปรุงโมเดล เพิ่มแนวคิด, วิธีการใหม่ ๆ ปรับปรุงให้โมเดลฐานมีความทันสมัย เพิ่มประสิทธิภาพในแง่ของความแม่นยำ (4) วัดประสิทธิภาพของโมเดล ทั้งในแง่ของความแม่นยำ และในแง่ของความเร็ว (5) ทดสอบโมเดล ด้วยการนำไปใช้งานจริง ที่ศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย



ภาพที่ 25 ขั้นตอนการดำเนินงาน

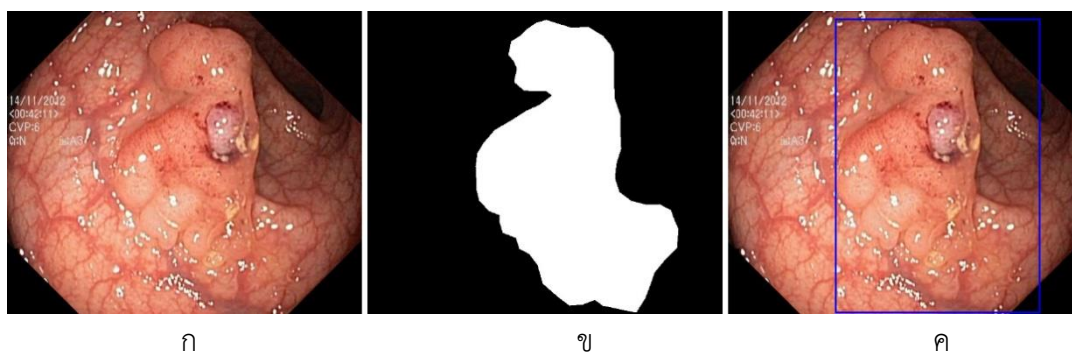
#### 4.1 การเตรียมข้อมูล (Data Preparation)

##### 4.1.1 คลังข้อมูลคควาเซอร์ (Kvasir Dataset)

คลังข้อมูลประกอบไปด้วยภาพตั้งเนื้อทางเดินอาหารส่วนล่าง จำนวน 1,000 ภาพ แต่ละภาพมีผลเฉลย มีขนาดหลากหลาย โดยมีขนาดเล็กที่สุด  $332 \times 487$  จุดสี ไปจนถึงใหญ่ที่สุด  $1920 \times 1072$  จุดสี ภาพถูกเข้ารหัสด้วยรหัสเจแปค (.jpeg)

##### 4.1.2 คลังข้อมูลซีวีซี (CVC-Colon Dataset)

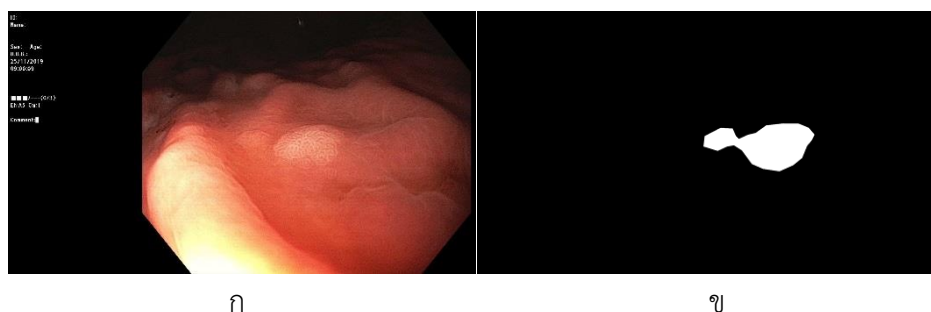
คลังข้อมูลประกอบไปด้วยภาพตั้งเนื้อทางเดินอาหารส่วนล่าง จำนวน 1,680 ภาพ แต่ละภาพมีผลเฉลย มีขนาดหลากหลาย โดยมีขนาด  $320 \times 480$  จุดสี ภาพถูกเข้ารหัสด้วยรหัสเจแปค (.jpeg)



ภาพที่ 26 ตัวอย่างภาพจากคลังข้อมูลคเวเซอร์ (ก) ภาพต้นฉบับ (ข) ผลเฉลย โดยพื้นที่สีขาวจะแทนบริเวณที่เป็นติ่งเนื้อ (ค) ผลเฉลย โดยบริเวณที่ถูกตีกรอบจะเป็นบริเวณที่เป็นติ่งเนื้อ (อ้างอิงจาก [21])

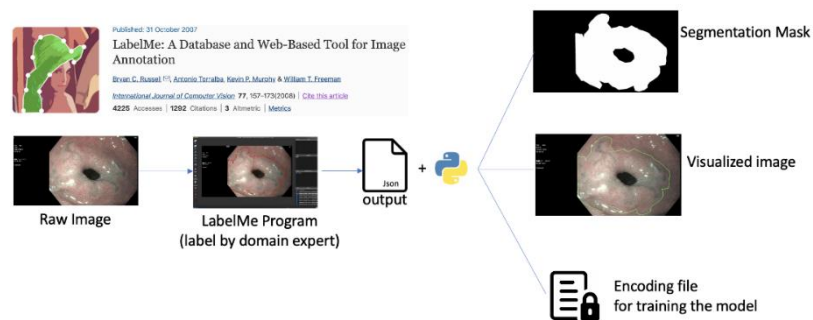
#### 4.1.3 คลังข้อมูลศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย

ผู้จัดทำได้รวบรวมข้อมูลจากคลังข้อมูลของโรงพยาบาลจุฬา โดยมีคนไข้ที่มีรอยโรคทั้งหมด 136 คน เริ่มเก็บข้อมูลตั้งแต่เดือนมกราคม 2559 จนถึงเดือนธันวาคม 2563 โดยคลังข้อมูลจะประกอบไปด้วยภาพติ่งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหาร จำนวน 802 ภาพ แต่ละภาพมีผลเฉลยที่ทางผู้วิจัยได้ร่วมมือกับทางทีมแพทย์ศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัยตลอดทั้งโครงการ มีขนาดของภาพอยู่ที่  $1920 \times 1080$  จุดสี โดยมีรุ่นกล้องส่องกระเพาะเป็น โอลิมปัส ชื่อรุ่น Olympus® EVIS EXERA III GIF-HQ190 gastroscope (Olympus Medical System Corps, Tokyo, Japan) ภาพถูกเข้ารหัสด้วยรหัสเจแปค มีชนิดของภาพทั้งหมดสองชนิด คือ ภาพชนิดแสงขาว (white light) และภาพชนิดเอ็นบีไอ (Narrow band image, NBI) [39] ในอัตราส่วน ร้อยละ 40 ต่อ 60



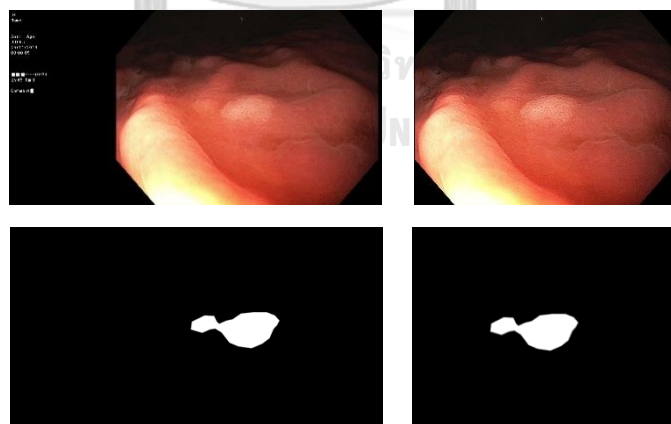
ภาพที่ 27 ตัวอย่างภาพจากคลังข้อมูลศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย (ก) ภาพต้นฉบับ (ข) ผลเฉลยที่ได้จากการใช้เครื่องมือเลเบลมี

ในการตีผลเฉลยบริเวณที่เป็นติ่งเนื้อชนิดเซลล์แบ่งตัวแบบผิดปกติในกระเพาะอาหารนั้น ทางผู้วิจัยได้เลือกใช้เครื่องมือ เลเบลมี (LabelMe) [40] ในการรวบรวมผลเฉลยจากคลังข้อมูลคนไข้ โดยตัวเครื่องมือจะสร้างผลเฉลยเป็นเอกสาร เจสัน (.json) ซึ่งสามารถนำมาเปลี่ยนแปลงให้เป็นข้อมูลผลเฉลยที่เป็นไฟล์ภาพได้ผ่านการเขียนโปรแกรมไพธอน



ภาพที่ 28 ขั้นตอนการเตรียมข้อมูลผ่านเครื่องมือ LabelMe

โดยภาพที่ได้จากการรวบรวมนั้น ยังคงมีพื้นที่ด้านซ้ายที่ไม่เกี่ยวข้องกับการทำหัตถการการส่องกล้อง ดังนั้น ก่อนที่จะนำภาพเหล่านี้ไปใช้ในการสอนโมเดล จะต้องครอบตัดบริเวณที่ไม่สนใจทิ้ง เพื่อเป็นการขจัดปัญหาโมเดลเรียนรู้พื้นที่ที่ไม่สนใจ



ก

ข

ภาพที่ 29 ขั้นตอนการครอบตัดบริเวณที่ไม่สนใจ (ก) ภาพต้นฉบับและผลเฉลย (ข) ภาพต้นฉบับและผลเฉลย หลังจากถูกครอบตัดแล้ว

#### 4.1.4 การกำหนดชุดตัวอย่างสอน

จะกำหนดโดยใช้อัตราส่วนหลักของ คลังฝึกสอน (Training Dataset) ต่อคลังตรวจสอบ ความถูกต้อง (Validating Dataset) ต่อคลังทดสอบ (Testing Dataset) เป็นอัตราส่วนร้อยละ 70 ต่อ 10 ต่อ 20 โดยคำนึงถึงคนไข้คนเดียวกัน จะถูกรวมอยู่ในคลังข้อมูลชุดเดียวกันเท่านั้น คลังข้อมูลชุดนี้ มีจำนวนภาพทั้งหมด 802 ภาพ แบ่งเป็น ภาพแสงขาว 307 ภาพ และภาพชนิดเอ็นบีไอ (Narrow Band Image, NBI) 495 ภาพ สำหรับการประเมินค่า Error นั้น ทางผู้วิจัยได้นำข้อมูล กระเพาะอาหารที่ไม่มีรอยโรค 160 ภาพ มาทดสอบหาค่าผิดพลาดจากการจำแนกส่วนเกินของโมเดล โดยข้อมูลที่ใช้จะมาจากคนไข้ปกติที่ไม่เจอโรคใด ๆ จากการทำหัตถการการส่องกล้อง โดยจำนวนรูปที่ใช้ทดสอบค่า Error จะเท่ากับกับจำนวนภาพบนบนชุดทดสอบ ในการรวบรวมผลลัพธ์การทดลอง ทางผู้วิจัยได้ใช้เทคนิค 3 folds cross validation เพื่อให้การทดลองมีความน่าเชื่อถือ และเพื่อป้องกันประสิทธิภาพที่แท้จริงของโมเดล

ในส่วนเทคนิคการทำ Hard negative sample จะมีการเติมภาพที่ไม่มีรอยโรคเพิ่มเข้ามา เพื่อฝึกสอนโมเดลร่วมกับภาพรอยโรค มีจำนวนภาพทั้งหมดก่อนทำการฝึกสอนทั้งสิ้น 838 ภาพ โดยภาพเหล่านี้ จะถูกคัดเพื่อนำไปฝึกสอน จำนวน 320 ภาพ โดยจะแบ่งเป็นชุดฝึกสอน จำนวน 280 ภาพ และชุดตรวจสอบ 40 ภาพ ทั้งนี้ ภาพที่ไม่มีรอยโรคที่เพิ่มเข้ามานั้น จะไม่มีความเกี่ยวข้อง (Independent) กับชุดทดสอบที่กล่าวไปข้างต้น

#### 4.2 การประมวลผลก่อน (Data Preprocessing)

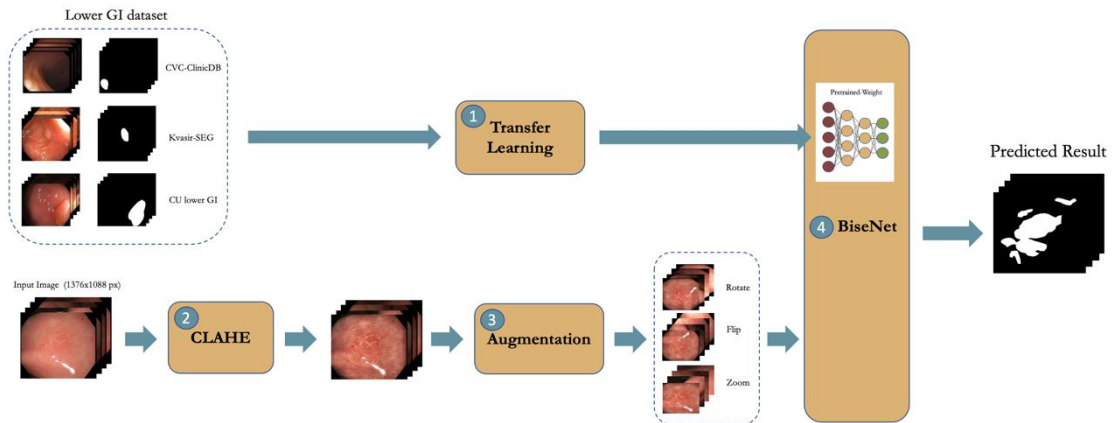
ในการประมวลผลก่อน ผู้จัดทำจะนำข้อมูลที่ผ่านการเตรียมข้อมูลจาก 4.1 มาเข้าสู่กระบวนการแคลช เพื่อเพิ่มรายละเอียดให้กับภาพ หลังจากนั้นจะนำภาพในชุดฝึกสอน และชุดตรวจสอบ ไปผ่านกระบวนการเพิ่มข้อมูล โดยใช้เทคนิคจาก 2.4 เพื่อเพิ่มปริมาณรูป โดยผู้จัดทำจะเพิ่มข้อมูลเป็นจำนวน 10 เท่า ก่อนนำไปฝึกสอนนั้น จะมีการแปลงภาพให้กลายเป็นเวกเตอร์โดยใช้กระบวนการปรับค่าปกติจาก 2.9 เพื่อให้การฝึกสอนเป็นไปอย่างราบรื่นและไม่เกิดข้อผิดพลาดจากการคำนวณที่มากเกินไป

#### 4.3 การปรับปรุงโมเดล (Model Implementation)

ในการปรับปรุงโมเดล ผู้จัดทำได้เพิ่มเทคนิคต่าง ๆ เพื่อเพิ่มความแม่นยำให้กับโมเดล โดยส่วนแรกจะปรับปรุงส่วนขาออกของโมเดล BiSeNet จากที่เคยทำนายออกมา 20 ความเป็นไปได้ตามข้อมูล cityscapes ให้ทำนายอยู่บนความเป็นไปได้เพียงสองอย่างเท่านั้น คือเป็นรอยโรค และไม่เป็นรอยโรค บนข้อมูลของโรงพยาบาลจุฬา ในส่วนถัดมาจะเพิ่มหลักการการทำแคลช การเพิ่มปริมาณ



ข้อมูล และการนำภาพปกติเข้ามาร่วมฝึกสอนโมเดลให้มีความแม่นยำขึ้น อ้างอิงจาก 2.4 2.5 3.3 ตามลำดับ



ภาพที่ 30 ภาพรวมกระแสนงานของโมเดล โดยตัวเลขจะแทนเทคนิคต่าง ๆ ที่ผู้จัดทำใช้ 1) การเรียนรู้แบบโอนถ่าย 2) การปรับภาพโดยแคลช 3) การเพิ่มข้อมูล และ 4) โมเดล BiSeNet

#### 4.4 การวัดประสิทธิภาพโมเดล (Model Evaluation)

จะใช้มาตรวัดสองค่าในการประเมินประสิทธิภาพของโมเดล คือมาตรวัดพื้นที่ทับซ้อนส่วนพื้นที่รวม เพื่อประเมินประสิทธิภาพในแง่ของความแม่นยำ และมาตรวัดความเร็วในการประมวลผลของภาพต่อวินาที เพื่อวัดประสิทธิภาพในแง่ของความเร็ว ดังที่กล่าวไว้ที่ 2.8

#### 4.5 การทดสอบโมเดล (Model Deployment)

ตัวโมเดลที่พัฒนาเสร็จแล้ว จะถูกนำไปติดตั้งเพื่อใช้งานจริง ที่ศูนย์ส่องกล้อง โรงพยาบาลจุฬาลงกรณ์มหาวิทยาลัย โดยจะมีแบบสอบถามความพึงพอใจในการใช้งานให้กับทีมแพทย์ เพื่อประเมินประสิทธิภาพในแง่ของการนำไปใช้จริง โดยจะมีการประเมิน และปรับปรุงโมเดลระหว่างช่วงทดสอบนี้เป็นระยะ

## บทที่ 5

### ผลการทดลอง

ในเนื้อหาบทนี้ จะแบ่งผลการทดลองออกเป็นสองส่วน โดยส่วนแรก 5.1 จะเป็นผลการทดลองบนชุดทดสอบทั้งสามชุด ในตารางที่ 1 จะกล่าวถึงผลการทดลองโดยละเอียด ในตารางที่ 2 จะกล่าวถึงค่าเฉลี่ยของแต่ละโมเดลบนชุดทดสอบทั้งสามชุด และตารางที่ 3 จะกล่าวถึงค่าความมั่นใจของโมเดลบนชุดทดสอบทั้งสามชุดที่ร้อยละ 95

ในส่วน 5.2 จะเป็นตัวอย่างผลลัพธ์ของชุดทดสอบ โดยจะเรียงลำดับตามตัวอักษร (ก) ภาพต้นฉบับ (ข) ผลเฉลย (ค) ผลลัพธ์จากโมเดล U-Net (ง) ผลลัพธ์จากโมเดล DeepLabV3 (จ) ผลลัพธ์จากโมเดล BiSeNet (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (TL) (ช) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (TL) + CLAHE (ซ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (TL) + CLAHE + Augmentation (AUG) (ฅ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (TL) + CLAHE + Augmentation (AUG) + Hard negative sample (HNS)

จากผลลัพธ์การทดลอง โมเดล BiSeNet ที่ผ่านกระบวนการประมวลผลก่อน จะได้ค่าเฉลี่ย IoU ที่สูงที่สุด เมื่อเทียบกับโมเดลตัวอื่น ๆ ที่ร้อยละ 54.87 ซึ่งมากกว่าตัว baseline ที่ร้อยละ 54.15 และตัว BiSeNet มาตรฐานที่ไม่ผ่านการปรับปรุง ที่ร้อยละ 50.15 โดยได้ค่า Error น้อยกว่าร้อยละ 1 ดังนั้น ในแง่ของประสิทธิภาพการจำแนกนี้ ตัวโมเดลของผู้จัดทำ สามารถทำได้ดี ในแง่ของการจำแนกรอยโรค และจำแนกบริเวณที่ปกติ อีกทั้งยังสามารถทำงานได้ตามเวลาจริง โดยที่ยังคงประสิทธิภาพไว้เทียบเท่ากับโมเดล baseline

ส่วนผลลัพธ์การคัดแยกนั้น โมเดล BiSeNet ที่ผ่านกระบวนการประมวลผลก่อน ได้ผลลัพธ์ค่า Sensitivity มากกว่าร้อยละ 90 แต่โมเดลนั้น ยังมีจุดอ่อนที่ค่า Specificity ยังคงได้น้อย เป็นผลให้ค่า F1 Score มีค่าที่น้อยตามไปด้วย ดังนั้น ทางผู้จัดทำจึงได้เพิ่มเทคนิคการใช้ภาพปกติที่ไม่มีรอยโรค ผสมผสานกับเทคนิคการเรียนรู้แบบกึ่งมีผู้สอน (Semi supervised) มาช่วยเพิ่มความแม่นยำมากขึ้น ในผลลัพธ์สุดท้าย วิธีการ Hard negative sample สามารถเพิ่มความแม่นยำในฝั่ง Specificity ที่สูงขึ้นมาก โดยสูญเสียความแม่นยำฝั่ง Sensitivity ลงไปเพียงเล็กน้อย ดังนั้น ค่าเฉลี่ย F1 Score ของโมเดลปรับปรุง ได้ร้อยละ 90.47 สูงกว่าทั้งโมเดล baseline ที่ร้อยละ 83.54 และโมเดล BiSeNet มาตรฐาน ที่ร้อยละ 84.06



ในงานวิจัยชิ้นนี้ ทางผู้จัดทำได้สรุปผลการทดลองในแต่ละประเด็นเพื่อทำการสรุปเทคนิคแต่ละหัวข้อ ประกอบไปด้วย ผลกระทบจากการใช้เทคนิคการเรียนรู้แบบโอนถ่าย ผลกระทบจากการใช้แคลช ผลกระทบจากการใช้เทคนิคการเพิ่มข้อมูล ผลกระทบจากการใช้ Hard negative sample และผลกระทบจากการใช้โมเดล BiSeNet

ผลกระทบจากการใช้เทคนิคการเรียนรู้แบบโอนถ่าย

ด้วยข้อมูลรอยโรคที่น้อยในคลังข้อมูล ทางผู้จัดทำจึงได้นำข้อมูลที่มีความใกล้เคียงกันที่มีเยอะกว่า อย่างข้อมูลระบบทางเดินอาหารส่วนล่าง ทั้งในงานแข่ง และข้อมูลจากศูนย์ส่องกล้อง ที่มีความคล้ายคลึงเรื่องโทนสีของภาพ และรุ่นกล้องที่ใช้ โดยการเพิ่มเดินภาพเข้าไปร่วมสอนก่อน (Pretrained) ทำให้โมเดลมีค่า negative predictive value เพิ่มขึ้นจากร้อยละ 87 เป็นสูงสุดร้อยละ 92 ในผลการทดลองแบบประมวลผลในขั้นตอนของการใส่ augmentation ร่วม และมีค่าเฉลี่ยสูงสุดที่ร้อยละ 80.42 สูงกว่าค่าเฉลี่ยของโมเดล baseline ทั้งสองตัว แต่ในทางเดียวกัน การใช้เทคนิคการเรียนรู้แบบโอนถ่าย กลับไม่ค่อยเพิ่มความแม่นยำในด้านของ Accuracy และ F1-Score มากนัก ถึงแม้ว่าการใช้เทคนิคนี้จะทำได้ดีในแง่ของการเพิ่ม Specificity ให้แก่โมเดล แต่ในทางกลับกัน การใช้เทคนิคนี้เพียงอย่างเดียวอาจไม่เพียงพอต่อค่าความแม่นยำที่คาดหวังไว้

ผลกระทบจากการใช้แคลช

การใช้แคลชได้ผลลัพธ์ที่ดีที่สุดในเรื่องของการเพิ่มความแม่นยำให้แก่ตัวโมเดล โดยประสิทธิภาพการจำแนกของโมเดลในค่า Sensitivity เพิ่มขึ้นจากร้อยละ 82 เป็นร้อยละ 89 และมี mIoU เพิ่มขึ้นจากร้อยละ 46 เป็นร้อยละ 55 ในชุดข้อมูลที่ 1 และได้ค่าเฉลี่ยรวมของ Sensitivity และ mIoU อยู่ที่ร้อยละ 91 และร้อยละ 95 สูงกว่าค่าเฉลี่ยของโมเดล Baseline โดยผลลัพธ์ที่ออกมานี้อาจเกิดขึ้นจากการที่โมเดลที่ผ่านการทำแคลชนั้นให้ข้อมูลที่มีความละเอียดเพิ่มขึ้น คล้ายกับการทำ NBI หรือ IEE ในขณะที่ส่องกล้อง ซึ่งการใช้เทคนิคนี้ ได้ความแม่นยำที่สูงมากพอที่จะไม่ต้องพึ่งพาโหมดแสงดำในระหว่างส่องกล้อง

ผลกระทบจากการใช้เทคนิคการเพิ่มข้อมูล

ทางผู้จัดทำได้ใช้เทคนิคนี้ในการเพิ่มข้อมูล 10 เท่าตัว โดยการสร้างข้อมูลใหม่ขึ้นมา นั้นจะมีพื้นฐานมาจากรูปในคลังข้อมูล โดยการเพิ่มข้อมูลมีส่วนช่วยในการเพิ่ม Sensitivity เมื่อเปรียบเทียบกับโมเดลก่อนหน้านี้ร้อยละ 4 (เพิ่มขึ้นจากร้อยละ 89 เป็นร้อยละ 93) และค่า mIoU เพิ่มขึ้นร้อยละ 2 (เพิ่มจากร้อยละ 55 เป็นร้อยละ 57)

ผลกระทบจากการใช้ Hard negative sample

ในขั้นตอนสุดท้าย ผู้จัดทำได้เพิ่มภาพถ่ายปกติเข้าไปในตัวโมเดล ผ่านการใช้เทคนิคแบบกึ่งมีผู้สอน เพื่อชดเชยค่า Specificity ที่ตกลงไปเป็นอย่างมาก โดยพยายามหาจุดสมดุลที่จะเพิ่มค่าความแม่นยำ และค่า F1-Score ให้มากขึ้น การนำเทคนิคสุดท้ายนี้เข้ามาใช้ ทำให้ค่าเฉลี่ย Accuracy และค่า F1-Score สูงขึ้นถึงร้อยละ 91 และร้อยละ 90 ตามลำดับ สูงกว่าค่าเฉลี่ยของโมเดลก่อนหน้าที่ทำได้เพียงร้อยละ 86 ทั้ง Accuracy และ F1-Score โดยการเพิ่มขึ้นมานี้ เกิดขึ้นจากการเพิ่มขึ้นของค่า specificity จากร้อยละ 80 เป็นร้อยละ 98 (ผลเฉลี่ย) โดยสูญเสีย Sensitivity ลงจากร้อยละ 91 เป็นร้อยละ 85 (ผลเฉลี่ย) สิ่งที่ได้เห็นได้เด่นชัดจากการปรับปรุงโดยเทคนิคนี้ คือการทำนายบริเวณที่ไม่ใช่รอยโรคตกลงเป็นอย่างมาก (อ้างอิงจากภาพถ่าย negative ในบทที่ 5) โดยสะท้อนถึงค่าเฉลี่ย Error ที่ลดลงจากร้อยละ 0.81 เป็นร้อยละ 0.08 ในชุดทดสอบทั้งสามชุด

ผลกระทบจากการใช้โมเดล BiSeNet

โมเดล BiSeNet เป็นโมเดลที่พยายามหาจุดสมดุล โดยตั้งเงื่อนไขว่าต้องทำงานตามเวลาจริงได้ และสูญเสียความแม่นยำให้น้อยที่สุด ดังนั้นแล้ว ผู้จัดทำจึงเลือกที่จะปรับปรุงโมเดลตัวนี้ โดยการเพิ่มเทคนิคดังที่กล่าวไปก่อนหน้านี้เพื่อชดเชยความแม่นยำที่สูญเสียจากการลดขนาดของโมเดลลงเมื่อเทียบกับโมเดล Baseline อย่าง โมเดล U-Net และ DeepLabV3+ ที่ถูกออกแบบมาให้จำแนกภาพโดยเฉพาะ โดยความเร็วในการประมวลผลโดยเฉลี่ยของโมเดลนั้นได้ผลลัพธ์ที่สูงมากกว่าโมเดล Baseline มากถึง 10-20 เท่า (จากผลเฉลี่ยความเร็วประเมินผลที่ร้อยละ 31.73 เทียบกับ 3.94 และ 2.35) ซึ่งเพียงพอต่อความต้องการขั้นต่ำที่ 25 เฟรมต่อวินาที

ด้วยวิธีการทั้งหมดนี้ ทำให้โมเดล Bisenet ที่ผ่านการปรับปรุง สามารถทำงานได้ตามเวลาจริง อีกทั้งยังมีความแม่นยำเทียบเท่าโมเดล Baseline ที่ได้รับความนิยมในงานแพทย์อีกด้วย

ตารางที่ 1 ผลการทดลองในแต่ละชุดการทดลอง

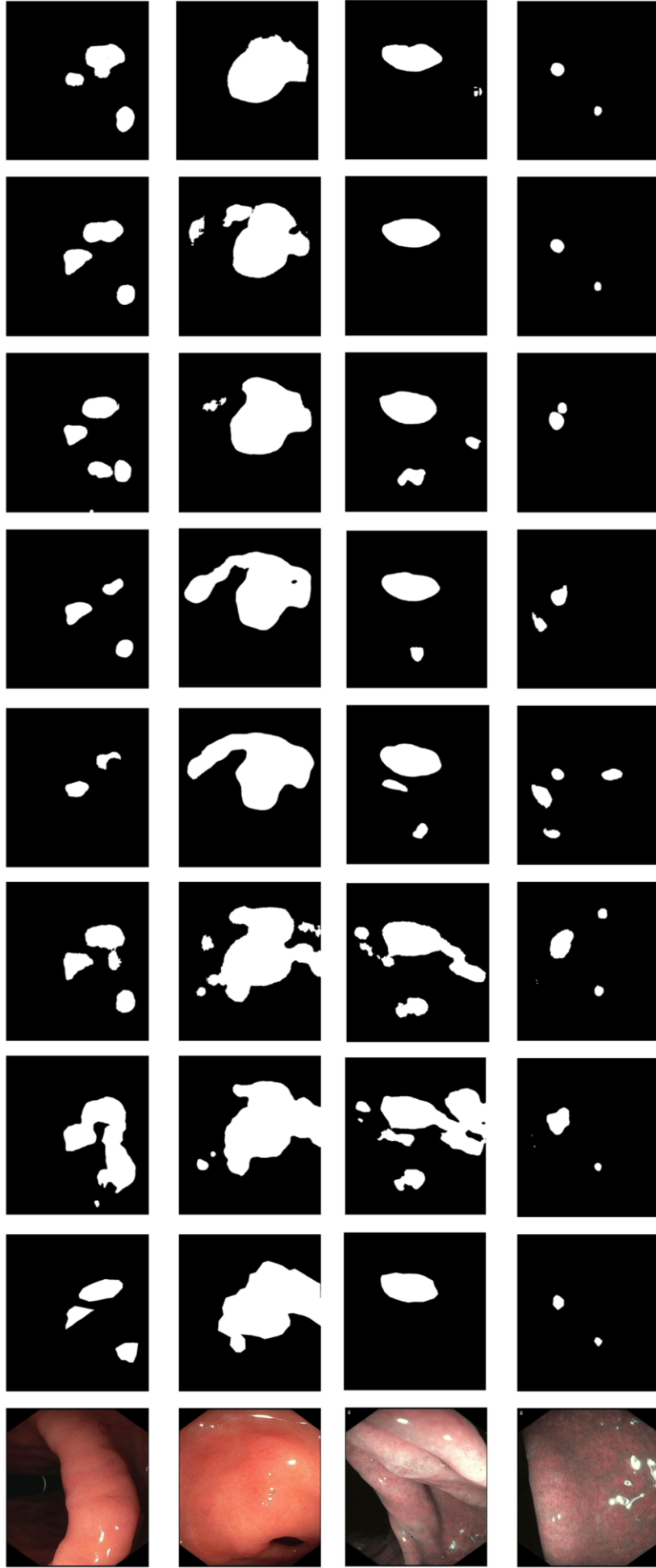
ชุดการทดลองที่ 1	TP	TN	FP	FN	SENS	SPEC	PPV	NPV	ACC	F1	IoU	Error	FPS
U-Net	140	100	60	20	0.8750	0.6250	0.7000	0.8333	0.7500	0.7778	0.5302	0.0181	3.49
DeepLabV3+	134	112	48	26	0.8375	0.7000	0.7365	0.8116	0.7688	0.7836	0.4922	0.0179	2.20
BiSeNet	131	140	20	29	0.8188	0.8750	0.8675	0.8284	0.8469	0.8424	0.4594	0.0046	34.02
BiSeNet+TL	128	147	13	32	0.8000	0.9188	0.9078	0.8212	0.8594	0.8505	0.4729	0.0033	33.33
BiSeNet+TL+CLAHE	143	118	42	17	0.8938	0.7375	0.7730	0.8741	0.8156	0.8290	0.5494	0.0097	31.83
BiSeNet+TL+CLAHE+AUG	149	128	32	11	0.9313	0.8000	0.8232	0.9209	0.8656	0.8739	0.5704	0.0096	31.53
BiSeNet+TL+CLAHE+AUG+HNS	138	157	3	22	0.8625	0.9813	0.9787	0.8771	0.9219	0.9169	0.5297	0.0006	31.78
ชุดการทดลองที่ 2	TP	TN	FP	FN	SENS	SPEC	PPV	NPV	ACC	F1	IoU	Error	FPS
U-Net	139	116	44	21	0.8688	0.7250	0.7596	0.8467	0.7969	0.8105	0.5433	0.0093	4.33
DeepLabV3+	135	102	58	25	0.8438	0.6375	0.6995	0.8031	0.7406	0.7649	0.5235	0.0127	2.34
BiSeNet	139	113	47	21	0.8688	0.7063	0.7473	0.8433	0.7875	0.8035	0.4980	0.0109	33.77
BiSeNet+TL	135	111	49	25	0.8438	0.6938	0.7337	0.8162	0.7688	0.7849	0.5064	0.0108	34.00
BiSeNet+TL+CLAHE	136	125	35	24	0.8500	0.7813	0.7953	0.8389	0.8156	0.8218	0.5240	0.0094	31.24
BiSeNet+TL+CLAHE+AUG	142	127	33	18	0.8875	0.7938	0.8114	0.8759	0.8406	0.8478	0.5436	0.0087	31.25
BiSeNet+TL+CLAHE+AUG+HNS	135	157	3	25	0.8438	0.9813	0.9783	0.8626	0.9125	0.9060	0.5189	0.0011	31.76
ชุดการทดลองที่ 3	TP	TN	FP	FN	SENS	SPEC	PPV	NPV	ACC	F1	IoU	Error	FPS
U-Net	151	142	18	9	0.9438	0.8875	0.8935	0.9404	0.9156	0.9179	0.5511	0.0041	3.99
DeepLabV3+	149	145	15	11	0.9313	0.9063	0.9085	0.9295	0.9188	0.9198	0.5350	0.0198	2.51
BiSeNet	148	136	24	12	0.9250	0.8500	0.8605	0.9189	0.8875	0.8916	0.5471	0.0048	33.74
BiSeNet+TL	148	128	32	12	0.9250	0.8000	0.8222	0.9143	0.8625	0.8706	0.5493	0.0070	33.61
BiSeNet+TL+CLAHE	143	110	50	17	0.8938	0.6875	0.7409	0.8661	0.7906	0.8102	0.5188	0.0102	31.01
BiSeNet+TL+CLAHE+AUG	145	130	30	15	0.9063	0.8125	0.8286	0.8966	0.8594	0.8657	0.5320	0.0059	31.52
BiSeNet+TL+CLAHE+AUG+HNS	131	157	3	29	0.8188	0.9813	0.9776	0.8441	0.9000	0.8912	0.5144	0.0007	31.64

ตารางที่ 2 ค่าเฉลี่ยของผลการทดลองของทั้งสามชุดการทดลอง

โมเดล	SENS	SPEC	PPV	NPV	ACC	F1	IoU	Error	FPS
U-Net	0.8958	0.7458	0.7844	0.8735	0.8208	0.8354	0.5415	0.0105	3.94
DeepLabV3+	0.8708	0.7479	0.7814	0.8481	0.8094	0.8228	0.5169	0.0168	2.35
BiSeNet	0.8708	0.8104	0.8251	0.8635	0.8406	0.8458	0.5015	0.0068	33.84
BiSeNet+TL	0.8563	0.8042	0.8212	0.8506	0.8302	0.8353	0.5095	0.0070	33.65
BiSeNet+TL+CLAHE	0.8792	0.7354	0.7697	0.8597	0.8073	0.8203	0.5307	0.0098	31.36
BiSeNet+TL+CLAHE+AUG	0.9083	0.8021	0.8211	0.8978	0.8552	0.8624	0.5487	0.0081	31.43
BiSeNet+TL+CLAHE+AUG+HNS	0.8471	0.9813	0.9782	0.8613	0.9115	0.9047	0.5210	0.0008	31.73

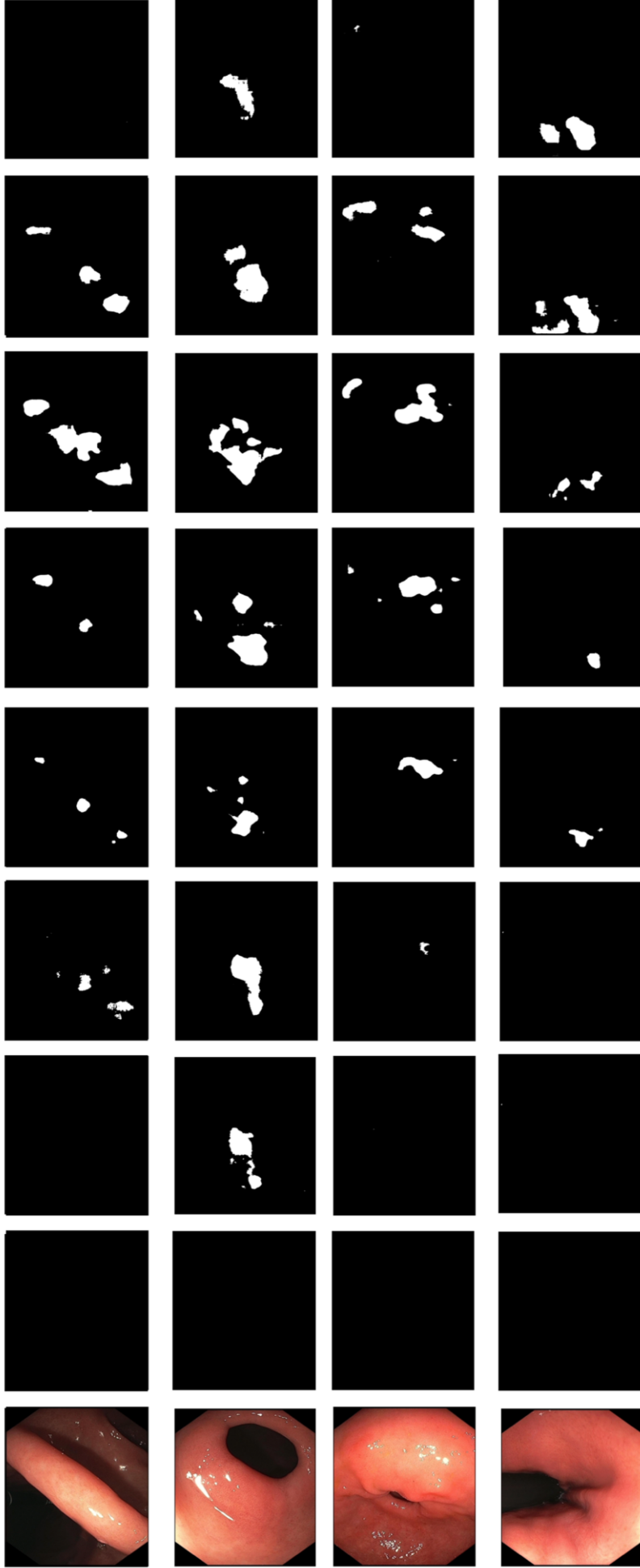
ตารางที่ 3 ค่าความมั่นใจ (Confidence Interval) ที่ร้อยละ 0.95 ของผลการทดลองทั้งสามชุดการทดลอง

โมเดล	SENS	SPEC	PPV	NPV	ACC	F1	IoU	Error	FPS
U-Net	0.0471	0.1499	0.1121	0.0660	0.0966	0.0830	0.0119	0.0080	0.4781
DeepLabV3+	0.0593	0.1591	0.1263	0.0799	0.1084	0.0957	0.0251	0.0042	0.1757
BiSeNet	0.0602	0.1031	0.0763	0.0549	0.0569	0.0500	0.0498	0.0041	0.1740
BiSeNet+TL	0.0718	0.1274	0.0985	0.0625	0.0603	0.0507	0.0433	0.0042	0.3808
BiSeNet+TL+CLAHE	0.0286	0.0531	0.0309	0.0209	0.0163	0.0107	0.0185	0.0004	0.4786
BiSeNet+TL+CLAHE+AUG	0.0248	0.0108	0.0099	0.0255	0.0147	0.0151	0.0223	0.0022	0.1798
BiSeNet+TL+CLAHE+AUG+HNS	0.0248	0.0000	0.0006	0.0187	0.0124	0.0146	0.0089	0.0003	0.0857



(ก) (ข) (ค) (ง) (จ) (ฉ) (ช) (ฉ) (ค) (ด) (ข) (ง) (ล)

ภาพที่ 31 ตัวอย่างผลลัพธ์จากการทดสอบตัวอย่างภาพที่มีรอยโรค บนคลังทดสอบ (ก) ภาพต้นฉบับ (ข) ผลเฉลย (ค) ผลลัพธ์จากโมเดล U-Net (ง) ผลลัพธ์จากโมเดล DeepLabV3 (จ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE (ช) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation +



(ก) (ข) (ค) (ง) (จ) (ฉ) (ฉ) (ฉ) (ฉ) (ฉ) (ฉ) (ฉ)

ภาพที่ 32 ตัวอย่างผลลัพธ์จากการทดสอบด้วยภาพที่ไม่มีรอยโรค บนคลังทดสอบ (ก) ภาพต้นฉบับ (ข) ผลเฉลี่ย (ค) ผลลัพธ์จากโมเดล U-Net (ง) ผลลัพธ์จากโมเดล DeepLabV3 (จ) ผลลัพธ์จากโมเดล BiSeNet (ฉ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning (ช) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE (ซ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation (ฌ) ผลลัพธ์จากโมเดล BiSeNet + Transfer learning + CLAHE + Augmentation + Hard negative sample

## บทที่ 6

### สรุปผลการทดลอง

เทคนิคการเรียนรู้เชิงลึก ในการตรวจจับ และวิเคราะห์ (Computer-aided detection & diagnosis, CADe/CADx) โรคต่าง ๆ ภายในทางเดินอาหารโดยใช้โครงข่ายประสาท ได้ถูกพัฒนาอย่างรวดเร็วในช่วงปี 2018 [41] โดยโมเดลส่วนใหญ่ได้เข้ามามีบทบาทสำคัญในการตรวจจับติ่งเนื้อ และโรคต่าง ๆ ในทางเดินอาหารส่วนล่างก่อนเป็นอันดับแรก โดยในปัจจุบัน โมเดลเหล่านี้สามารถทำงานได้ตามเวลาจริง และมีค่า Sensitivity มากถึงร้อยละ 90 [42]

มีงานวิจัยหลายชิ้นในแวดวงการแพทย์ ที่ใช้การเรียนรู้เชิงลึกในการตรวจจับรอยโรคภายในทางเดินอาหารส่วนบน แต่ส่วนใหญ่จะมุ่งเน้นไปทางพัฒนาโมเดลตรวจจับมะเร็งกระเพาะอาหาร [43-46] มีเพียงส่วนน้อยที่ตรวจจับรอยโรคก่อนมะเร็ง ผลงานที่ใกล้เคียงที่สุด โดยนักวิจัย Xu M. et al. ได้ใช้โมเดลการเรียนรู้เชิงลึก 4 ชนิด ประกอบไปด้วย ResNet-50 [47] VGG-16 [48] DenseNet-169 [49] EffcientNet-B4 [50] เพื่อตรวจจับรอยโรคระหว่างทำหัตถการ แต่การใช้งานจริงนั้น ทางแพทย์จะต้องหยุดกล้องชั่วคราวเพื่อให้โมเดลสามารถตรวจจับบริเวณที่มีรอยโรคในขณะนั้นได้ ซึ่งมีความยุ่งยากในทางปฏิบัติ [51] อีกทั้งยังมีข้อจำกัดของโมเดล ที่สามารถใช้งานได้กับภาพแสงดำเท่านั้น ทำให้ไม่สามารถใช้งานได้จริงในกรณีที่การส่องกล้องส่วนใหญ่ จะใช้แสงขาวในการทำหัตถการ

โมเดล BiSeNet นั้น ถือว่าเป็นโมเดลที่สามารถใช้งานได้ตามเวลาจริง และยังคงไว้ซึ่งความแม่นยำ โดยการใช้โมเดล BiSeNet แบบดั้งเดิมนั้น ได้ความเร็วในการประมวลผลมากกว่าโมเดล Baseline U-Net และ DeeplabV3+ มากถึง 10 เท่า แต่ในแง่ความแม่นยำของการตัดแยกนั้น BiSeNet กลับทำได้เพียงร้อยละ 86 เท่านั้น ซึ่งไม่เพียงพอต่อค่ามาตรฐานของ Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI) ที่กำหนดเอาไว้ว่า โมเดลที่ดีนั้น ควรจะมีค่า NPV ร้อยละ 90 [52] ดังนั้น การเพิ่มเทคนิคต่าง ๆ เข้าไปนั้น จึงจำเป็นในการบรรลุค่ามาตรฐานนี้ โดยผลลัพธ์ท้ายสุด โมเดล BiSeNet + Transfer learning + CLAHE + AUG ได้ผลลัพธ์ NPV อยู่ร้อยละ 92.09 ซึ่งมากกว่าค่ามาตรฐาน PIVI กำหนดไว้ อีกทั้งค่าเฉลี่ย NPV ที่ได้จากการทดลองทั้ง 3 ชุดข้อมูล ร้อยละ 89.78 ซึ่งใกล้เคียงกับค่ามาตรฐาน PIVI

ในส่วนของการจำแนก โมเดล BiSeNet ของทางผู้จัดทำ ได้ผลลัพธ์ค่าเฉลี่ย IoU อยู่ที่ร้อยละ 52.10 ซึ่งเพียงพอต่อการบอกขอบเขตคร่าว ๆ ในระหว่างการทำหัตถการ โดยเป้าหมายหลักของ โมเดลจะเป็นการเตือนทีมแพทย์ให้ไม่พลาดเมื่อกล้องทำการจับรอยโรคตามเวลาจริง

งานวิจัยชิ้นนี้ แสดงให้เห็นว่า การเพิ่มเทคนิคต่าง ๆ เข้าไปในตัวโมเดล BiSeNet ดั้งเดิม เช่น การทำการเรียนรู้แบบโอนถ่าย การทำแคลช การเพิ่มข้อมูล และการใช้การเรียนรู้แบบกึ่งผู้สอนนั้น ทำให้โมเดลที่ผ่านการปรับปรุง ได้ค่าเฉลี่ย F1-Score ในชุดทดสอบทั้งสามชุด มากกว่าร้อยละ 90 และยังสามารถใช้งานได้จริง โดยมีประสิทธิภาพการแสดงผล มากกว่า 25 เฟรมต่อวินาที ในอนาคต นั้น งานวิจัยชิ้นนี้สามารถต่อยอดได้ในอีกสองทาง หนึ่งคือ สามารถนำผลลัพธ์จากการตัดแยก มาช่วย คัดเลือกภาพที่จะนำไปจำแนกต่อ เพื่อเพิ่มประสิทธิภาพในการประมวลผลด้านความเร็วให้มากยิ่งขึ้น โดยไม่จำเป็นต้องจำแนกทุกเฟรมแบบปกติ สองคือ การนำค่าสูญเสียจากการทำการตัดแยก มาเฉลี่ย รวมกับการจำแนก เพื่อสะท้อนค่าที่แท้จริงก่อนที่จะนำไปทำกระบวนการย้อนกลับ (Back propagation) เพื่อให้ค่าสูญเสียที่เกิดขึ้น มีความเที่ยงตรง และอาจช่วยเพิ่มความแม่นยำให้กับโมเดล

ประโยชน์ของงานวิจัยชิ้นนี้ ตัวโมเดลของผู้จัดทำสามารถตรวจจับรอยโรคได้ตามเวลาจริง ไม่ ว่าจะเป็นภาพแสงขาว หรือแสงดำ อีกทั้งยังคงไว้ซึ่งความแม่นยำเทียบเท่ากับโมเดล benchmark ทางผู้จัดทำคาดหวังว่า งานวิจัยชิ้นนี้ จะช่วยเพิ่มความแม่นยำให้แก่แพทย์ฝึกหัด ให้สามารถตรวจจับ รอยโรคได้เทียบเท่าแพทย์ผู้ชำนาญ โดยไม่จำเป็นต้องเปลี่ยนเป็นแสงดำในการตรวจจับ อีกทั้งยังเป็น แนวทางใหม่ ๆ ให้กับกรวิจัยการเรียนรู้เชิงลึก ในระบบทางเดินอาหารส่วนบนต่อไป



## บรรณานุกรม

1. Fox, J.G. and T.C. Wang, *Inflammation, atrophy, and gastric cancer*. J Clin Invest, 2007. 117(1): p. 60-9.
2. Lim, J.H., et al., *Correlation between Endoscopic and Histological Diagnoses of Gastric Intestinal Metaplasia*. Gut Liver, 2013. 7(1): p. 41-50.
3. Panteris, V., et al., *Diagnostic capabilities of high-definition white light endoscopy for the diagnosis of gastric intestinal metaplasia and correlation with histologic and clinical data*. Eur J Gastroenterol Hepatol, 2014. 26(6): p. 594-601.
4. Pittayanon, R., et al., *The learning curve of gastric intestinal metaplasia interpretation on the images obtained by probe-based confocal laser endomicroscopy*. Diagn Ther Endosc, 2012. 2012: p. 278045.
5. Dixon, M.F., et al., *Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994*. Am J Surg Pathol, 1996. 20(10): p. 1161-81.
6. Ang, T.L., et al., *A multicenter randomized comparison between high-definition white light endoscopy and narrow band imaging for detection of gastric lesions*. Eur J Gastroenterol Hepatol, 2015. 27(12): p. 1473-8.
7. Savarino, E., et al., *Narrow-band imaging with magnifying endoscopy is accurate for detecting gastric intestinal metaplasia*. World J Gastroenterol, 2013. 19(17): p. 2668-75.
8. Wu, C.C.H., et al., *A prospective randomized tandem gastroscopy pilot study of linked color imaging versus white light imaging for detection of upper gastrointestinal lesions*. J Gastroenterol Hepatol, 2021.
9. Sun, M., et al., *Accurate Gastric Cancer Segmentation in Digital Pathology Images Using Deformable Convolution and Multi-Scale Embedding Networks*. IEEE Access, 2019. 7: p. 75530 - 75541.
10. Li, Y., et al. *GT-Net: A Deep Learning Network for Gastric Tumor Diagnosis*. in 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). 2018.

11. Chen, L.-C., et al. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. in *Computer Vision – ECCV 2018*. 2018. Cham: Springer International Publishing.
12. Ronneberger, O., P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015. Cham: Springer International Publishing.
13. Read, P. and M.-P. Meyer, *Restoration of motion picture film*. 1 ed. 2000: Butterworth-Heinemann.
14. Rodriguez-Diaz, E., et al., *Real-time artificial intelligence-based histologic classification of colorectal polyps with augmented visualization*. *Gastrointest Endosc*, 2021. 93(3): p. 662-670.
15. Wang, C., et al. *Localizing and Identifying Intestinal Metaplasia Based on Deep Learning in Oesophagoscope*. in *2019 8th International Symposium on Next Generation Electronics (ISNE)*. 2019.
16. Sun, X., et al. *Colorectal Polyp Segmentation by U-Net with Dilation Convolution*. in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. 2019.
17. Jason, Y., et al., *How transferable are features in deep neural networks?*, in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. 2014, MIT Press: Montreal, Canada. p. 3320–3328.
18. Yu, C., et al. *BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation*. in *Computer Vision – ECCV 2018*. 2018. Cham: Springer International Publishing.
19. Cordts, M., et al. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
20. Bernal, J., et al., *WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians*. *Comput Med Imaging Graph*, 2015. 43: p. 99-111.

21. Jha, D., et al., *Kvasir-SEG: A Segmented Polyp Dataset*. 2019, Springer, Cham. p. 451-462.
22. Zuiderveld, K. *Contrast Limited Adaptive Histogram Equalization*. in *Graphics Gems*. 1994. Elsevier.
23. Lecun, Y., et al., *Gradient-based learning applied to document recognition*. *Proceedings of the IEEE*, 1998. 86(11): p. 2278-2324.
24. Andrew Ng, Kian Katanforoosh, and Younes Mourri. *Deep Learning Specialization*. 2020; Available from: <https://www.coursera.org/specializations/deep-learning>.
25. Buslaev, A., et al., *Albumentations: Fast and Flexible Image Augmentations*. *Information*, 2020. 11(2).
26. Deng, J., et al., *ImageNet: a Large-Scale Hierarchical Image Database*. 2009. 248-255.
27. Lin, T.-Y., et al., *Microsoft COCO: Common Objects in Context*. 2014.
28. Everingham, M., et al., *The Pascal Visual Object Classes (VOC) challenge*. *International Journal of Computer Vision*, 2010. 88: p. 303-338.
29. Hashimoto, R., et al., *Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video)*. *Gastrointest Endosc*, 2020. 91(6): p. 1264-1271 e1.
30. Morales, G., et al., *Automatic Segmentation of Mauritia flexuosa in Unmanned Aerial Vehicle (UAV) Imagery Using Deep Learning*. *Forests*, 2018. 9: p. 736.
31. Chen, L.-C., et al., *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016. PP.
32. Chen, L.-C., et al., *Rethinking Atrous Convolution for Semantic Image Segmentation*. 2017.
33. Chollet, F., *Xception: Deep Learning with Depthwise Separable Convolutions*. 2017. 1800-1807.
34. Zhao, H., et al., *ICNet for Real-Time Semantic Segmentation on High-Resolution Images*. 2017.

35. Paszke, A., et al., *ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation*. 2016.
36. Szegedy, C., et al., *Going Deeper with Convolutions*. 2014.
37. Hummel, R., *Image enhancement by histogram transformation*. *Computer Graphics and Image Processing*, 1977. 6(2): p. 184-195.
38. Pizer, S.M., et al., *Adaptive histogram equalization and its variations*. *Computer Vision, Graphics, and Image Processing*, 1987. 39(3): p. 355-368.
39. Gono, K., *Narrow Band Imaging: Technology Basis and Research and Development History*. *Clin Endosc*, 2015. 48(6): p. 476-80.
40. Wada, K., *labelme: Image Polygonal Annotation with Python*. 2016.
41. Goncalves, W.G.E., et al., *Deep learning in gastric tissue diseases: a systematic review*. *BMJ Open Gastroenterol*, 2020. 7(1): p. e000371.
42. Mori, Y., et al., *Artificial intelligence in colonoscopy - Now on the market. What's next?* *J Gastroenterol Hepatol*, 2021. 36(1): p. 7-11.
43. Hirasawa, T., et al., *Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images*. *Gastric Cancer*, 2018. 21(4): p. 653-660.
44. Suzuki, H., et al., *Artificial intelligence for cancer detection of the upper gastrointestinal tract*. *Dig Endosc*, 2021. 33(2): p. 254-262.
45. Li, L., et al., *Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging*. *Gastric Cancer*, 2020. 23(1): p. 126-132.
46. Zhang, L., et al., *Diagnosis of gastric lesions through a deep convolutional neural network*. *Dig Endosc*, 2020.
47. He, K., et al. *Deep Residual Learning for Image Recognition*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
48. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, in *2015 International Conference on Learning Representations (ICLR)*. 2015.
49. Huang, G., et al. *Densely Connected Convolutional Networks*. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

50. Tan, M. and Q. Le, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, in *Proceedings of the 36th International Conference on Machine Learning*, C. Kamalika and S. Ruslan, Editors. 2019, PMLR: Proceedings of Machine Learning Research. p. 6105-6114.
51. Xu, M., et al., *Artificial intelligence in diagnosis of gastric precancerous conditions by image-enhanced endoscopy: a multicenter, diagnostic study (with video)*. *Gastrointest Endosc*, 2021.
52. Committee, A.T., et al., *ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps*. *Gastrointest Endosc*, 2015. 81(3): p. 502 e1-502 e16.





จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## ประวัติผู้เขียน

ชื่อ-สกุล	Vitchaya Siripoppohn
วัน เดือน ปี เกิด	19 December 1995
สถานที่เกิด	Ramathibodi hospital, Bangkok
วุฒิการศึกษา	Actuarial Science, Mahidol Univesity.
ที่อยู่ปัจจุบัน	147/10, Shanthiphap village, Chohare road, Muangmor sub-district, Muang District, Phrae, 54000.



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY