



การเรียนรู้เชิงลึกโดยโครงข่ายประสาทเทียมคอนโวลูชันสำหรับกำหนดขอบเขตอัตโนมัติ
บนอวัยวะที่มีความเสี่ยงของมะเร็งศีรษะและลำคอ



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตร์มหาบัณฑิต
สาขาวิชาฟิสิกส์การแพทย์
ปีการศึกษา 2563
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การเรียนรู้เชิงลึกโดยโครงข่ายประสาทเทียมคอนโวลูชันสำหรับกำหนดขอบเขตอัตโนมัติ
บนอวัยวะที่มีความเสี่ยงของมะเร็งศีรษะและลำคอ



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาฟิสิกส์การแพทย์
ปีการศึกษา 2563
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง "การเรียนรู้เชิงลึกโดยโครงข่ายประสาทเทียมคอนโวลูชันสำหรับกำหนดขอบเขต

อัตโนมัติบนอวัยวะที่มีความเสี่ยงของมะเร็งศีรษะและลำคอ"

ของ รัฐพงษ์ นันทาแพร์

ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต สาขาวิชาฟิสิกส์การแพทย์

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการสอบวิทยานิพนธ์

(ดร.ทวีป แสงแห่งธรรม)

..... ประธานที่ปรึกษาวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.ฐิติพงศ์ แก้วเหล็ก)

..... กรรมการผู้ทรงคุณวุฒิภายใน

(ผู้ช่วยศาสตราจารย์ ดร.นันทวัฒน์ อูดี)

อนุมัติ

.....
(ศาสตราจารย์ ดร.ไพศาล มุณีสว่าง)

คณบดีบัณฑิตวิทยาลัย

| | |
|------------------------|---|
| ชื่อเรื่อง | การเรียนรู้เชิงลึกโดยโครงข่ายประสาทเทียมคอนโวลูชันสำหรับ กำหนดขอบเขตอัตโนมัติบนอวัยวะที่มีความเสี่ยงของมะเร็งศีรษะและ ลำคอ |
| ผู้วิจัย | รัฐพงษ์ นันทาแพร์ |
| ประธานที่ปรึกษา | ผู้ช่วยศาสตราจารย์ ดร. จูติพงษ์ แก้วเหล็ก |
| ประเภทสารนิพนธ์ | วิทยานิพนธ์ วท.ม. สาขาวิชาฟิสิกส์การแพทย์, มหาวิทยาลัยนเรศวร, 2563 |
| คำสำคัญ | การเรียนรู้เชิงลึก, ภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษา, มะเร็ง ศีรษะและลำคอ, สถาปัตยกรรมยูเน็ต, สถาปัตยกรรมพีเจอร์พีระมิต เน็ตเวิร์ค (เอฟพีเอ็น) |

บทคัดย่อ

การกำหนดขอบเขตอวัยวะที่มีความเสี่ยงในผู้ป่วยมะเร็งศีรษะและลำคอที่รักษาด้วยวิธีการฉายรังสี มีข้อจำกัด เช่น ใช้เวลากำหนดขอบเขตนาน มีความแปรปรวนภายในและระหว่างแพทย์ผู้กำหนดขอบเขตรวมถึงประสบการณ์ของแพทย์ ดังนั้นการศึกษานี้จึงมีจุดมุ่งหมายเพื่อพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ ประเมินผลการทำนายขอบเขต และวิเคราะห์ปัจจัยที่มีผลต่อการทำนายของโมเดลที่พัฒนาขึ้น ดำเนินการพัฒนาโมเดลบนพื้นฐานรูปแบบสถาปัตยกรรมยูเน็ต และพีเจอร์พีระมิตเน็ตเวิร์ค ร่วมกับโมเดลแบคโบน 4 โมเดล ได้แก่ วีจีจี19 เรสเน็ต50 ซี-เรสเน็ต50 และอินเซปชันวี3 ใช้รูปแบบการสอน 2 วิธี ได้แก่ วิธีปรับจูน และวิธีถ่ายทอดการเรียนรู้ สอนโมเดลโดยใช้ข้อมูลภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษา และข้อมูลการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงจากฐานข้อมูลทิสไอเอ จำนวน 102 ชุดข้อมูล ประเมินประสิทธิภาพการทำนายของโมเดลด้วยค่าสัมประสิทธิ์ความเหมือน (ดีเอสซี) ค่าระยะห่าง 95 เปอร์เซ็นไทล์ฮาวดรอฟ (95%เอชดี) ประเมินเวลาสอนโมเดล เวลาทำนายมาส์ก และวิเคราะห์ปัจจัยที่มีผลต่อประสิทธิภาพการทำนายของโมเดลการเรียนรู้เชิงลึกที่พัฒนาขึ้น ผลการประเมินพบว่าโมเดลที่ให้ผลการทำนายที่ดีในแต่ละรูปแบบโมเดลแบคโบน ได้แก่ โมเดลทียูวีจีจี19 เอฟยูเรสเน็ต50 เอฟเอฟซี-เรสเน็ต50 และเอฟยูอินเซปชันวี3 ผลการประเมินค่าดีเอสซี มากกว่า 0.80 ในอวัยวะที่มีความเสี่ยงก้านสมอง ดวงตา ทั้ง 2 ข้าง กระดูกขากรรไกรล่าง และมีค่าดีเอสซี มากกว่า 0.68 ใน ต่อม้ำลายข้างกหนูทั้ง 2 ข้าง และไขสันหลัง ผลการประเมินค่า 95%เอชดี น้อยกว่า 2 มิลลิเมตร ในอวัยวะที่มีความเสี่ยงก้านสมอง

ดวงตาทั้ง 2 ข้าง ไซส์นหลัง และค่า 95%เอชดี น้อยกว่า 4.1 มิลลิเมตร ในอวัยวะที่มีความเสี่ยง กระจกขากรรไกรล่าง และต่อมน้ำลายข้างกกหูทั้ง 2 ข้าง เวลาทำนายมาस्कเฉลี่ยเท่ากับ 1.286, 1.348, 1.370 และ 1.403 วินาทีต่อภาพ สำหรับ โมเดลที่ยูวีจีจี19 เอพยูเรสเน็ต50 เอพเอฟซี-เรส เน็ต50 และเอพยูอินเซปชันวี3 ตามลำดับ ปัจจัยที่มีผลต่อการทำนายของโมเดลที่พัฒนาขึ้นได้แก่ คุณภาพของภาพซีทีสำหรับสอนโมเดล ความถูกต้องของขอบเขตมากส์อ้างอิงในชุดข้อมูลสอน และ ปริมาณข้อมูลสำหรับสอนโมเดล



| | |
|-----------------------|--|
| Title | DEEP LEARNING WITH CONVOLUTIONAL NEURAL NETWORK FOR AUTO-DELINEATION ON ORGANS AT RISK OF HEAD AND NECK CANCER |
| Author | RUTTHAPONG NANTAPRAE |
| Advisor | Assistant Professor Titipong Kaewlek, Ph.D. |
| Academic Paper | Thesis M.S. in Medical Physics, Naresuan University, 2020 |
| Keywords | Deep Learning, CT simulation image, Head and Neck cancer, U-Net architecture, Feature Pyramid Network (FPN) architecture |

ABSTRACT

The limitation of organs at risk (OARs) delineation process in head and neck (H&N) cancer patient, who has been treated by radiotherapy. For instance, it is time-consuming, the variation of intra- and inter-observer and depending on the experience of radiation oncologist's delineated. Consequently, this study aims to develop the deep learning model for auto-delineation of OARs on CT simulation images of H&N cancer patient, to evaluate the prediction results, and to analyze the factors affecting the prediction of the models were developed. The model development base on U-Net and Feature pyramid network architectures were modified by 4 backbone models: VGG19, ResNet50, SE-ResNet50, and InceptionV3. Training models used 2 training methods: transfer learning and fine-tune methods. The CT simulation image and DICOM RT contour file total of 102 dataset collected from The cancer image archive database used to train the models. The model performances were evaluated by Dice Similarity Coefficients (DSC), 95th percentile Hausdorff Distance (95%HD), models training time, models prediction time, and analysis result of factors that affected the accuracy of the model prediction. The results showed that the best models were classified by each backbone models: TUVGG19, FUResNet50, FFSE-ResNet50, and FUIncV3. The results of averages DSC value were more than 0.8 for brain stem, left and right eye globes, and mandible.

more than 0.68 for the left and right parotid glands and spinal cord. The result of averages 95%HD value were less than 2 millimeters for brain stem, left and right eye globes, and spinal cord. Less than 4.1 millimeters for the mandible, left and right parotid glands. The result for averages time of model predictions were 1.286, 1.348, 1.370, and 1.403 seconds per image for the TUVGG19, FUResNet50, FFSE-ResNet50, and FUIncV3 models, respectively. Factors affected the model's prediction included CT image quality for training models, the accuracy of training mask reference, and the number of the dataset for training should be large.



ประกาศคุณูปการ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ข้าพเจ้าขอขอบคุณ

ผู้ช่วยศาสตราจารย์ ดร.ฐิติพงศ์ แก้วเหล็ก อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาสละเวลาให้คำปรึกษา ตรวจสอบ และชี้แนะจุดแก้ไขข้อบกพร่องวิทยานิพนธ์

ดร.ทวีป แสงแห่งธรรม ประธานกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.นันทวัฒน์ อุติ ดร.ชัชฎาพงศ์ บุตรดี และคณาจารย์ภาควิชารังสีเทคนิค คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร ที่ให้ข้อเสนอแนะที่เป็นประโยชน์ ทำให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

พี่ และเพื่อน ร่วมหลักสูตรทุกท่านที่ให้ความช่วยเหลือ และให้กำลังใจตลอดการทำวิทยานิพนธ์ ขอขอบคุณทุกท่านที่มีส่วนสนับสนุนการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วง โดยไม่ได้กล่าวนามไว้ ณ ที่นี้

สุดท้ายนี้ขอขอบคุณ คุณพ่อ คุณแม่ สำหรับการดูแลด้วยความรัก การสนับสนุนในทุกด้านอย่างดีที่สุด และคอยเป็นกำลังใจที่ดีเสมอมา

คุณค่าและประโยชน์ที่เกิดจากวิทยานิพนธ์ฉบับนี้ ขอมอบและอุทิศแด่ผู้มีพระคุณทุกท่าน

รัฐพงษ์ นันทาแพร์

สารบัญ

| | หน้า |
|---|------|
| บทคัดย่อภาษาไทย..... | ค |
| บทคัดย่อภาษาอังกฤษ..... | จ |
| ประกาศคุณูปการ..... | ช |
| สารบัญ..... | ซ |
| สารบัญตาราง..... | ฅ |
| สารบัญภาพ..... | ฉ |
| สัญลักษณ์และอักษรย่อ..... | ฒ |
| บทที่ 1 บทนำ..... | 1 |
| ความเป็นมาและความสำคัญของปัญหา..... | 1 |
| จุดมุ่งหมายของการศึกษา..... | 2 |
| ขอบเขตการศึกษา..... | 3 |
| สมมุติฐานของการศึกษา..... | 3 |
| บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... | 4 |
| ทฤษฎีที่เกี่ยวข้อง..... | 4 |
| การเรียนรู้เชิงลึก..... | 4 |
| เพอร์เซปตรอน..... | 4 |
| ค่าถ่วงน้ำหนัก..... | 5 |
| ค่าเบี่ยงเบน..... | 5 |
| ฟังก์ชันกระตุ้น..... | 5 |

| | |
|---|----|
| โครงข่ายประสาทเทียม..... | 7 |
| การเรียนรู้ของโครงข่ายประสาทเทียม..... | 7 |
| อัลกอริทึมปรับปรุงประสิทธิภาพ..... | 9 |
| โครงข่ายประสาทเทียมคอนโวลูชัน..... | 9 |
| สถาปัตยกรรม U-net..... | 11 |
| สถาปัตยกรรม Features Pyramid Networks | 12 |
| Backbone models..... | 13 |
| Transfer learning and Fine-tuning..... | 23 |
| โรคมะเร็งศีรษะและลำคอ | 24 |
| อวัยวะที่มีความเสี่ยง | 25 |
| งานวิจัยที่เกี่ยวข้อง..... | 26 |
| Ibragimov B. และ Xing L. (2017) | 26 |
| Wang Y. และคณะ (2017)..... | 26 |
| Wu X. และคณะ (2018) | 26 |
| Tong N. และคณะ (2018) | 27 |
| Zhu W. และคณะ (2018)..... | 28 |
| Men K. และคณะ (2019)..... | 28 |
| van der Veen J. และคณะ (2018)..... | 28 |
| บทที่ 3 วิธีดำเนินงานวิจัย | 30 |
| ประชากรและกลุ่มตัวอย่าง..... | 30 |
| เครื่องมือที่ใช้ในการศึกษา | 32 |
| วิธีดำเนินการศึกษา..... | 32 |

| | |
|---|----|
| 1. การเตรียมข้อมูล image และ mask..... | 33 |
| 2. การพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยง | 38 |
| 3. การทดสอบและประเมินผล..... | 52 |
| บทที่ 4 ผลการศึกษา..... | 56 |
| ผลการทำนาย mask..... | 56 |
| ผลการประเมินค่าสัมประสิทธิ์ความเหมือน (DSC)..... | 62 |
| ผลการประเมินค่าระยะห่าง 95 เปอร์เซ็นต์ไทล์ฮาวดรอป (95%HD)..... | 66 |
| ผลการประเมินเวลา..... | 70 |
| บทที่ 5 บทสรุป..... | 72 |
| อภิปรายผลการทำนาย mask..... | 72 |
| อภิปรายผลค่าเฉลี่ย DSC และค่าเฉลี่ย 95%HD..... | 75 |
| อภิปรายผลการประเมินเวลา..... | 80 |
| ข้อจำกัด..... | 81 |
| ข้อเสนอแนะ..... | 81 |
| สรุป..... | 82 |
| บรรณานุกรม..... | 85 |
| ภาคผนวก ก ผลการประเมินค่า DSC..... | 90 |
| ภาคผนวก ข ผลการประเมินค่า 95%HD..... | 92 |
| ประวัติผู้วิจัย..... | 94 |

สารบัญตาราง

| | หน้า |
|---|------|
| ตาราง 1 ชนิดและสมการทางคณิตศาสตร์ของฟังก์ชันกระตุ้น | 6 |
| ตาราง 2 ค่าขีดจำกัดปริมาณรังสีสูงสุดของอวัยวะที่มีความเสี่ยงบริเวณศีรษะและลำคอ. | 25 |
| ตาราง 3 ข้อมูลคุณลักษณะผู้ป่วยของชุดข้อมูลภาพ 102 ชุดข้อมูล..... | 30 |
| ตาราง 4 การกระจายชุดข้อมูลสอน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ | 37 |
| ตาราง 5 จำนวนพารามิเตอร์ของโมเดลการเรียนรู้เชิงลึก..... | 49 |
| ตาราง 6 ผลการประเมินเวลาสอนโมเดลและเวลาทำนาย mask ของโมเดล..... | 71 |
| ตาราง 7 ผลค่าเฉลี่ย DSC จากโมเดลที่ให้ผลการทำนายที่ดีของการศึกษานี้เปรียบเทียบกับ ค่าเฉลี่ย DSC จากการศึกษาอื่นก่อนหน้า | 78 |
| ตาราง 8 ผลค่าเฉลี่ย 95%HD จากโมเดลที่ให้ผลการทำนายที่ดีของการศึกษานี้ เปรียบเทียบกับค่าเฉลี่ย 95%HD จากการศึกษาอื่นก่อนหน้า..... | 79 |
| ตาราง 9 การเปรียบเทียบระยะเวลาสอนโมเดลจากวิธีการสอนและรูปแบบสถาปัตยกรรม โครงข่ายเดียวกันที่โมเดล backbone ต่างกัน | 80 |
| ตาราง 10 สรุปผลชนิดของอวัยวะที่มีความเสี่ยงที่มีค่าเฉลี่ย DSC สูงสุด ค่าเฉลี่ย 95%HD ต่ำสุด เวลาสอน และเวลาทำนาย mask ของโมเดลที่ให้ผลการทำนายที่ดีในแต่ละรูปแบบ โมเดล backbone | 83 |

สารบัญภาพ

| | หน้า |
|---|------|
| ภาพ 1 เพอร์เซปตรอน..... | 4 |
| ภาพ 2 โครงข่ายประสาทเทียมแบบหลายชั้นเพอร์เซปตรอน | 7 |
| ภาพ 3 รูปแบบการ convolution ข้อมูลภาพด้วยเมทริกถ่วงน้ำหนัก..... | 11 |
| ภาพ 4 โครงสร้างและการดำเนินการของสถาปัตยกรรม U-net..... | 12 |
| ภาพ 5 โครงสร้างและการดำเนินการของสถาปัตยกรรม FPN | 13 |
| ภาพ 6 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ VGG19..... | 14 |
| ภาพ 7 การเชื่อมต่อชั้นดำเนินการของชุดดำเนินการ convolution กลุ่มที่ 1 และ กลุ่มที่ 2 โมเดล backbone แบบ ResNet50 | 15 |
| ภาพ 8 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ ResNet50..... | 16 |
| ภาพ 9 การเชื่อมต่อชั้นดำเนินการของชุดดำเนินการ convolution รูปแบบที่ 1 และ รูปแบบที่ 2 โมเดล backbone แบบ SE-ResNet50..... | 17 |
| ภาพ 10 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ SE-ResNet50..... | 18 |
| ภาพ 11 การเชื่อมต่อชั้นดำเนินการของ module 1 โมเดล backbone แบบ InceptionV3 | 19 |
| ภาพ 12 การเชื่อมต่อชั้นดำเนินการของ module 2 โมเดล backbone แบบ InceptionV3 | 20 |
| ภาพ 13 การเชื่อมต่อชั้นดำเนินการของ module 3 โมเดล backbone แบบ InceptionV3 | 22 |
| ภาพ 14 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ InceptionV3..... | 23 |
| ภาพ 15 แผนภาพการเตรียมข้อมูล CT image..... | 34 |

| | |
|---|----|
| ภาพ 16 แผนภาพการแปลงข้อมูล DICOM RT structure เป็นข้อมูล binary mask array .. | 35 |
| ภาพ 17 แผนภาพการเตรียมไฟล์ mask นำเข้า | 36 |
| ภาพ 18 การกระจายความถี่ mask อวัยวะที่มีความเสี่ยง ของชุดข้อมูลสอน ชุดข้อมูล ตรวจสอบ และชุดข้อมูลทดสอบ | 38 |
| ภาพ 19 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ VGG19..... | 39 |
| ภาพ 20 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ ResNet50..... | 40 |
| ภาพ 21 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ SE-ResNet50 | 41 |
| ภาพ 22 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ InceptionV3 | 42 |
| ภาพ 23 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ VGG19 | 43 |
| ภาพ 24 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ ResNet50 | 45 |
| ภาพ 25 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ SE-ResNet50..... | 46 |
| ภาพ 26 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ InceptionV3..... | 48 |
| ภาพ 27 แผนภาพการทำงานของโปรแกรมในขั้นตอนการสอนโมเดล | 51 |
| ภาพ 28 แผนภาพการทำงานของโปรแกรมในขั้นตอนการทำนาย mask..... | 54 |
| ภาพ 29 แผนภาพการทำงานของโปรแกรมในขั้นตอนการประเมินผลการทดสอบ..... | 55 |
| ภาพ 30 ผลการทำนายขอบเขตของ 7 อวัยวะที่มีความเสี่ยงจากโมเดล TUVGG19..... | 56 |
| ภาพ 31 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง brain stem | 57 |
| ภาพ 32 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง eye globes | 58 |
| ภาพ 33 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง mandible | 59 |

| | |
|---|----|
| ภาพ 34 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง parotid glands..... | 60 |
| ภาพ 35 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง spinal cord | 61 |
| ภาพ 36 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ VGG19 | 62 |
| ภาพ 37 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ ResNet50 | 63 |
| ภาพ 38 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ SE-ResNet50..... | 64 |
| ภาพ 39 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ InceptionV3..... | 65 |
| ภาพ 40 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ VGG19 | 66 |
| ภาพ 41 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ ResNet50 | 67 |
| ภาพ 42 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ SE-ResNet50..... | 68 |
| ภาพ 43 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดย โมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ InceptionV3..... | 69 |
| ภาพ 44 เปรียบเทียบผลการทำนาย mask ผิดหรือคลาดเคลื่อน บนภาพ soft tissue window กับภาพ brain window | 73 |

สัญลักษณ์และอักษรย่อ

| | |
|------------|---|
| 95%HD | 95 percentile Hausdorff distance |
| Add | addition method |
| avg pool | average pooling layer |
| batchNorm | batch normalization layer |
| BST | brain stem |
| BU | bottom-up pathway |
| concat | concatenation |
| CNN | convolutional neural networks |
| conv | convolution layer |
| conv G1 | convolution group 1 |
| conv G2 | convolution group 2 |
| conv G3 | convolution group 3 |
| conv G4 | convolution group 4 |
| conv (1,1) | convolution kernel size 1×1 layer |
| conv (3,3) | convolution kernel size 3×3 layer |
| conv-batch | convolutional layer follow by batch normalization layer |
| CT | computed tomography |
| DICOM | digital imaging and communications in medicine |
| DSC | Dice similarity coefficient |
| EYE_Lt | left eye globe |
| EYE_Rt | right eye globe |
| FCNN | fully convolutional neural networks |
| FPN | feature pyramid networks |
| Gy | Gray |
| IncepV3 | InceptionV3 backbone model |
| L-I | lateral connection 1 |
| L-II | lateral connection 2 |
| MAN | mandible |

| | |
|------------------|--|
| max pool | max pooling layer |
| Mix(concat) | Merging layers more than 2 layers by concatenation method |
| Multiply | Multiply method |
| OARs | Organs at risk |
| padding | padding layer |
| Pre | prediction |
| PRO_Lt | left parotid gland |
| PRO_Rt | right parotid gland |
| Ref | reference |
| ReLU | rectified linear activation function |
| ResNet50 | Deep residual networks backbone model |
| RTOG | The radiation therapy oncology group |
| SE-ResNet50 | squeeze-and-excitation networks backbone model |
| softmax | soft max activation function layer |
| SPC | spinal cord |
| TCGA-HNSC | The cancer genome atlas head-neck squamous cell carcinoma data set |
| TD | top-down pathway |
| upsampling (2,2) | upsampling kernel size 2x2 layer |
| upsampling (8,8) | upsampling kernel size 8x8 layer |
| upsampling (4,4) | upsampling kernel size 4x4 layer |
| VGG19 | visual geometry group 19 backbone model |

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การเรียนรู้เชิงลึก (Deep learning) เป็นส่วนหนึ่งของศาสตร์ด้านปัญญาประดิษฐ์ (Artificial Intelligent) มีจุดมุ่งหมายเพื่อสร้างความฉลาดรูปแบบหนึ่งให้คอมพิวเตอร์สามารถประมวลผลข้อมูลในรูปแบบนามธรรมจากระดับพื้นฐานถึงระดับสูงได้ โดยมีแนวคิดจากรูปแบบการทำงานของระบบประสาทสมองของมนุษย์ที่มีการทำงานเชื่อมต่อของเซลล์ประสาทกว่า 1,000 ล้านเซลล์ ส่งผลให้มนุษย์มีลำดับการคิด การทำความเข้าใจ และการแสดงออกที่มีความซับซ้อน (1-3) รูปแบบแนวคิดนี้ได้รับความสนใจและถูกพัฒนาเพื่อใช้แก้ไขปัญหาในด้านต่าง ๆ เช่น ด้าน computer vision ด้าน natural language processing รวมถึงงานด้านการแพทย์และสาธารณสุข เป็นต้น

การพัฒนาโมเดลการเรียนรู้เชิงลึกและใช้ประโยชน์ด้าน computer vision ถูกนำมาใช้เพื่อแก้ไขปัญหาการประมวลผลภาพและวิดีโอ เช่น การแบ่งส่วนภาพ (4) การจำแนกภาพ (5, 6) เป็นต้น การพัฒนาและใช้ประโยชน์ด้าน natural language processing เช่น การแปลด้วยคอมพิวเตอร์ (7) การสร้างแบบจำลองภาษา (8) เป็นต้น การพัฒนาและใช้ประโยชน์ทางด้านการแพทย์ถูกใช้เพื่อช่วยแก้ปัญหาบางประการ ทำให้ลดระยะเวลาและกระบวนการทำงานมีประสิทธิภาพมากขึ้น เช่น การใช้โมเดลการเรียนรู้เชิงลึกเพื่อช่วยตรวจจบบรอยโรค (9) ส่งผลให้การวินิจฉัยมีความแม่นยำมากยิ่งขึ้น การนำมาช่วยในการค้นคว้าโมเลกุลยาชนิดใหม่ (10, 11) การนำมาช่วยในกระบวนการรักษาผู้ป่วยมะเร็งที่ใช้วิธีการฉายรังสี เช่น การตรวจจบบการเคลื่อนไหวของก้อนมะเร็งในระหว่างจำลองการรักษา (12) การกำหนดขอบเขตก้อนมะเร็งและอวัยวะที่มีความเสี่ยง (13-18) จนกระทั่งถึงการติดตามผลการรักษา (19-21)

การรักษาผู้ป่วยมะเร็งด้วยการฉายรังสีเป็นหนึ่งในวิธีการรักษาที่ได้รับการยอมรับสำหรับบริเวณที่มีความซับซ้อนทางกายวิภาคศาสตร์ เช่น บริเวณศีรษะและลำคอ เนื่องจากวิธีการรักษาอื่น เช่น การผ่าตัด ทำได้ยากและอาจมีผลกระทบหลังการรักษามากกว่า บริเวณศีรษะและลำคอมีตำแหน่งที่จะเกิดโรคมะเร็งได้หลายตำแหน่ง เช่น มะเร็งหลังโพรงจมูก หลอดอาหาร กล่องเสียง ต่อมไทรอยด์ ลิ้น และต่อมน้ำลายข้างกกหู ซึ่งโรคมะเร็งชนิดต่าง ๆ เหล่านี้พบได้มากในผู้ป่วยมะเร็งชาวไทย จากสถิติผู้ป่วยมะเร็งที่เข้ารับบริการที่สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ ปี 2562 (22) พบว่ามีจำนวนผู้ป่วยมะเร็งชนิดต่าง ๆ ที่กล่าวข้างต้นรวมมากถึง 453 ราย

กระบวนการรักษาผู้ป่วยมะเร็งเรื้อรังด้วยการฉายรังสีประกอบด้วยหลายขั้นตอน (23) ได้แก่ การจัดทำจำลองการรักษา การกำหนดขอบเขตเป้าหมายและอวัยวะปกติโดยรอบ การคำนวณปริมาณรังสี การประกันคุณภาพก่อนการฉายรังสี การฉายรังสี และการติดตามการรักษา ซึ่งแต่ละกระบวนการดังกล่าวมีความเกี่ยวข้องกัน หากมีความคลาดเคลื่อนของหนึ่งในกระบวนการดังกล่าว ก็จะส่งผลกระทบต่อผลการรักษาทั้งกระบวนการ การศึกษานี้ให้ความสนใจในขั้นตอนการกำหนดขอบเขตของอวัยวะปกติโดยรอบก่อนมะเร็ง เนื่องจากการฉายรังสีเพื่อทำลายเซลล์มะเร็งที่บริเวณศีรษะและลำคอจะทำให้อวัยวะปกติโดยรอบได้รับปริมาณรังสีระหว่างการฉายรังสี หากได้รับในปริมาณที่เกิดขีดจำกัดจะส่งผลให้สูญเสียการทำงานปกติได้ ซึ่งก็จะส่งผลต่อคุณภาพชีวิตของผู้ป่วยหลังได้รับการรักษามะเร็งด้วยวิธีการฉายรังสี เช่น การเกิดภาวะน้ำลายแห้ง (Xerostomia) (24) การตายของกระดูกขากรรไกรหลังการฉายรังสี (post extraction osteoradionecrosis) (25) การเกิดความผิดปกติต่อดวงตาและระบบประสาทตา (26) การเกิดอาการผิดปกติจากการสูญเสียการทำงานของก้านสมองและไขสันหลัง (27) เป็นต้น

การปฏิบัติงานทางคลินิกโดยทั่วไปการกำหนดขอบเขตอวัยวะปกติที่มีความเสี่ยง จะถูกกำหนดโดยแพทย์รังสีรักษา แต่ก็มีข้อจำกัดคือ ขั้นตอนการกำหนดขอบเขตที่ใช้ระยะเวลานาน มีความแปรปรวนระหว่างแพทย์ผู้กำหนดขอบเขต และมีปัจจัยด้านประสบการณ์ของแพทย์ร่วมด้วย (28) จากการทบทวนวรรณกรรมมีหลายการศึกษาก่อนหน้าได้พัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบริเวณศีรษะและลำคอ โดยมีเป้าหมายเพื่อลดข้อจำกัดในการกำหนดขอบเขตที่กล่าวข้างต้น ซึ่งส่วนมากใช้โครงสร้างสถาปัตยกรรมโครงข่ายรูปแบบ U-net ในขณะที่เดียวกันก็มีการพัฒนารูปแบบสถาปัตยกรรมโครงข่ายอีกหลายรูปแบบที่น่าสนใจ รวมถึงรูปแบบวิธีการสอนโมเดลการเรียนรู้เชิงลึก ดังนั้นการศึกษานี้จึงมีแนวคิดที่จะพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาที่ใช้รูปแบบสถาปัตยกรรมโครงข่าย U-net (29) และ Features Pyramid Networks (FPN) (30) แล้วทำการประเมินประสิทธิภาพโมเดลที่พัฒนาขึ้น พร้อมทั้งศึกษาปัจจัยที่มีผลต่อการทำนายของโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงที่พัฒนาขึ้น

จุดมุ่งหมายของการศึกษา

1. เพื่อพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ โดยใช้รูปแบบสถาปัตยกรรมโครงข่าย U-Net และ FPN

2. เพื่อประเมินเปรียบเทียบประสิทธิภาพความถูกต้อง และเวลาการทำนายของโมเดลการเรียนรู้เชิงลึกสำหรับการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอที่พัฒนาขึ้น และวิเคราะห์ปัจจัยที่มีผลต่อประสิทธิภาพการทำงานของโมเดล

ขอบเขตการศึกษา

พัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ โดยใช้รูปแบบสถาปัตยกรรมโครงข่าย U-Net และ FPN ใช้ข้อมูลภาพเอกซเรย์คอมพิวเตอร์ และข้อมูล Radiation therapy structure รูปแบบไฟล์ DICOM จากฐานข้อมูล The cancer image archive ในชุดข้อมูล Head-Neck-PET-CT (31) จาก 4 สถาบันที่แตกต่างกันใน Québec ประเทศแคนาดา จำนวน 101 ราย และจากชุดข้อมูล TCGA-HNSC (32) จำนวน 1 ราย สำหรับเตรียมชุดข้อมูลสอน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ เปรียบเทียบประสิทธิภาพการทำนายขอบเขตอวัยวะที่มีความเสี่ยงของโมเดลที่พัฒนาขึ้นจากค่าสัมประสิทธิ์ความเหมือน (DSC) ค่าระยะห่าง 95 เปอร์เซ็นไทล์ฮาวดรอป (95%HD) เวลาสอนโมเดล เวลาทำนายผล mask และประเมินปัจจัยอื่นที่มีผลต่อประสิทธิภาพการทำนายของโมเดลการเรียนรู้เชิงลึก

สมมุติฐานของการศึกษา

โมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอที่พัฒนาขึ้น สามารถทำนายขอบเขตอวัยวะที่มีความเสี่ยงแบบอัตโนมัติได้ดี และใช้เวลาในการกำหนดขอบเขตอัตโนมัติอย่างเหมาะสม

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

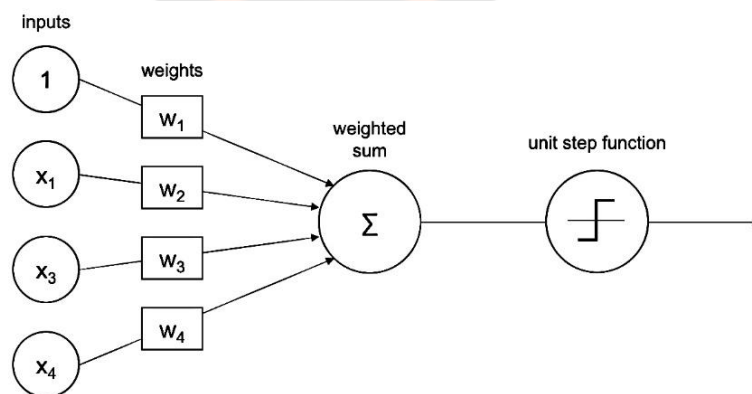
ทฤษฎีที่เกี่ยวข้อง

การเรียนรู้เชิงลึก (1-3)

การเรียนรู้เชิงลึกเป็นส่วนหนึ่งของศาสตร์ด้านปัญญาประดิษฐ์ ถูกสร้างขึ้นจากแนวคิดการทำงานของเซลล์ประสาทสมองของมนุษย์ที่มีการทำงานเชื่อมต่อกันมากถึง 1000 ล้านเซลล์ ส่งผลให้มนุษย์มีลำดับการคิด การแสดงออกที่มีความซับซ้อน โดยแนวคิดนี้ถูกพัฒนาเป็นส่วนหนึ่งของการประมวลผลของคอมพิวเตอร์เพื่อให้คอมพิวเตอร์สามารถประมวลผลชุดข้อมูลได้ เช่น ข้อมูลภาพ เสียง วิดีโอ ข้อความ เป็นต้น โดยข้อมูลเหล่านี้จะถูกค้นหาคุณลักษณะเฉพาะของข้อมูลโดยอัตโนมัติเมื่อเริ่มการการทำงานของตัวประมวลผล โดยไม่จำเป็นต้องอาศัยการกำหนดเงื่อนไขจากผู้ใช้ จากนั้นการทำงานของตัวประมวลผลจะนำไปสู่การทำนายผลลัพธ์ ผลลัพธ์จากการทำนายจะถูกนำไปหาค่าความคลาดเคลื่อนกับค่าความจริง ค่าความคลาดเคลื่อนที่ได้ถูกนำมาใช้ปรับปรุงตัวประมวลผล การประมวลผลที่มีขั้นตอนการค้นหาคุณลักษณะหลายขั้นนี้ถูกจัดให้เป็นรูปแบบของการเรียนรู้เชิงลึก

เพอร์เซปตรอน (3)

เพอร์เซปตรอนเป็นหน่วยการประมวลผลที่เล็กที่สุดของการเรียนรู้เชิงลึก มีแนวคิดการทำงานคล้ายกับเซลล์ประสาทสมองของมนุษย์ประกอบด้วย ส่วนรับข้อมูลนำเข้า (input) ส่วนค่าถ่วงน้ำหนัก (weights) ส่วนรวมผลคูณของค่าข้อมูลนำเข้าและค่าถ่วงน้ำหนัก (weighted sum) ส่วนค่าเบี่ยงเบน (bias) และส่วนฟังก์ชันกระตุ้น (activation function) แสดงดังภาพ 1



ภาพ 1 เพอร์เซปตรอน (3)

ค่าถ่วงน้ำหนัก

ค่าถ่วงน้ำหนัก (weight) เป็นค่าพารามิเตอร์ที่เป็นส่วนประกอบของเพอร์เซปตรอน ซึ่งจะถูกรับเปลี่ยนค่าเมื่อมีการเรียนรู้ของโมเดล โดยค่าถ่วงน้ำหนักจะถูกคูณเข้ากับค่าข้อมูลสำหรับแต่ละหน่วยของข้อมูลนำเข้า เพื่อแสดงระดับความสำคัญของคุณลักษณะค่าข้อมูลนำเข้า ก่อนที่จะถูกรวมและส่งผ่านฟังก์ชันกระตุ้นต่อไป

ค่าเบี่ยงเบน

ค่าเบี่ยงเบน (bias) เป็นค่าพารามิเตอร์ที่มี 1 ตัวต่อ 1 เพอร์เซปตรอน สามารถถูกปรับเปลี่ยนค่าได้เมื่อมีการเรียนรู้ของโมเดล โดยค่าเบี่ยงเบนจะถูกรวมเข้ากับผลรวมของผลคูณระหว่างค่าถ่วงน้ำหนักกับค่าข้อมูลนำเข้าทั้งหมดที่เข้าสู่เพอร์เซปตรอน ความสำคัญของค่าเบี่ยงเบนใช้สำหรับบอกความสำคัญของเพอร์เซปตรอนในโครงข่ายที่ประกอบด้วยเพอร์เซปตรอนหลายตัว

ฟังก์ชันกระตุ้น (1)

ฟังก์ชันกระตุ้น (activation function) เป็นส่วนหนึ่งของเพอร์เซปตรอนทำหน้าที่เป็นจุดกำหนดการส่งผ่านค่าข้อมูลไปยังเพอร์เซปตรอนอื่นหรือไม่ด้วยฟังก์ชันทางคณิตศาสตร์ ช่วยให้การส่งผ่านค่าข้อมูลที่ผ่านมาการดำเนินการทางคณิตศาสตร์ด้วยค่าถ่วงน้ำหนักและค่าเบี่ยงเบนไปยังเพอร์เซปตรอนอื่นไม่เป็นเชิงเส้น และในระหว่างการสอนโมเดลฟังก์ชันกระตุ้นมีบทบาทสำคัญในการปรับแก้เดียน คุณสมบัติของฟังก์ชันกระตุ้นดังกล่าวนี้ช่วยให้โครงข่ายที่มีการเชื่อมต่อแบบลึกสามารถเรียนรู้ฟังก์ชันที่ซับซ้อนได้ โดยฟังก์ชันกระตุ้นมีหลายชนิด ดังแสดงในตารางที่ 1

ฟังก์ชัน Sigmoid มีการดำเนินการทางคณิตศาสตร์ดังสมการในตารางที่ 1 ผลลัพธ์จากการดำเนินการของฟังก์ชันมีค่าระหว่าง 0 ถึง 1 ข้อจำกัดของฟังก์ชัน Sigmoid คือ ระดับความชันของฟังก์ชันน้อย ส่งผลให้การลู่เข้าช้า

ฟังก์ชัน Tanh มีการดำเนินการทางคณิตศาสตร์ดังสมการในตาราง 1 ผลลัพธ์จากการดำเนินการของฟังก์ชันมีค่าระหว่าง -1 ถึง 1 ค่ากึ่งกลางอยู่ที่ 0 ส่งผลให้การ optimization นั้นทำได้ง่าย แต่ก็ยังมีข้อจำกัดที่ระดับความชันของฟังก์ชันน้อย

ฟังก์ชัน Rectified linear unit (ReLU) มีการดำเนินการทางคณิตศาสตร์ดังสมการในตาราง 1 ผลลัพธ์จากการดำเนินการของฟังก์ชันมีค่าเท่ากับ $[0, \infty)$ จุดเด่นของฟังก์ชันคือ มีระดับความชันส่งผลให้การลู่เข้าดีกว่าเมื่อเปรียบเทียบกับฟังก์ชัน Sigmoid และ Tanh ข้อจำกัดคือ ใช้ได้เฉพาะชั้นการประมวลผลที่มองไม่เห็น (Hidden layer) ไม่สามารถใช้ในชั้นข้อมูลขาออกสุดท้ายของโมเดลได้ จำเป็นต้องใช้ฟังก์ชัน SoftMax ในชั้นสุดท้ายแทนเพื่อคำนวณความน่าจะเป็น และอีก 1 ข้อจำกัด

ของฟังก์ชัน ReLU คือ สามารถทำให้เกิดปัญหาการตายของเซลล์ประสาทได้ หากมีค่าข้อมูลสูงไหลผ่าน ReLU อาจทำให้การถ่วงน้ำหนักถูกปรับปรุงมาก ส่งผลให้เซลล์ประสาทหยุดทำงานเมื่อมีชุดข้อมูลใหม่ไหลผ่าน

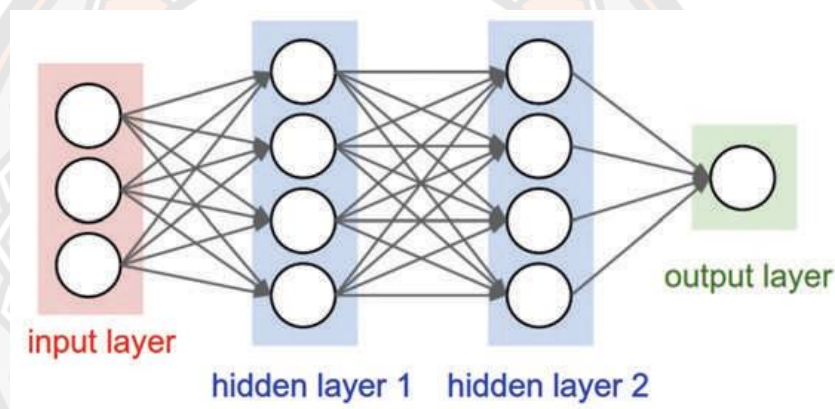
ฟังก์ชัน Softmax เป็นฟังก์ชันที่มีการดำเนินการทางคณิตศาสตร์ โดยผลลัพธ์อยู่ในรูปแบบความน่าจะเป็น ฟังก์ชัน softmax จะถูกใช้ในชั้นสุดท้ายเมื่อใช้ฟังก์ชัน ReLU สำหรับการแก้ปัญหาการจำแนกประเภท

ตาราง 1 ชนิดและสมการทางคณิตศาสตร์ของฟังก์ชันกระตุ้น (1)

| Name | Equation | Derivative |
|----------------------------|--|--|
| Binary Step | $\sigma(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases}$ | $\sigma'(x) = \begin{cases} 0, & x \neq 0 \\ ?, & x = 0 \end{cases}$ |
| Identity | $\sigma(x) = x$ | $\sigma'(x) = 1$ |
| Sigmoid | $\sigma(x) = \frac{1}{1 + e^{-x}}$ | $\sigma'(x) = \sigma(x)(1 - \sigma(x))$ |
| Tanh | $\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ | $\sigma'(x) = 1 - \sigma(x)^2$ |
| Rectified Linear (ReLU) | $\sigma(x) = \max(0, x)$ | $\sigma'(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ |
| Leaky ReLU | $\sigma(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases}$ | $\sigma'(x) = \begin{cases} 1, & x \geq 0 \\ a, & x < 0 \end{cases}$ |

โครงข่ายประสาทเทียม (1)

โครงข่ายประสาทเทียม ประกอบด้วยชั้นการดำเนินการหลัก 3 ชั้น ได้แก่ ชั้นข้อมูลนำเข้า (input layer) ชั้นการประมวลผลซ่อน (hidden layer) และชั้นผลลัพธ์ (output layer) โดยในชั้นข้อมูลนำเข้ามีโหนดรับค่าข้อมูลที่จะทำการประมวลผล ชั้นการประมวลผลซ่อน ประกอบไปด้วยการเรียงตัวของเพอร์เซปตรอนหลายเพอร์เซปตรอน โดยแต่ละตัวจะเชื่อมกับทุกโหนดในชั้นข้อมูลนำเข้า และเชื่อมกับทุกเพอร์เซปตรอนในชั้นการประมวลผลซ่อนชั้นถัดไป โดยในโครงข่ายประสาทเทียมจะมีชั้นการประมวลผลซ่อนหลายชั้นได้ขึ้นอยู่กับความต้องการประมวลผล ชั้นผลลัพธ์จะเป็นชั้นสุดท้ายของโครงข่ายประสาทเทียมโดยจะมีจำนวนโหนดตามรูปแบบของปัญหาที่ต้องการประมวลผล โดยแต่ละโหนดจะเชื่อมกับทุกโหนดในชั้นการประมวลผลซ่อนสุดท้าย แสดงการเชื่อมต่อของโครงข่ายประสาทเทียมดังภาพที่ 2



ภาพ 2 โครงข่ายประสาทเทียมแบบหลายชั้นเพอร์เซปตรอน (1)

การเรียนรู้ของโครงข่ายประสาทเทียม (1)

การเรียนรู้ของโครงข่ายประสาทเทียมประกอบด้วย 2 ขั้นตอนหลัก ได้แก่ การดำเนินการไปข้างหน้า (forward propagation) และการดำเนินการย้อนกลับ (back propagation) โดยสามารถอธิบายสองขั้นตอนนี้จากเพอร์เซปตรอน 2 ตัว โดยกำหนดให้ a_0 และ y เป็นตัวแทนของข้อมูลขาเข้า และข้อมูลขาออกอ้างอิง ตามลำดับ กำหนดการเชื่อมต่อด้วยค่าการถ่วงน้ำหนักและค่าเบี่ยงเบนด้วย (w_1, b_1) และ (w_2, b_2) และกำหนดให้ σ แทนฟังก์ชันกระตุ้น

การดำเนินการไปข้างหน้า ค่าถ่วงน้ำหนักจะถูกกำหนดแบบสุ่มหรือใช้ค่าถ่วงน้ำหนักจากแบบจำลองที่ถูกสอนมาก่อน จะถูกดำเนินการคูณกับค่าข้อมูลนำเข้าและรวมเข้ากับค่าเบี่ยงเบนของแต่ละเพอร์เซปตรอนจากนั้นจะผ่านฟังก์ชันกระตุ้นไปยังชั้นถัดไป แสดงดังสมการ 1 - 4

$$Z_1 = a_0 * w_1 + b_1 \quad (1)$$

$$a_1 = \sigma(Z_1) \quad (2)$$

$$Z_2 = a_1 * w_2 + b_2 \quad (3)$$

$$a_2 = \sigma(Z_2) \quad (4)$$

การดำเนินการย้อนกลับ เป็นขั้นตอนการเรียนรู้ของโครงข่ายผ่านค่าความคลาดเคลื่อนเพื่อปรับปรุงค่าถ่วงน้ำหนักและค่าเบี่ยงเบน โดยผลลัพธ์ที่ได้จากขั้นตอนการดำเนินการไปข้างหน้า จะถูกนำไปคำนวณหาค่าความคลาดเคลื่อนกับค่าคำตอบอ้างอิงตั้งสมการ 5 เมื่อ ∇a คือ ค่าความคลาดเคลื่อน

$$\nabla a = a_2 - y \quad (5)$$

โดยค่าความคลาดเคลื่อนที่คำนวณได้จะถูกนำมาปรับปรุงค่าถ่วงน้ำหนักและค่าเบี่ยงเบน โดยดำเนินการปรับค่าตามแนวเกรเดียน เริ่มจากการคำนวณค่าความคลาดเคลื่อนของค่าถ่วงน้ำหนัก และค่าเบี่ยงเบนตามสมการ 6 ถึง 11

กำหนดให้ ∇Z_2 เป็นตัวปรับค่า b_2 และ w_2 และ ∇Z_1 เป็นตัวปรับค่า b_1 และ w_1

$$\nabla Z_2 = \nabla a * \sigma'(Z_2) \quad (6)$$

$$\nabla b_2 = \nabla Z_2 \quad (7)$$

$$\nabla w_2 = a_1^T * \nabla Z_2 \quad (8)$$

$$\nabla Z_1 = \nabla a * \sigma'(Z_1) \quad (9)$$

$$\nabla b_1 = \nabla Z_1 \quad (10)$$

$$\nabla w_1 = a_0^T * \nabla Z_1 \quad (11)$$

เมื่อผลการปรับค่าถูกคำนวณแล้วจะทำการอัปเดตพารามิเตอร์ (ค่าถ่วงน้ำหนัก และค่าเบี่ยงเบนของเพอร์เซปตรอน) ส่วนใหญ่มักเป็นตัวอย่างเทคนิค gradient descent โดยที่ n แสดงถึงอัตราการเรียนรู้ แสดงการอัปเดตพารามิเตอร์ตั้งสมการ 12 ถึง 15

$$w_1 \leftarrow w_1 \leftarrow n \cdot \nabla w_1 \quad (12)$$

$$b_1 \leftarrow b_1 \leftarrow n \cdot \nabla b_1 \quad (13)$$

$$w_2 \leftarrow w_2 \leftarrow n \cdot \nabla w_2 \quad (14)$$

$$b_2 \leftarrow b_2 \leftarrow n \cdot \nabla b_2 \quad (15)$$

อัลกอริทึมปรับปรุงประสิทธิภาพ (1)

อัลกอริทึมปรับปรุงประสิทธิภาพ (Optimization algorithms) เป็นส่วนสำคัญสำหรับการเรียนรู้ของโครงข่ายประสาทเทียม ทำหน้าที่สำหรับการปรับปรุงค่าให้โมเดลมีการเรียนรู้ที่เหมาะสม ลดความคลาดเคลื่อน หรือการหาตำแหน่งที่ทำให้มีข้อผิดพลาดน้อยที่สุด รวมถึงขั้นตอนการปรับค่าพารามิเตอร์ของโครงข่าย ตัวอย่างของอัลกอริทึมปรับปรุงประสิทธิภาพพื้นฐาน คือ gradient descents นอกจากนี้ยังมีอีกหลายอัลกอริทึมที่ทำงานคล้ายกัน เช่น Adagrad optimizer, Momentum optimizer, Adam optimizer, Ftrl optimizer RMSProp optimizer เป็นต้น

โครงข่ายประสาทเทียมคอนโวลูชัน (1, 3, 33, 34)

โครงข่ายประสาทเทียมคอนโวลูชัน เป็นโครงข่ายประสาทเทียมชนิดหนึ่งที่มีการดำเนินการประมวลผลทางคณิตศาสตร์ที่เรียกว่า คอนโวลูชัน โดยโครงข่ายประสาทเทียมชนิดนี้ประมวลผลได้ดีกับข้อมูลที่มีความสัมพันธ์เชิงพื้นที่แบบกริด ประกอบด้วยโครงข่ายของเพอร์เซพตรอนหลายตัวเรียงต่อกันหลายชั้น มีการเชื่อมต่อภายในระหว่างเพอร์เซพตรอนของชั้นที่อยู่ติดกัน องค์ประกอบโครงสร้างพื้นฐานของโครงข่ายประสาทเทียมคอนโวลูชัน ประกอบไปด้วยชั้นดำเนินการดังนี้

Convolutional layer เป็นชั้นที่มีการดำเนินการ คอนโวลูชัน แสดงขั้นตอนดังภาพที่ 3 ในภาพประกอบด้วยเมทริกข้อมูลภาพนำเข้า (Input image) และเมทริกถ่วงน้ำหนัก (weight) หรือที่รู้จักทั่วไปในชื่อ kernel / filter การคอนโวลูชัน เริ่มจากการนำเมทริกถ่วงน้ำหนักวางทับเหนือเมทริกข้อมูลภาพนำเข้า ค่าข้อมูลในช่องเมทริกที่ซ้อนกันจะถูกดำเนินการคูณ จากนั้นนำผลคูณแต่ละช่องรวมเข้าด้วยกัน ค่าผลรวมที่ได้จะถูกวางลงในเมทริกใหม่ (เมทริกผลลัพธ์ของการคอนโวลูชัน) ที่ตำแหน่งกึ่งกลางเมทริกถ่วงน้ำหนัก เมทริกถ่วงน้ำหนักจะถูกเลื่อนไปยังตำแหน่งพิกเซลถัดไป (การเลื่อนที่ของเมทริกถ่วงน้ำหนักครั้งละ 1 พิกเซล จะเรียกว่า 1 stride) กระบวนการทำซ้ำจนกว่าพิกเซลของเมทริกภาพทั้งหมดจะได้รับการประมวลผล จากภาพที่ 3 แสดงการ stride ของเมทริกถ่วงน้ำหนักเท่ากับ 2 เมื่อมีจำนวน stride ที่มากขึ้นจะทำให้ขนาดภาพเล็กลงอย่างรวดเร็ว ดังนั้นหากต้องการรักษาขนาดเมทริกภาพให้มีขนาดคงเดิมสามารถเพิ่มค่าพิกเซล 0 ในแนวแถวและคอลัมน์ที่บริเวณขอบภาพ การเพิ่มค่าพิกเซลเพื่อรักษาขนาดภาพในลักษณะนี้เรียกว่า same padding

Activation layer เป็นชั้นดำเนินการของฟังก์ชันกระตุ้น มีบทบาทเสมือนประตูทางผ่านของสัญญาณระหว่างชั้นก่อนหน้าและชั้นถัดไป การควบคุมของชั้นฟังก์ชันกระตุ้นนี้ขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์ที่ผู้ใช้งานกำหนด เช่น ฟังก์ชัน ReLU สำหรับชั้นการประมวลผลซ่อน ฟังก์ชัน soft max สำหรับชั้นสุดท้ายของโครงข่ายเพื่อทำนายความน่าจะเป็นของผลลัพธ์ เป็นต้น

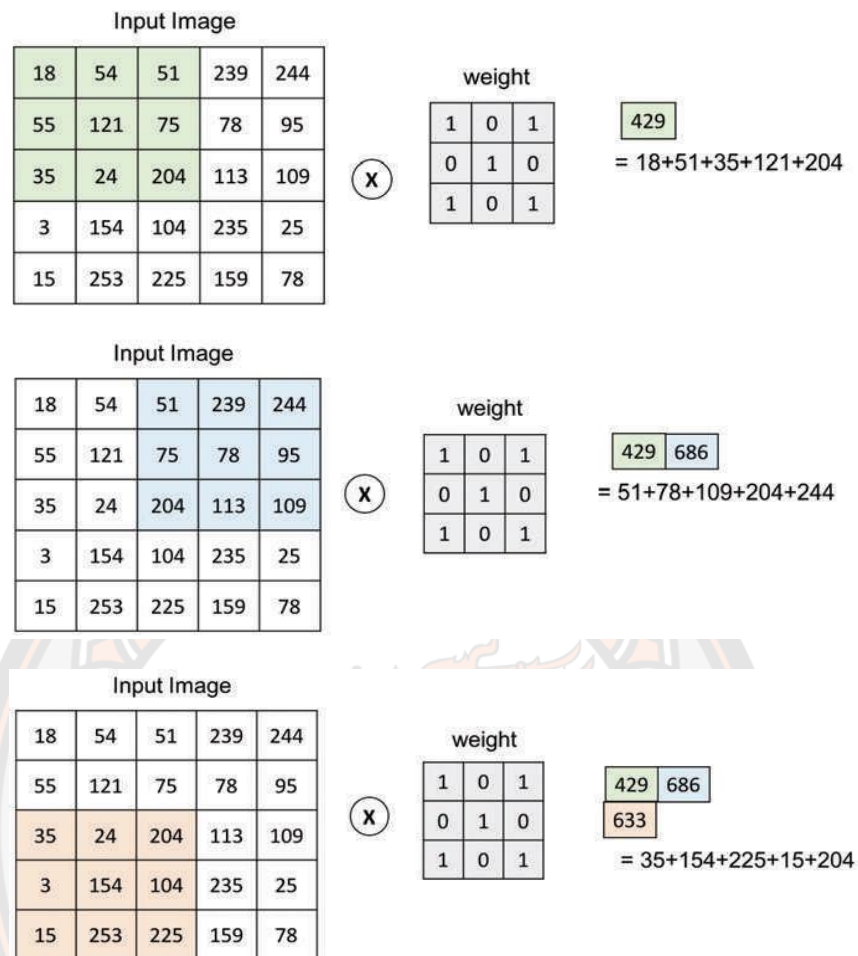
Pooling or Sub-sampling layer เป็นชั้นดำเนินการลดขนาดเชิงพื้นที่ด้วยวิธีการสุ่มตัวอย่าง (sampling) เพื่อลดจำนวนพารามิเตอร์และการคำนวณภายในโครงข่าย รูปแบบการดำเนินการของชั้น Pooling ใช้วิธีการวางเมทริกสุ่มตัวอย่าง เช่น max pooling, average pooling เป็นต้น โดย max pooling คือ การสุ่มตัวอย่างค่าที่มากที่สุดในขอบเขตของเมทริกเป็นตัวแทนข้อมูล average pooling คือ การสุ่มตัวอย่างค่าเฉลี่ยของข้อมูลในขอบเขตเมทริกเป็นตัวแทนข้อมูล ขนาดของเมทริกสุ่มตัวอย่างมีหลายขนาด เช่น ขนาด 3×3 , stride = 2 (overlapping pooling), ขนาด 2×2 pooling, stride = 2 เป็นต้น

Fully connected layer เป็นชั้นดำเนินการก่อนชั้นสุดท้ายของโครงข่าย มีการเชื่อมต่อแบบเต็มรูปแบบกับฟังก์ชันกระตุ้นทั้งหมดในชั้นก่อนหน้า แล้วถูกเชื่อมต่อกับชั้นสุดท้ายด้วยชั้นฟังก์ชัน soft max เพื่อแสดงผลความน่าจะเป็นของข้อมูลขาออก

Upsampling layer เป็นชั้นการดำเนินการถอดรหัสโดยใช้วิธีการ Deconvolution (35) เพื่อขยายค่าข้อมูลผลลัพธ์จากการดำเนินการ convolution ให้มีขนาดเท่ากับขนาดภาพนำเข้าเริ่มแรก และใช้ชั้น soft max activation function ในชั้นสุดท้ายของโครงข่าย เพื่อทำนายความน่าจะเป็นของแต่ละพิกเซลตามจำนวนคลาสหรือ channel ที่ต้องการ

Batch normalization layer เป็นชั้นที่มีการดำเนินการทางคณิตศาสตร์เพื่อทำการปรับข้อมูลให้เป็นมาตรฐาน โดยการปรับจะทำการรักษาค่าเฉลี่ย ข้อมูลขาออกให้เข้าใกล้ 0 และค่าเบี่ยงเบนมาตรฐานข้อมูลขาออกให้เข้าใกล้ 1 โดยในระหว่างการสอนชั้น batch normalization จะปรับข้อมูลขาออกให้เป็นมาตรฐาน โดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของชุดข้อมูลนำเข้าปัจจุบัน

Merging layers เป็นชั้นที่ดำเนินการผสมผลลัพธ์ระหว่างชั้นดำเนินการสองชั้นหรือมากกว่าเข้าด้วยกัน โดยรับข้อมูลนำเข้าในรูปแบบรายการของ tensor ที่มีรูปร่างทั้งหมดเหมือนกัน แล้วส่งคืนผลลัพธ์เป็นค่า tensor ตัวเดียว ที่เป็นการผสมกันของข้อมูลที่ถูกนำเข้ามาผสมกันทั้งหมด โดยใช้วิธีการดำเนินการของชั้นต่าง ๆ เช่น Concatenate layer เป็นชั้นการเชื่อมข้อมูลนำเข้าในรูปแบบรายการของ tensor, Add layer เป็นชั้นการรวมข้อมูลนำเข้าในรูปแบบรายการของ tensor, Multiply layer เป็นชั้นการคูณ (element-wise) ข้อมูลนำเข้าในรูปแบบรายการของ tensor เป็นต้น

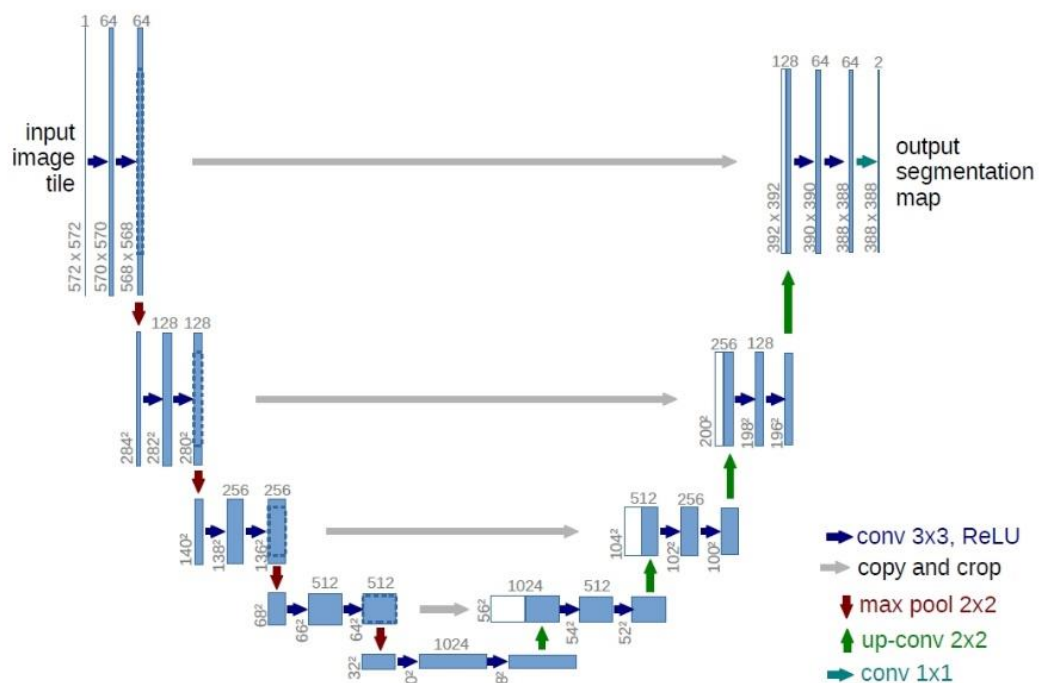


ภาพ 3 รูปแบบการ convolution ข้อมูลภาพด้วยเมทริกถ่วงน้ำหนัก (1)

สถาปัตยกรรม U-net (29)

สถาปัตยกรรม U-net พัฒนาจากสถาปัตยกรรมคอนโวลูชันแบบเต็มรูปแบบ fully convolutional neural networks (FCNN) โดยทำการปรับปรุงและขยายบางส่วนของสถาปัตยกรรม FCNN ทำให้สถาปัตยกรรม U-net ทำงานได้แม้จะใช้ข้อมูลรูปภาพในการสอนจำนวนน้อยได้ และยังคงให้ผลลัพธ์ที่ถูกต้องแม่นยำสำหรับงานด้านการแบ่งส่วนภาพ โดยรูปแบบสถาปัตยกรรมแสดงดังภาพที่ 4 ประกอบไปด้วยส่วน contracting (ภาพที่ 4 ด้านซ้าย) และส่วน expansive (ภาพที่ 4 ด้านขวา) โดยส่วน contracting เป็นสถาปัตยกรรมโครงข่ายแบบคอนโวลูชัน ประกอบด้วยการประยุกต์ใช้ชั้นคอนโวลูชันขนาด 3×3 สองชั้นแต่ละชั้นตามด้วยฟังก์ชันกระตุ้น Rectified linear unit (ReLU) และชั้น max pooling ขนาด 2×2 โดย stride เท่ากับ 2 สำหรับดำเนินการ downsampling โดยที่แต่ละขั้นตอนของการ downsampling มีการเพิ่มจำนวนของ feature channels เป็น 2 เท่า ซึ่งในส่วนนี้มีเป้าหมายเพื่อทำการค้นหาองค์ประกอบภาพด้วยการ

เข้ารหัส (Encoder) อีกส่วนหนึ่งคือ ส่วน expansive ทุกขั้นตอนในส่วนนี้ประกอบด้วย การ upsampling ของ feature map โดยชั้นคอนโวลูชัน ขนาด 2×2 จำนวนครึ่งหนึ่งของ feature channels มีการเชื่อมต่อกับ crop feature map ที่สัมพันธ์กับส่วน contracting และชั้นคอนโวลูชัน ขนาด 3×3 สองชั้น แต่ละชั้นตามด้วยฟังก์ชันกระตุ้น ReLU ในส่วนของการ cropping เป็นสิ่งสำคัญ เนื่องจากมีการสูญเสียพิกเซลบริเวณขอบสำหรับการคอนโวลูชันแต่ละครั้ง และในขั้นสุดท้ายของสถาปัตยกรรม เป็นชั้นคอนโวลูชันขนาด 1×1 ถูกนำมาใช้เพื่อทำการ map แต่ละ feature vector จำนวน 64 องค์กรประกอบ กับจำนวนคลาสที่ต้องการ โดยรวมแล้วทั้งโครงข่ายมีชั้นคอนโวลูชัน 23 ชั้น



ภาพ 4 โครงสร้างและการดำเนินการของสถาปัตยกรรม U-net(29)

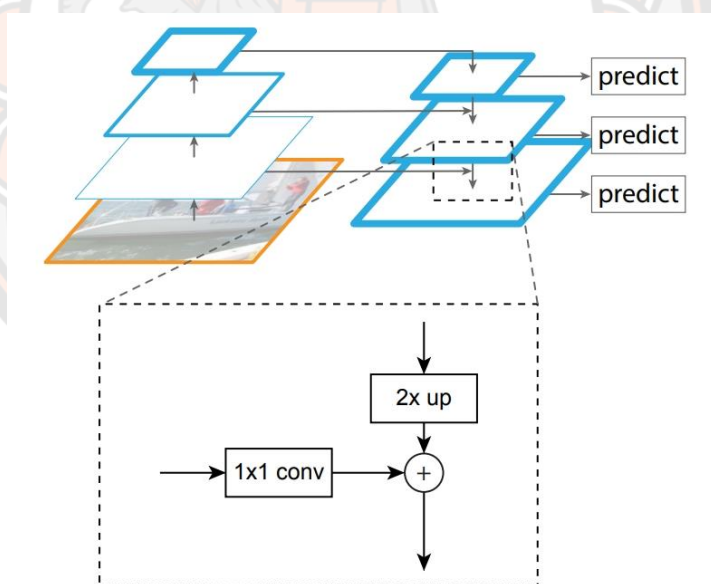
สถาปัตยกรรม Features Pyramid Networks (30)

สถาปัตยกรรม FPN เป็นสถาปัตยกรรมที่นำเสนอการใช้ประโยชน์จากความหลายมาตราส่วนบนระดับพีระมิดของโครงข่ายคอนโวลูชันเพื่อสร้างพีระมิดคุณลักษณะ โดยมีการเพิ่มการเชื่อมต่อด้านข้างเพื่อสร้างแผนภาพคุณลักษณะที่มีความหมายระดับสูงในทุกๆระดับชั้นพีระมิด รูปแบบการเชื่อมต่อของสถาปัตยกรรม FPN มี 3 รูปแบบได้แก่ การเชื่อมต่อในแนว bottom-up pathway การเชื่อมต่อในแนว top-down pathway และการเชื่อมต่อในแนว lateral connection โดยแต่ละแนวการเชื่อมต่อมีรายละเอียดดังนี้

bottom-up pathway ในขั้นนี้เป็นการประมวลผลแบบการดำเนินการไปข้างหน้า (feedforward computation) ของโครงข่าย backbone คอนโวลูชัน เพื่อทำการแยกคุณลักษณะภาพ โดยจะได้ภาพที่มีหลายมาตรส่วนที่ระดับพีระมิตต่างกัน โดยที่ระดับพีระมิตสูงความละเอียดเชิงพื้นที่จะลดลง ค่าความหมายคุณลักษณะภาพสูงขึ้น ผลลัพธ์จากขั้นสุดท้ายของแต่ละระดับพีระมิต ถูกใช้เป็นแผนภาพคุณลักษณะเพื่อดำเนินการแนว lateral แล้วผสานเข้ากับการดำเนินการในแนว top-down pathway

top-down pathway ภาพอ้างอิงจากผลลัพธ์สุดท้ายของแต่ละระดับพีระมิตในแนว bottom-up pathway ถูกใช้เป็นแผนภาพเริ่มต้นเพื่อทำการ upsamplig เพิ่มขนาดเชิงพื้นที่ให้มีขนาดเพิ่มเป็น 2 เท่า และข้อมูลจะถูกปรับความละเอียดเชิงพื้นที่ด้วยการผสาน (Addition method) เข้ากับผลลัพธ์จากการ convolution ของการเชื่อมต่อในแนว lateral connection จากนั้นผลลัพธ์ที่ได้จากการผสานคุณลักษณะแต่ละระดับพีระมิตในแนว top-down pathway จะถูก convolution เพื่อสร้างแผนภาพคุณลักษณะสุดท้ายรวมถึงการลด aliasing effect จากการ upsampling

โครงสร้างสถาปัตยกรรม FPN และการดำเนินการแสดงดังภาพ 5



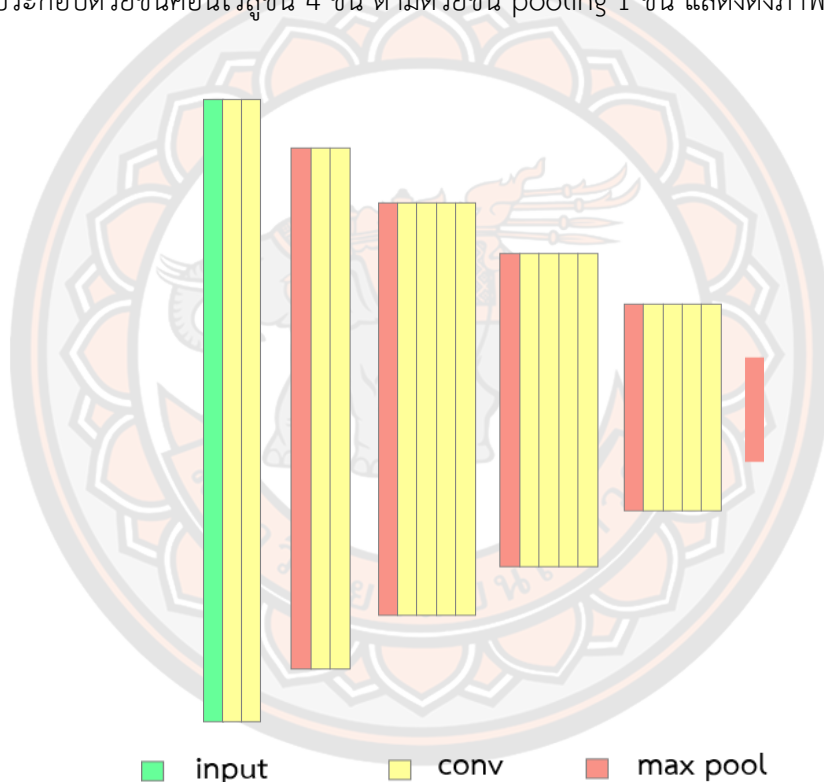
ภาพ 5 โครงสร้างและการดำเนินการของสถาปัตยกรรม FPN(30)

Backbone models

การแบ่งส่วนภาพ (segmentation) ด้วยโมเดลการเรียนรู้เชิงลึก มีส่วนการแยกแผนภาพคุณลักษณะ (feature extractor) ซึ่งเป็นส่วนสำคัญของโมเดล โดยสามารถใช้โมเดลการเรียนรู้เชิงลึกสำหรับการจำแนกภาพ (classification) ที่ถูกสอนด้วยชุดข้อมูลภาพขนาดใหญ่ มีค่าถ่วงน้ำหนักที่ดี

และไม่มีชั้นสุดท้ายของโมเดล (dense layer) มาใช้แทนส่วนการแยกแผนภาพคุณลักษณะในโมเดล การเรียนรู้เชิงลึกสำหรับการแบ่งส่วนภาพได้ โมเดลการจำแนกที่จะนำมาใช้แทนส่วนสำคัญดังกล่าวนี้ จึงถูกเรียกว่า โมเดล backbone ซึ่งการใช้โมเดล backbone ที่มีค่าถ่วงน้ำหนักที่ดีจะช่วยเพิ่ม ประสิทธิภาพ และลดเวลาการสอนโมเดลได้ เช่น VGG19 model (36), ResNet50 model (37), SE-ResNet50 model (38), InceptionV3 model (39) เป็นต้น โดยรายละเอียดการเชื่อมต่อ ชั้นดำเนินการของแต่ละรูปแบบโมเดล backbone แสดงดังนี้

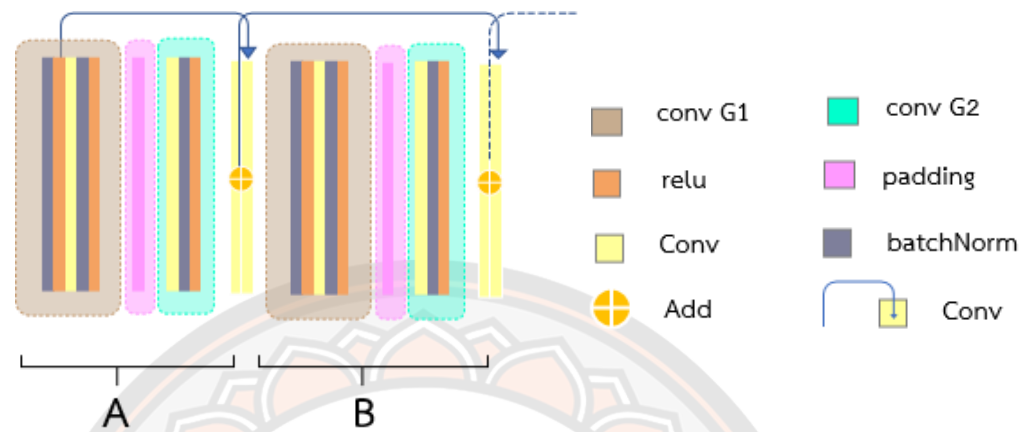
โมเดล backbone แบบ VGG19 ประกอบด้วย 5 block ใน block ที่ 1 และ 2 ประกอบด้วยชั้นคอนโวลูชัน 2 ชั้นเชื่อมต่อกับชั้น pooling 1 ชั้น ใน block ที่ 3 ถึง 5 ประกอบด้วยชั้นคอนโวลูชัน 4 ชั้น ตามด้วยชั้น pooling 1 ชั้น แสดงดังภาพ 6



ภาพ 6 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ VGG19

โมเดล backbone แบบ ResNet50 ประกอบด้วย 6 block ใน block ที่ 1 ประกอบด้วยชั้น input—batch normalization—padding เชื่อมต่ออย่างละ 1 ชั้น และ ต่อเข้ากับ block ที่ 2 ประกอบด้วยชั้น convolution—batch normalization—activation—padding—pooling และเชื่อมต่อกับ block 3 โดยใน block ที่ 3 ถึง 6 เป็น รูปแบบการเชื่อมต่อเฉพาะของโมเดล backbone แบบ ResNet50 ประกอบด้วย 4 stage แต่ละ stage อยู่ใน block ที่ 3 ถึง 6 ตามลำดับ ลักษณะการเชื่อมต่อประกอบด้วย

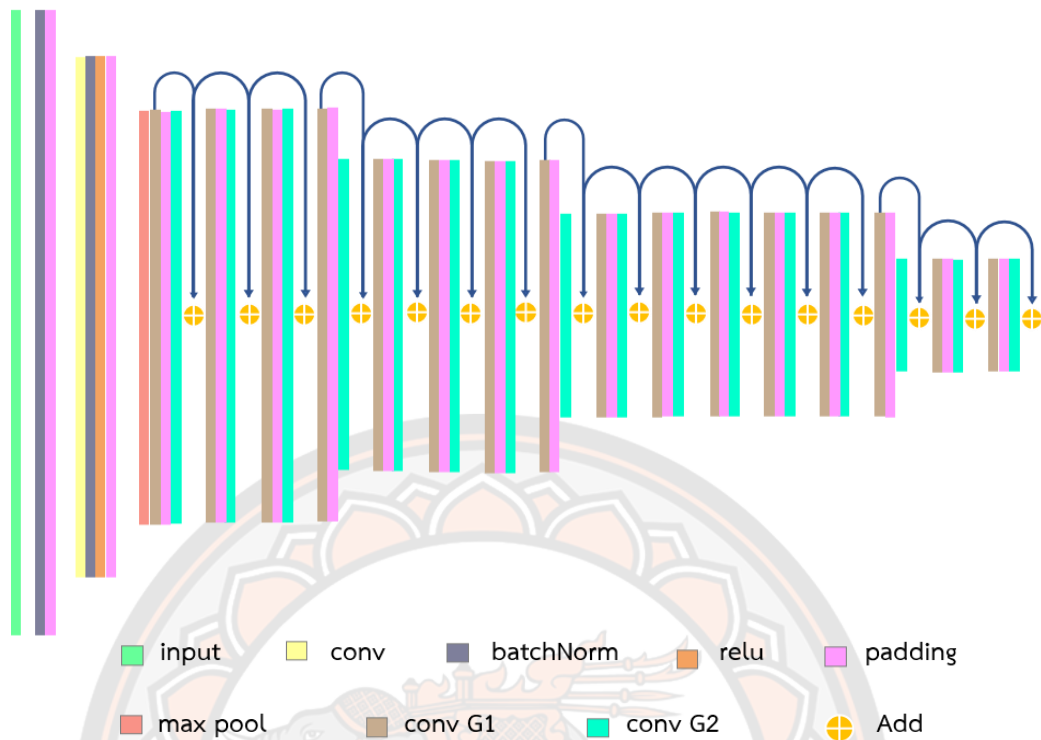
ชุดดำเนินการ convolution ซึ่ง 1 ชุดประกอบด้วย 2 กลุ่ม convolution แสดงดังภาพที่ 7B



ภาพ 7 การเชื่อมต่อชั้นดำเนินการของชุดดำเนินการ convolution กลุ่มที่ 1 และ กลุ่มที่ 2 โมเดล backbone แบบ ResNet50

กลุ่มที่ 1 มีการเชื่อมต่อของชั้น batch normalization—activation (Relu)—convolution—batch normalization—activation กลุ่มที่ 2 มีการเชื่อมต่อของชั้น convolution—batch normalization—activation ตามลำดับ ทั้งสองกลุ่มเชื่อมต่อกันด้วยชั้น padding 1 ชั้น การดำเนินการขั้นสุดท้ายของแต่ละชุดดำเนินการ convolution จะทำการคอนโวลูชันขั้นสุดท้ายของกลุ่มที่ 2 รวม (addition) เข้ากับผลจากการรวม (addition) ของชุดการ convolution ก่อนหน้าภายใน stage เดียวกัน ยกเว้นในชุดดำเนินการ convolution แรกของแต่ละ stage ที่ขั้นสุดท้ายจะทำการคอนโวลูชันขั้นสุดท้ายของกลุ่มที่ 2 รวมเข้ากับผลการคอนโวลูชันจากชั้น activation แรกของกลุ่มที่ 1 แสดงดังภาพที่ 7A

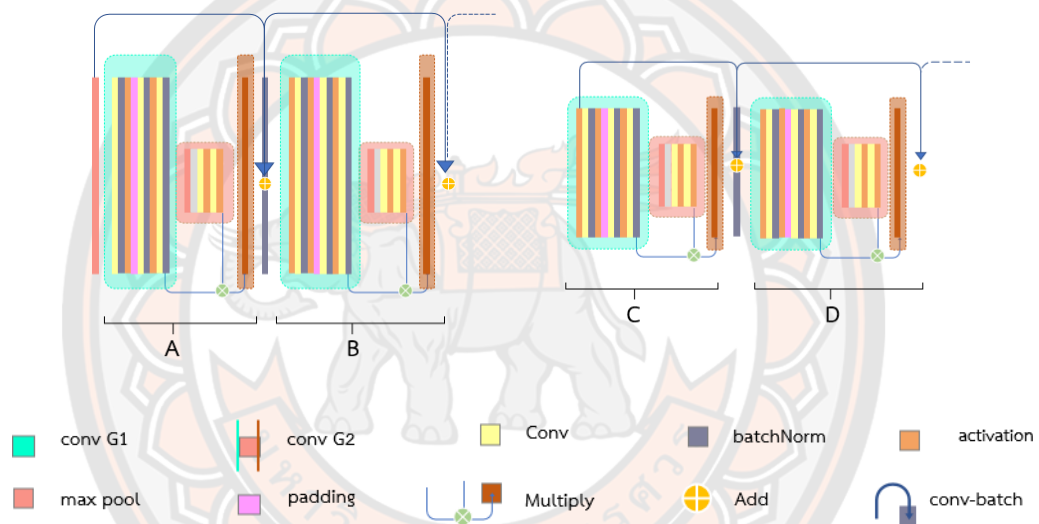
จำนวนชุดดำเนินการ convolution ใน stage 1 (block 3) มีจำนวน 3 ชุด stage 2 (block 4) มีจำนวน 4 ชุด stage 3 (block 5) มีจำนวน 6 ชุด และ stage 4 (block 6) มีจำนวน 3 ชุด แสดงดังภาพที่ 8



ภาพ 8 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ ResNet50

โมเดล backbone แบบ SE-ResNet50 ประกอบด้วย 6 block ใน block ที่ 1 ประกอบด้วยชั้น input เชื่อมเข้ากับชั้น padding และต่อเข้ากับ block ที่ 2 ประกอบด้วย การเชื่อมต่อกันของชั้น convolution—batch normalization—activation—padding—max pooling อย่างละ 1 ชั้น และเชื่อมต่อกับ block 3 โดยใน block ที่ 3 ถึง 6 เป็นรูปแบบการเชื่อมต่อเฉพาะของโมเดล backbone แบบ SE-ResNet50 ประกอบด้วย 4 stage โดยแต่ละ stage อยู่ใน block ที่ 3 ถึง 6 ตามลำดับ ลักษณะการเชื่อมต่อ ประกอบด้วยชุดดำเนินการ convolution 2 รูปแบบ รูปแบบที่ 1 ชุดดำเนินการ convolution แรกของแต่ละ stage ประกอบด้วย 2 กลุ่ม convolution แสดงดังภาพที่ 9C กลุ่มที่ 1 มีการเชื่อมต่อของชั้นดำเนินการต่าง ๆ ดังนี้ activation—convolution — batch normalization — activation— padding — convolution — batch normalization —activation —convolution — batch normalization กลุ่มที่ 2 มีการเชื่อมต่อของชั้นดำเนินการ ดังนี้ lambda—convolution—activation—convolution—activation ทั้งสองกลุ่มเชื่อมต่อกันด้วยชั้น max pooling การดำเนินการขั้นตอนสุดท้ายของชุดดำเนินการ convolution รูปแบบที่ 1 ทำการ convolution—batch normalization

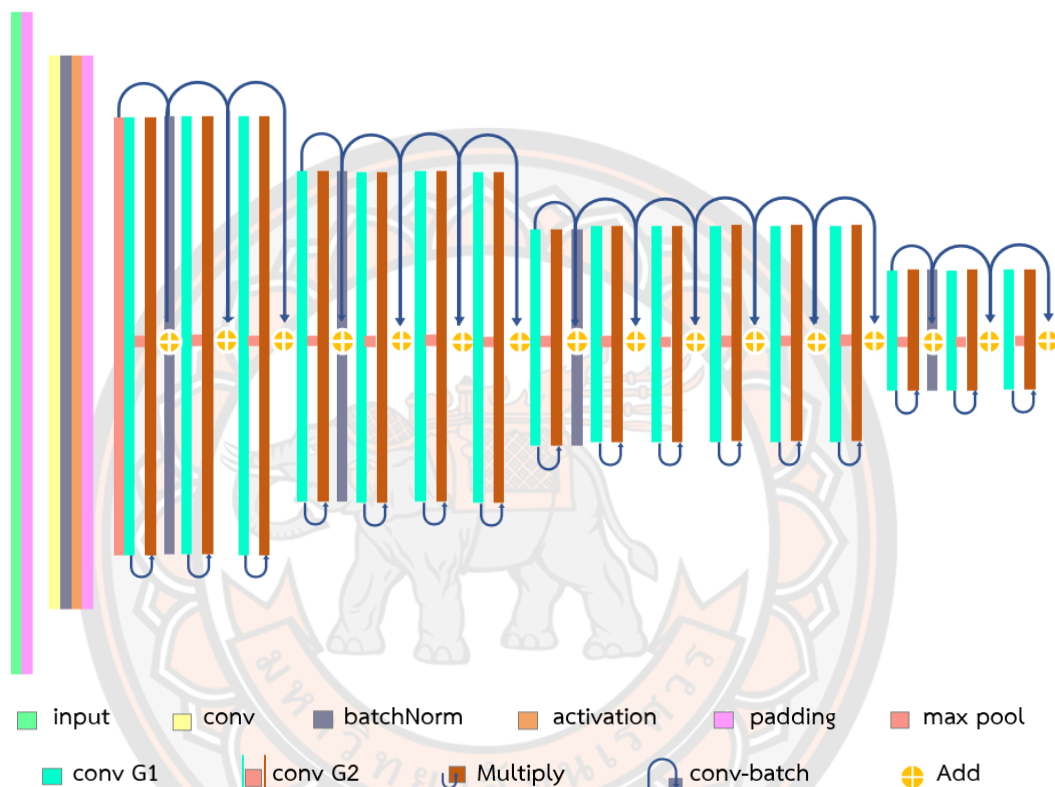
ชั้น activation แรกของกลุ่มที่ 1 รวม (addition) เข้ากับผลจากการคูณ (multiply) ของชั้น batch normalization สุดท้ายของกลุ่มที่ 1 และชั้น activation สุดท้ายของกลุ่มที่ 2 ยกเว้นในชุดดำเนินการ convolution รูปแบบที่ 1 ของ stage 1 (block 3) แสดงดังภาพที่ 9A จะเห็นได้ว่าชุดดำเนินการ convolution นี้ไม่มีชั้น activation ก่อนหน้าชั้น convolution แรก ดังนั้นการดำเนินการขั้นสุดท้ายของชุดดำเนินการนี้จะทำการ convolution — batch normalization ชั้น Max pooling ของ block 2 เพื่อรวม (addition) เข้ากับผลจากการคูณ (multiply) ของชั้น batch normalization สุดท้ายของกลุ่มที่ 1 และชั้น activation สุดท้ายของกลุ่มที่ 2



ภาพ 9 การเชื่อมต่อชั้นดำเนินการของชุดดำเนินการ convolution รูปแบบที่ 1 และ รูปแบบที่ 2 โมเดล backbone แบบ SE-ResNet50

ชุดดำเนินการ convolution รูปแบบที่ 2 ชุดดำเนินการ convolution อื่นในแต่ละ stage ประกอบด้วย 2 กลุ่ม convolution แสดงดังภาพที่ 9B และ 9D ทั้ง 2 กลุ่ม convolution มีชั้นดำเนินการเช่นเดียวกับชุดดำเนินการ convolution รูปแบบที่ 1 (ยกเว้นรูปแบบที่ 1 ของ stage 1 (block 3)) โดยที่การดำเนินการในขั้นตอนสุดท้ายจะใช้ผลของการรวม (addition) ของชุดดำเนินการ convolution ก่อนหน้า (แทนการ convolution—batch normalization ชั้น activation แรกของกลุ่มที่ 1) เพื่อนำผลมารวม (addition) เข้ากับผลจากการคูณ (multiply) ของชั้น batch normalization สุดท้ายของกลุ่มที่ 1 และชั้น activation สุดท้ายของกลุ่มที่ 2 ของชุดดำเนินการนี้

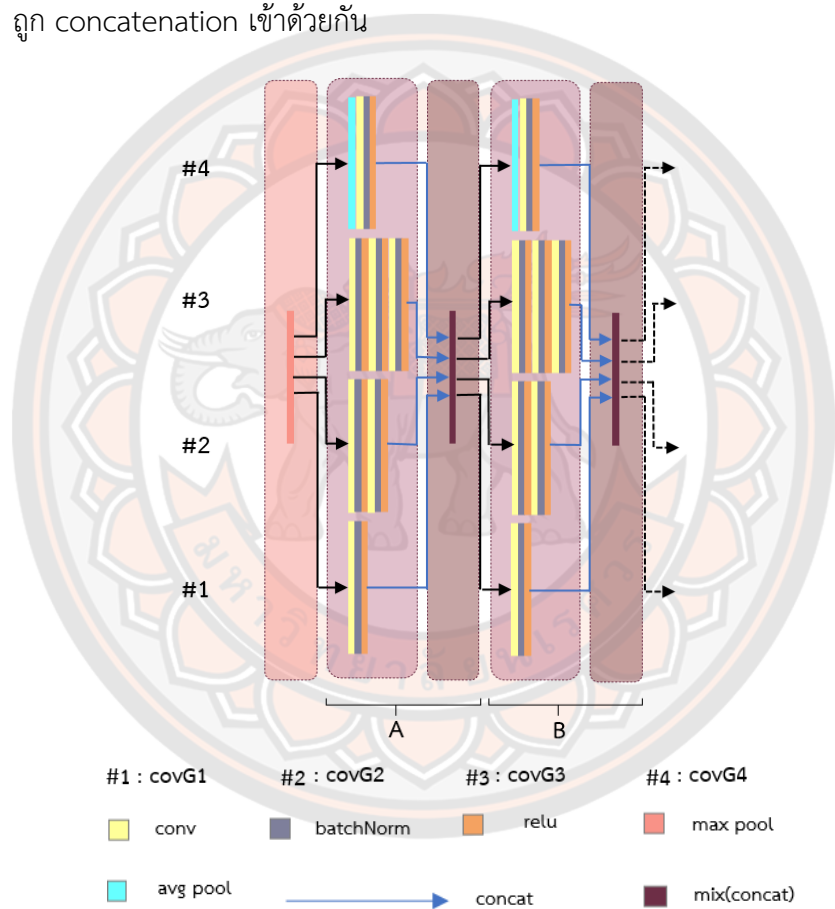
จำนวนชุดดำเนินการ convolution ใน stage 1 (block 3) มีจำนวน 3 ชุด stage 2 (block 4) มีจำนวน 4 ชุด stage 3 (block 5) มีจำนวน 6 ชุด และ stage 4 (block 6) มีจำนวน 3 ชุด โดยชุดดำเนินการ convolution แรกของแต่ละ block ใช้รูปแบบ 1 แล้วตามด้วยรูปแบบที่ 2 จนครบตามจำนวนดังกล่าวข้างต้น แสดงดังภาพที่ 10



ภาพ 10 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ SE-ResNet50

โมเดล backbone แบบ InceptionV3 ประกอบด้วย 6 block ใน block ที่ 1 ประกอบด้วยชั้น input ต่อเข้ากับ block 2 ซึ่งต่อเข้ากับ block 3 ประกอบด้วย การเชื่อมต่อกันของชั้น convolution—batch normalization—activation จำนวน 3 ชุด และ 2 ชุด ใน block 2 และ 3 ตามลำดับ แล้วตามด้วยชั้น max pooling ในชั้นสุดท้าย สำหรับแต่ละ block การเชื่อมต่อของ block ที่ 4 ถึง 6 เป็นรูปแบบการเชื่อมต่อเฉพาะของ โมเดล backbone แบบ InceptionV3 ที่มีลักษณะการเชื่อมต่อแบบ module จำนวน 3 module โดยแต่ละ module อยู่ใน block ที่ 4 ถึง 6 ตามลำดับ

module 1 มีรูปแบบการเชื่อมต่อของชั้นดำเนินการแสดงดังภาพ 11 ซึ่งถูกเชื่อมต่อจากชั้น max pooling ของ block 3 ประกอบด้วย 4 กลุ่มดำเนินการ (#1, #2, #3, #4) มีการเชื่อมต่อของชั้น convolution—batch normalization—activation ในแต่ละกลุ่มดำเนินการจำนวน 1 ชุด 2 ชุด และ 3 ชุด ในกลุ่มการดำเนินการที่ 1, 2 และ 3 ตามลำดับ ในกลุ่มที่ 4 มีการเชื่อมต่อของชั้น average pooling ในชั้นแรกตามด้วยการเชื่อมต่อของชั้น convolution—batch normalization—activation การดำเนินการชั้นตอนสุดท้ายของ module 1 ผลจากชั้น activation สุดท้ายของแต่ละกลุ่มดำเนินการถูก concatenation เข้าด้วยกัน



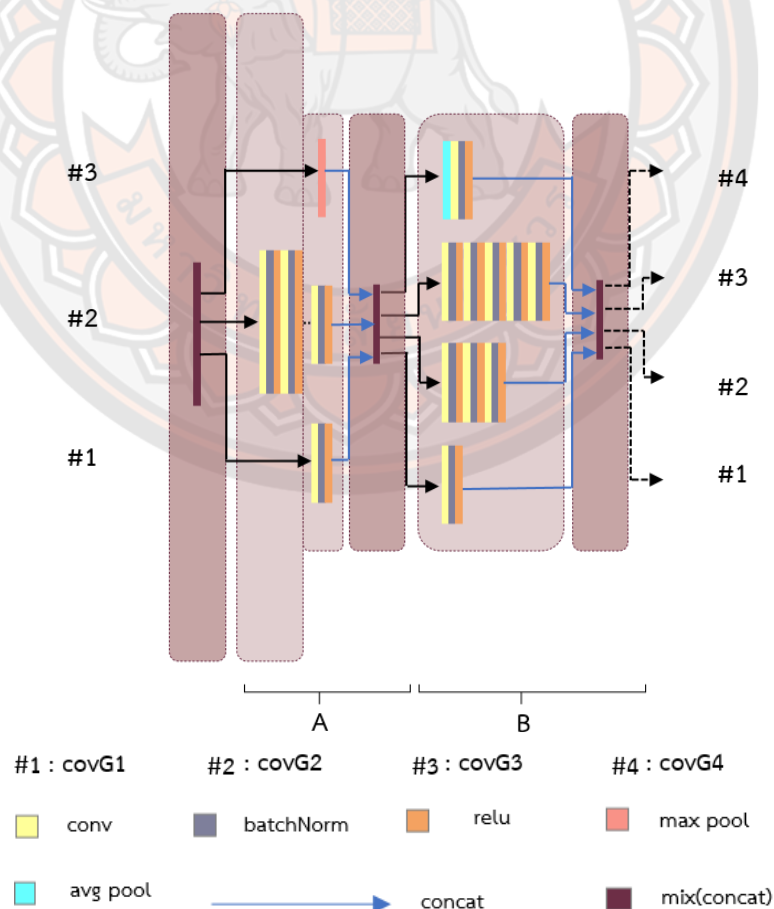
ภาพ 11 การเชื่อมต่อชั้นดำเนินการของ module 1 โมเดล backbone แบบ InceptionV3

module 2 มีรูปแบบการเชื่อมต่อของชั้นดำเนินการ 2 รูปแบบ

รูปแบบที่ 1 (ภาพที่ 12A) การดำเนินการ convolution ชุดแรกของ block 5 ซึ่งถูกเชื่อมต่อกับชั้น concatenation สุดท้ายของ block 4 ประกอบด้วย 3 กลุ่มดำเนินการ (#1, #2, #3) มีการเชื่อมต่อของชั้น convolution—batch normalization—activation

ในแต่ละกลุ่มดำเนินการจำนวน 1 ชุด และ 3 ชุด ในกลุ่มการดำเนินการที่ 1 และ 2 ตามลำดับ ในกลุ่มที่ 3 ประกอบด้วยการดำเนินการของชั้น max pooling การดำเนินการขั้นสุดท้ายของ module 2 รูปแบบที่ 1 (ภาพที่ 12A) ผลจากขั้นสุดท้ายของแต่ละกลุ่มดำเนินการถูก concatenation เข้าด้วยกัน

รูปแบบที่ 2 (ภาพที่ 12B) การดำเนินการ convolution อื่น ๆ ของ block 5 ซึ่งถูกเชื่อมต่อกับชั้น concatenation ของการดำเนินการก่อนหน้าประกอบด้วย 4 กลุ่มดำเนินการ (#1, #2, #3, #4) มีการเชื่อมต่อของชั้น convolution—batch normalization—activation ในแต่ละกลุ่มดำเนินการจำนวน 1 ชุด 3 ชุด และ 5 ชุด ในกลุ่มการดำเนินการที่ 1 2 และ 3 ตามลำดับ ในกลุ่มที่ 4 มีการเชื่อมต่อของชั้น Average pooling ในชั้นแรกตามด้วยการเชื่อมต่อของชั้น convolution—batch normalization—activation ขั้นสุดท้ายของ module 2 รูปแบบที่ 2 (ภาพที่ 12B) ผลจากการดำเนินการในชั้น activation สุดท้ายของแต่ละกลุ่มดำเนินการถูก concatenation เข้าด้วยกัน

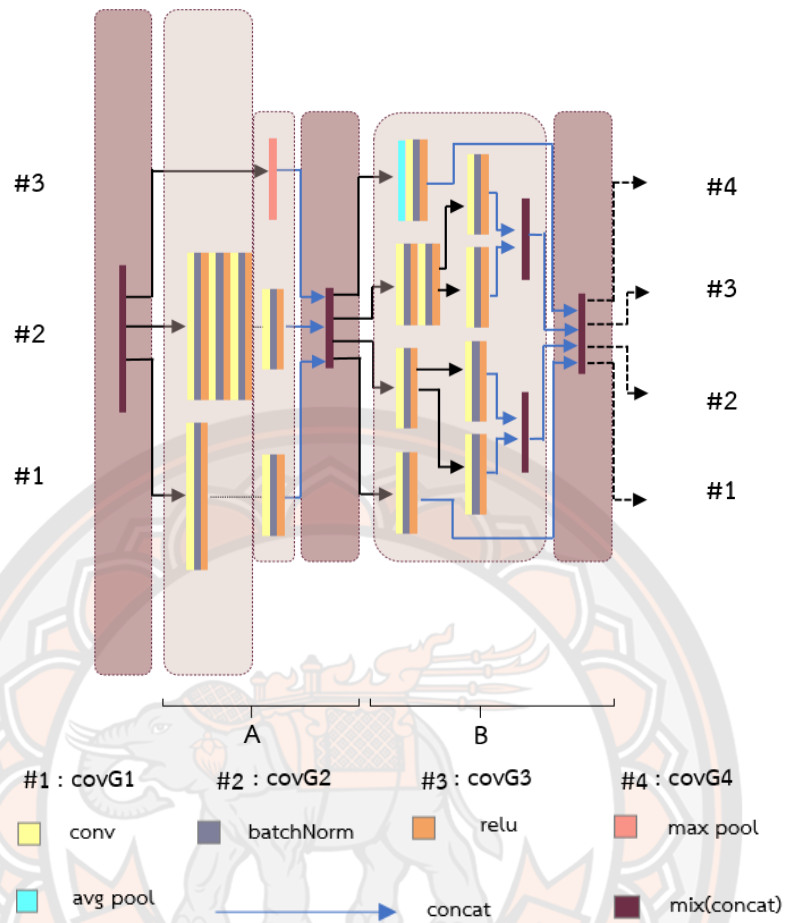


ภาพ 12 การเชื่อมต่อชั้นดำเนินการของ module 2 โมเดล backbone แบบ InceptionV3

module 3 มีรูปแบบการเชื่อมต่อของชั้นดำเนินการ 2 รูปแบบ

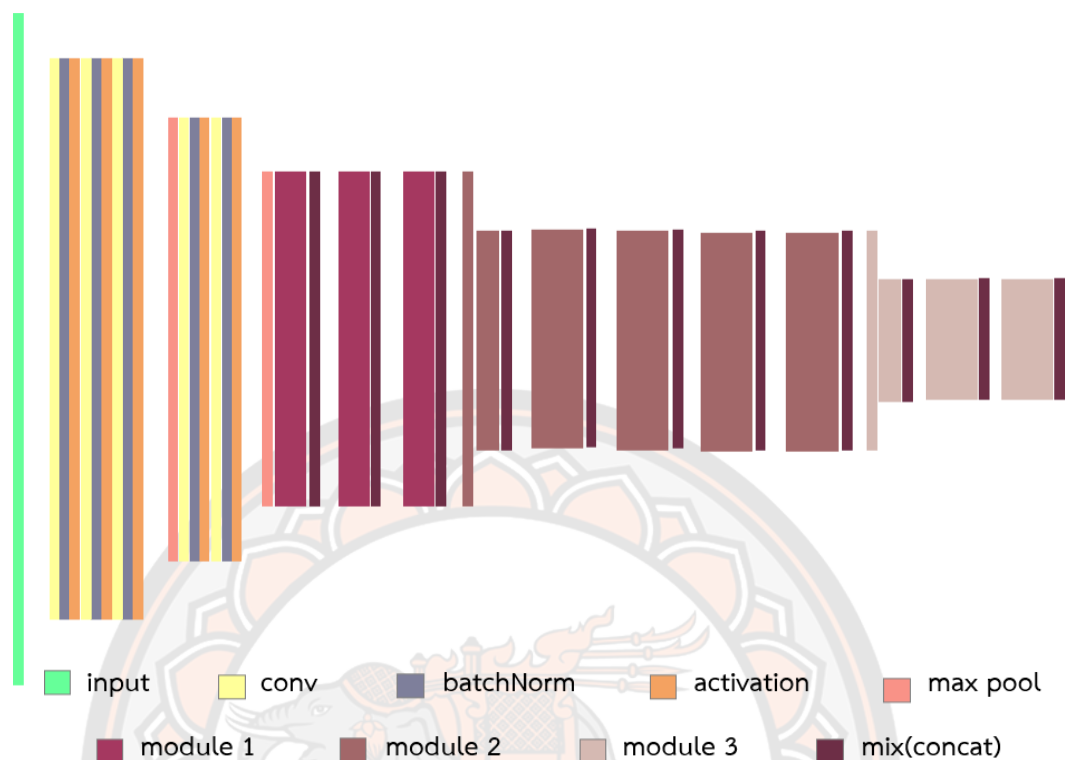
รูปแบบที่ 1 (ภาพที่ 13A) การดำเนินการ convolution ชุดแรกของ block 6 ซึ่งถูกเชื่อมต่อกับชั้น concatenation สุดท้ายของ block 5 ประกอบด้วย 3 กลุ่มดำเนินการ (#1, #2, #3) มีการเชื่อมต่อของชั้น convolution—batch normalization—activation ในแต่ละกลุ่มดำเนินการจำนวน 2 ชุด และ 4 ชุด ในกลุ่มการดำเนินการที่ 1 และ 2 ตามลำดับ ในกลุ่มที่ 3 ประกอบด้วยการดำเนินการของชั้น max pooling การดำเนินการชั้นตอนสุดท้ายของ module 3 รูปแบบที่ 1 (ภาพที่ 13A) ผลจากการดำเนินการในชั้นสุดท้ายของแต่ละกลุ่มดำเนินการถูก concatenation เข้าด้วยกัน

รูปแบบที่ 2 (ภาพที่ 13B) การดำเนินการ convolution อื่น ๆ ของ block 5 ซึ่งถูกเชื่อมต่อกับชั้น concatenation ของการดำเนินการก่อนหน้าประกอบด้วย 4 กลุ่มดำเนินการ (#1, #2, #3, #4) มีการเชื่อมต่อของชั้น convolution — batch normalization—activation ในแต่ละกลุ่มดำเนินการจำนวน 1 ชุด 1 ชุด และ 2 ชุด ในกลุ่มการดำเนินการที่ 1 2 และ 3 ตามลำดับ ในกลุ่มที่ 4 ประกอบด้วยการดำเนินการของชั้น average pooling ในชั้นแรกตามด้วยการเชื่อมต่อของชั้น convolution—batch normalization—activation โดยในกลุ่มที่ 2 และ 3 จะถูกดำเนินการ convolution ด้วยกลุ่มดำเนินการย่อยที่มีการต่อเชื่อมของชั้น convolution—batch normalization—activation อีกกลุ่มละ 2 กลุ่มย่อย และผลจากการดำเนินการของ 2 กลุ่มย่อยนี้จะถูก concatenation เข้าด้วยกันและใช้เป็นผลสุดท้ายของกลุ่มดำเนินการที่ 2 และ 3 การดำเนินการชั้นตอนสุดท้ายของ module 3 รูปแบบที่ 2 (ภาพที่ 13B) ผลจากชั้นสุดท้ายของแต่ละกลุ่มดำเนินการถูก concatenation เข้าด้วยกัน



ภาพ 13 การเชื่อมต่อชั้นดำเนินการของ module 3 โมเดล backbone แบบ InceptionV3

โดยใน block 4 ใช้รูปแบบของชั้นดำเนินการของ module 1 จำนวน 3 ชุด ใน block ที่ 5 ใช้รูปแบบการต่อเชื่อมกันของ module 2 รูปแบบที่ 1 จำนวน 1 ชุด รูปแบบที่ 2 จำนวน 4 ชุด และใน block ที่ 6 ใช้รูปแบบการต่อเชื่อมกันของ module 3 รูปแบบที่ 1 จำนวน 1 ชุด รูปแบบที่ 2 จำนวน 2 ชุด แสดงการเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ InceptionV3 ดังภาพที่ 14



ภาพ 14 การเชื่อมต่อชั้นดำเนินการของโมเดล backbone แบบ InceptionV3

Transfer learning and Fine-tuning (40)

การถ่ายทอดการเรียนรู้และการปรับจูน เป็นวิธีการใช้คุณลักษณะที่ถูกเรียนรู้จากปัญหาอื่นก่อนหน้าแล้ว นำมาใช้ประโยชน์กับการสอนโมเดลสำหรับปัญหารูปแบบใหม่ที่ใกล้เคียงกัน การสอนโมเดลใหม่ด้วยวิธีถ่ายทอดการเรียนรู้และวิธีปรับจูน โมเดลสำหรับปัญหาใหม่นี้จะต้องใช้ชั้นดำเนินการของโมเดลที่ถูกสอนมาก่อนแล้วเป็นส่วนหนึ่งของโมเดลใหม่ มีการเพิ่มชั้นดำเนินการตามรูปแบบปัญหาใหม่ จากนั้นทำการสอนโมเดล หากใช้วิธีถ่ายทอดการเรียนรู้ต้องทำการตรึงค่าพารามิเตอร์ในส่วนของโมเดลที่ถูกสอนมาก่อนแล้วในขณะที่สอนโมเดลใหม่ แต่ถ้าใช้วิธีการปรับจูนจะไม่ทำการตรึงค่าพารามิเตอร์ในส่วนของโมเดลที่ถูกสอนมาก่อนแล้ว แต่ค่าพารามิเตอร์ในส่วนนี้จะถูกใช้เป็นตัวตั้งต้นสำหรับการสอนโมเดลแทนการเริ่มต้นการสอนด้วยค่าตัวตั้งต้นแบบสุ่ม

โรคมะเร็งศีรษะและลำคอ (41-43)

โรคมะเร็งศีรษะและลำคอ มักเกิดจาก squamous cell ที่เรียงตัวตามพื้นผิวเยื่อเมือกต่าง ๆ ในศีรษะและลำคอ เช่น ปาก จมูก ลำคอ เป็นต้น เซลล์ squamous ที่ผิดปกติเป็นมะเร็งจะถูกเรียกว่า squamous cell carcinoma มะเร็งศีรษะและลำคอจะถูกแบ่งตามตำแหน่งที่พบก่อนมะเร็ง เช่น ช่องปาก คอหอย กล่องเสียง โพรงจมูก ต่อม้ำลาย เป็นต้น

สาเหตุของการเกิดมะเร็งบริเวณศีรษะและลำคอ ได้แก่ การสูบบุหรี่ การดื่มแอลกอฮอล์ การติดเชื้อไวรัส Human papilloma virus (HPV) โดยเฉพาะอย่างยิ่ง HPV ชนิดที่ 16 และปัจจัยเสี่ยงอื่น ๆ เช่น การเคี้ยวหมาก อาหารหมักดอง สุขภาพช่องปาก การประกอบอาชีพที่ได้รับ การสัมผัสแร่ใยหินและเส้นใยสังเคราะห์ การติดเชื้อไวรัส Epstein-barr เป็นปัจจัยเสี่ยงของมะเร็งหลังโพรงจมูก เป็นต้น

อาการของมะเร็งศีรษะและลำคอ เช่น การพบก้อนเนื้อส่วนต่าง ๆ ของศีรษะและลำคอ อาการกลืนลำบาก อาการเสียงเปลี่ยน มีเลือดออกทางจมูก อาการบวมใต้คางหรือบริเวณโดยรอบ เป็นต้น การวินิจฉัย แพทย์จะทำการซักประวัติ ตรวจร่างกาย การส่งตรวจพิเศษ เช่น การส่องกล้อง การตรวจทางรังสีวิทยา เป็นต้น การยืนยันผลผู้ป่วยมะเร็งยืนยันด้วยผลการตรวจทางพยาธิวิทยา การรักษา มะเร็งศีรษะและลำคอ ผู้ป่วยจะถูกวางแผนการรักษาสำหรับแต่ละรายมีหลายปัจจัยในการพิจารณา เช่น ตำแหน่งของก้อนมะเร็ง ระยะการดำเนินการของโรค อายุและสุขภาพของผู้ป่วย เป็นต้น วิธีการรักษาหลักได้แก่ การผ่าตัด การฉายรังสี และการให้ยาเคมีบำบัด หรืออาจใช้การรักษาร่วมกันหลายวิธี

การรักษาผู้ป่วยมะเร็งด้วยการฉายรังสีค่อนข้างได้รับการยอมรับ เนื่องจากเป็นการรักษาที่ไม่ลุกล้ำหากเทียบกับการผ่าตัด และมีผลข้างเคียงต่อระบบภูมิคุ้มกันของร่างกายน้อยกว่า การใช้ยาเคมีบำบัด แต่การใช้รังสีก็มีข้อจำกัดกล่าวคือ อวัยวะปกติที่อยู่รอบก้อนมะเร็งที่ต้องการฉายรังสีจะได้รับปริมาณรังสีด้วย ซึ่งอาจจะทำให้เกิดการสูญเสียการทำงานได้ และรังสีอาจเหนี่ยวนำให้เกิดโรคมะเร็งทุติยภูมิ การกำหนดปริมาณรังสีเพื่อการรักษาแบบหายขาดที่เป็นมาตรฐานจะกำหนดให้รังสี 2 Gy ต่อ fraction วันละ 1 fraction 5 วันต่อเนื่องใน 1 สัปดาห์ สำหรับปริมาณรังสีรวมกำหนดที่ 66-70 Gy สำหรับการรักษามะเร็งศีรษะและลำคอแบบหายขาด หากเป็นกรณีอื่น ๆ ต้องมีปรับเปลี่ยนตามคู่มือแนะนำ

อวัยวะที่มีความเสี่ยง (42)

อวัยวะที่มีความเสี่ยง (Organs at risk) คือ เนื้อเยื่อปกติโดยรอบก้อนมะเร็ง หรืออาจรวมอยู่ในปริมาณที่ทำให้การฉายรังสี ซึ่งมีความเสี่ยงที่รังสีอาจทำให้การทำงานของอวัยวะเหล่านั้นแย่งลง เช่น การเกิดภาวะน้ำลายแห้ง (Xerostomia) (24) การตายของกระดูกขากรรไกรหลังการฉายรังสี (Post extraction osteoradionecrosis) (25) การเกิดความผิดปกติต่อดวงตาและระบบประสาทตา (26) การเกิดอาการผิดปกติจากการสูญเสียการทำงานของก้านสมองและไขสันหลัง (27) เป็นต้น ทำให้การวางแผนการรักษาไม่ได้หมายความถึงแค่การพิจารณาขอบเขตปริมาณของก้อนมะเร็งและขอบเขตที่มีการขยายเท่านั้น แต่ยังรวมถึงอวัยวะที่มีความเสี่ยงอื่น ๆ ด้วย ดังนั้นจึงมีการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงตามโปรโตคอล เพื่อใช้ประเมินปริมาณรังสีและผลกระทบ อวัยวะที่มีความเสี่ยงของการฉายรังสีบริเวณศีรษะและลำคอ และขีดจำกัดปริมาณรังสีสูงสุดแสดงดังตารางที่ 2

ตาราง 2 ค่าขีดจำกัดปริมาณรังสีสูงสุดของอวัยวะที่มีความเสี่ยงบริเวณศีรษะและลำคอ

| อวัยวะที่มีความเสี่ยง | ขีดจำกัดปริมาณรังสีสูงสุด (RTOG 1016) |
|-----------------------------------|---------------------------------------|
| spinal cord | maximum dose 50 Gy |
| brain stem | maximum dose 52 Gy |
| eyes | mean dose < 35 Gy (RTOG 0225) |
| lens | maximum dose 10 Gy |
| optic nerve | maximum dose 54 Gy |
| optic chiasm | maximum dose 54 Gy |
| lips | mean dose < 20 Gy |
| oral cavity | mean dose < 30 Gy |
| parotid glands | mean dose < 26 Gy |
| contralateral submandibular gland | mean dose < 39 Gy |
| cervical esophagus | mean dose < 30 Gy |
| glottic and supraglottic larynx | mean dose < 20 Gy |
| mandible | maximum dose 66 Gy |

งานวิจัยที่เกี่ยวข้อง

Ibragimov B. และ Xing L. (2017) (13) พัฒนาอัลกอริทึมการเรียนรู้เชิงลึกสำหรับการแบ่งส่วนภาพอวัยวะที่มีความเสี่ยงในภาพ CT ศีรษะและลำคอ ทำการเปรียบเทียบประสิทธิภาพอัลกอริทึมที่พัฒนาขึ้นกับอัลกอริทึมทางการค้า และความแปรปรวนในการแบ่งส่วนภาพของผู้สังเกต (inter-observer variability) โครงข่ายประสาทเทียม convolution ถูกนำมาใช้เพื่อเรียนรู้รูปแบบความสัมพันธ์ของความเข้ม (Intensity) บนภาพ CT ที่ใช้สอน โดยแสดงรูปแบบความเข้มแบบบวกลบใน voxel ของ OARs ที่สนใจ และรูปแบบความเข้มแบบลบใน voxel อื่นโดยรอบ OARs ผ่านขั้นตอนการของโครงข่ายประสาทเทียมเพื่อตรวจจับภาพคุณลักษณะ โครงข่ายประสาทเทียมที่ผ่านการสอนแล้วถูกนำไปจำแนก voxel ส่วนที่สนใจในชุดภาพทดสอบซึ่งสัมพันธ์กับ OARs และทำการปรับปรุงผลการจำแนก voxel ด้วยอัลกอริทึม Markov random fields ขั้นตอนสุดท้ายทำการแยกส่วน voxel ที่มีการเชื่อมต่อกันมากที่สุดจะถูกจำแนกเป็นส่วนของ OARs ทำการ dilate-erode เพื่อกำจัดช่องว่าง ได้ผลลัพธ์ของการแบ่งส่วนภาพในชุดข้อมูลภาพทดสอบ ประสิทธิภาพของโครงข่ายประสาทเทียมถูกทดสอบด้วยการแบ่งส่วนภาพอวัยวะ spinal cord, mandible, parotid glands, submandibular glands, larynx, pharynx, eye globes, optic nerves และ optic chiasm โดยใช้ภาพ CT 50 ภาพ ผลการแบ่งส่วนภาพถูกคำนวณหาค่า DSC ผลค่า DSC มีค่าอยู่ระหว่าง 37.4% สำหรับ chiasm ถึง 89.5% สำหรับ mandible โดยภาพรวมผลที่ได้แสดงถึงประสิทธิภาพที่คล้ายกันหรือเหนือกว่างานอื่นก่อนหน้านี้

Wang Y. และคณะ (2017) (14) เสนอวิธีการแบ่งส่วนภาพอวัยวะที่มีความเสี่ยงบนภาพ CT ศีรษะและลำคอ โดยใช้โครงข่ายการแบ่งส่วน OARs บนพื้นฐานสถาปัตยกรรม 3D U-Net ประกอบด้วย 2 ขั้นตอน ได้แก่ 1.) การสร้างกรอบตำแหน่งของ OARs และ 2.) การแบ่งส่วนจากปริมาตรขนาดเล็กภายในกรอบจากขั้นตอนที่ 1 และทั้งสองขั้นตอนจะใช้สถาปัตยกรรมแยกเฉพาะสำหรับแต่ละขั้นตอน (ขั้นตอนที่ 1 LocNet, ขั้นตอนที่ 2 SegNet) ซึ่งการแบ่งส่วนย่อย 2 ขั้นตอนนี้ทำให้ในแต่ละส่วนย่อยทำงานได้ง่ายขึ้น ส่งผลให้โครงข่ายโดยรวมสมบูรณ์ขึ้น ใช้ชุดข้อมูลภาพจาก The MICCAI 2015 challenge ในการศึกษา เพื่อให้สามารถเปรียบเทียบผลกับวิธีการอื่นได้ โดยข้อมูลถูกแบ่งใช้สำหรับการสอน 33 ชุดข้อมูล ใช้สำหรับการประเมิน 15 ชุดข้อมูล ประเมินด้วยค่า 95%HD และค่า DSC ผลการประเมินค่า 95%HD จากวิธีการที่นำเสนอได้รับการจัดอันดับสูงสุด ใน 7 OARs จากทั้งหมด 9 OARs และเป็นอันดับ 2 ใน OARs อื่น ๆ ผลค่า DSC ได้รับการจัดอันดับสูงสุด ใน 5 OARs จากทั้งหมด 9 OARs และเป็นอันดับ 2 ใน OARs อื่น ๆ โดยมีความแตกต่างเล็กน้อยจากค่าที่ถูกจัดเป็นอันดับสูงสุดที่ OARs อื่น ๆ

Wu X. และคณะ (2018) (15) นำเสนอการใช้วิธี Automatic Anatomy Recognition (AAR) เพื่อกำหนดขอบเขตหลายวัตถุบนภาพ CT จำลองการรักษาบริเวณศีรษะและลำคออัตโนมัติ

ใช้ชุดข้อมูลภาพ CT จำลองการรักษาจำนวน 216 ชุด ประกอบด้วยภาพวัตถุตัวอย่าง 2600 ภาพ ซึ่งวัตถุตัวอย่างได้แก่ skin outer boundary, left and right parotid glands, left and right submandibular glands, cervical esophagus, supraglottic and glottic larynx, cervical spinal canal, mandible และ orohypopharynx constrictor muscle วิธีการศึกษาประกอบด้วย 3 ขั้นตอนได้แก่ 1.) การสร้างโมเดล fuzzy เพื่อหาลำดับชั้นที่เหมาะสมสำหรับการจัดเรียงวัตถุตามความสัมพันธ์ระหว่างวัตถุ ใช้ข้อมูลภาพที่มีคุณภาพดีจากกลุ่มประชากรภาพ CT ศีรษะและลำคอในการศึกษา 2.) การจดจำวัตถุ ใช้ประโยชน์จากข้อมูลทางกายวิภาคก่อนหน้านี้ที่ทำการเข้ารหัสลำดับชั้นต่าง ๆ เพื่อจัดวางตำแหน่งสำหรับแต่ละวัตถุ ขั้นตอนสุดท้ายใช้อัลกอริทึมการกำหนดขอบเขตแยกแผนภาพคุณลักษณะ เพื่อกำหนดขอบเขตตามผลการจดจำวัตถุ โดยวิธี AAR ที่ใช้ในการศึกษานี้ผ่านการปรับปรุงส่วนประกอบใน framework เช่น การสร้างแบบจำลองความสัมพันธ์ของขอบเขต การรวม texture กับ intensity เข้าด้วยกัน และการประเมินคุณภาพวัตถุ เป็นต้น ประเมินผลโดยพิจารณาจากค่าความคลาดเคลื่อนของตำแหน่งวัตถุ ค่า DSC และค่า HD ผลการศึกษาพบว่าความถูกต้องของการจดจำตำแหน่งโดยรวมมีความคลาดเคลื่อน 2 voxel ผลค่า DSC โดยรวมใกล้เคียงกับ 0.8 และค่า HD อยู่ภายใน 1 voxel แม้ในชุดข้อมูลภาพจะมีสิ่งแปลกปลอม (streak artifacts) เล็กน้อย (<4 สไลซ์)

Tong N. และคณะ (2018) (16) พัฒนาการแบ่งส่วนภาพอวัยวะที่มีความเสี่ยงอัตโนมัติบริเวณศีรษะและลำคอแบบใหม่ที่ใช้ FCNN ร่วมกับ shape representation model (SRM) โดยโมเดล SRM และ FCNN ถูกสอนด้วย 2 ขั้นตอน 1.) โมเดล SRM ถูกสอนให้เรียนรู้รูปแบบการแสดงออกของรูปร่างแฝงของ OARs จากชุดข้อมูลสอน 2.) ค่าพารามิเตอร์ที่ถูกสอนตามขั้นตอน 1 ของโมเดล SRM ถูกยึดตรึงไว้ เพื่อใช้สอนโมเดล FCNN โมเดลทั้งสองถูกรวมเข้าด้วยกันเพื่อใช้กำหนดขอบเขต OARs 9 อวัยวะ ได้แก่ brain stem, optic chiasm, mandible, optical nerves, parotids glands และ submandibular glands ทำการศึกษาด้วยข้อมูลภาพ CT 32 ชุดภาพ ถูกแบ่งเป็นชุดข้อมูลสอน 22 ชุดภาพ และสำหรับทดสอบโมเดล 10 ชุดภาพ ขั้นตอนการสอนใช้วิธีการ augmentation เพื่อเพิ่มชุดข้อมูลภาพ ประเมินความถูกต้องเชิงปริมาณของการแบ่งส่วนภาพด้วยค่า DSC, positive predictive value, sensitivity, average surface distance และ 95% maximum surface distance ผลค่า DSC เท่ากับ 0.870, 0.583, 0.937, 0.653, 0.689, 0.835, 0.832, 0.755 และ 0.813 สำหรับ brain stem, optic chiasm, mandible, left optic nerve, right optic nerve, left parotid, right parotid, left submandibular และ right submandibular ตามลำดับ ความถูกต้องของการแบ่งส่วน OARs ได้ผลการศึกษาที่ดีจากการใช้ SMR ร่วมกับ FCNN และใช้เวลาสั้นลงสำหรับการแบ่งส่วนภาพหลายอวัยวะ เมื่อเปรียบเทียบกับวิธีการอื่นที่ทันสมัย

Zhu W. และคณะ (2018) (17) พัฒนาโมเดล AnatomyNet ซึ่งเป็นโครงข่ายการเรียนรู้เชิงลึกที่มีรูปแบบการดำเนินการแบบ 3D convolution สำหรับการแบ่งส่วนปริมาตรทางกายวิภาคศาสตร์ศีรษะและลำคออัตโนมัติ ใช้รูปแบบสถาปัตยกรรมโครงข่ายแบบ 3D U-Net และพัฒนาส่วนสำคัญของโมเดล 3 ส่วน ได้แก่ 1.) รูปแบบการดำเนินการส่วนการเข้ารหัสใช้วิธีการเข้ารหัสแบบปริมาตรทั้งชุดภาพ CT แทนการเข้ารหัสแบบสไลซ์ CT, 2.) การรวม 3D squeeze and-excitation residual blocks เข้ากับชั้นดำเนินการส่วนการเข้ารหัส เพื่อเพิ่มการแสดงคุณลักษณะภาพ (feature) ที่ดีขึ้น และ 3.) ปรับปรุงฟังก์ชันการสูญเสีย (loss function) ใหม่ โดยทำการรวม Dice scores และ focal loss ทำให้การสอนโมเดลดีขึ้น ใช้ข้อมูลภาพ CT ศีรษะและลำคอ 216 ชุด เพื่อสอนโมเดล AnatomyNet ให้สามารถแบ่งส่วนภาพอวัยวะที่มีความเสี่ยง brain stem, chiasm, mandible, optic nerve left, optic nerve right, parotid gland left, parotid gland right, submandibular gland left และ submandibular gland right การทดสอบประสิทธิภาพโมเดล ใช้ข้อมูลภาพ CT จากชุดข้อมูล MICCAI 2015 เพื่อคำนวณค่าสัมประสิทธิ์ความเหมือน (DSC) และ ค่าระยะห่าง 95 เปอร์เซนต์ไทล์ฮาวดราฟ (95%HD) ผลการประเมิน ค่า DSC เพิ่มขึ้น 3.3% เมื่อเทียบกับผลจากการแข่งขัน MICCAI 2015 โมเดลที่พัฒนาสามารถประมวลผลได้เร็ว

Men K. และคณะ (2019) (18) นำเสนอการปรับปรุงความแม่นยำและประสิทธิภาพการแบ่งส่วนภาพอวัยวะที่มีความเสี่ยงในงานรังสีรักษาด้วยโครงข่ายใหม่ “convolutional neural networks cascades (CNN cascades)” ซึ่งประกอบด้วยการทำงาน 2 ส่วน ได้แก่ 1.) simple region detector (SRD) เป็นส่วนการทำงานของโมเดลเพื่อใช้หาความสัมพันธ์ระดับต้นสำหรับกำหนดตำแหน่งขอบเขตที่สนใจ (ROI) ของอวัยวะที่มีความเสี่ยง 2.) fine segmentation unit (FSU) ทำงานต่อเนื่องจากส่วนแรกโดยใช้ ROI ขนาดเล็กจากส่วนที่ 1 เป็นข้อมูลนำเข้าและใช้โครงข่ายระดับลึกสำหรับการแบ่งส่วนภาพ ใช้ข้อมูลผู้ป่วยมะเร็งศีรษะและลำคอ 100 ชุด (ภาพ CT 14,651 สไลด์) และข้อมูลการแบ่งส่วนภาพอ้างอิง เป็นกลุ่มประชากรในการศึกษา โมเดลถูกสอนให้แบ่งส่วนภาพอวัยวะที่มีความเสี่ยง brain stem, spinal cord, left eye, right eye, left parotid, right parotid และ mandible ประเมินผลความถูกต้องของโครงข่าย CNN cascades ด้วยค่า DSC และค่า HD ผลค่าเฉลี่ย DSC ของ OARs ทั้งหมดเท่ากับ 0.9 ค่าเฉลี่ย HD เท่ากับ 3 มิลลิเมตร ลดระยะเวลาได้ 48% และ 5% สำหรับโมเดล FSU และ U-Net ตามลำดับ

van der Veen J. และคณะ (2018) (28) ทำการตรวจสอบความถูกต้องของโครงข่ายประสาทเทียม 3 มิติ สำหรับการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงแบบกึ่งอัตโนมัติ ใช้โมเดล CNN เพื่อกำหนดขอบเขตอวัยวะที่มีความเสี่ยงแบบอัตโนมัติ จากนั้นนำผลขอบเขตอวัยวะที่มีความเสี่ยงที่ได้มาทำการปรับแก้โดยแพทย์รังสีรักษา ทำการเปรียบเทียบผลขอบเขตจากการปรับแก้กับขอบเขตจากการวาดด้วยมือ ศึกษาด้วยภาพเอกซเรย์คอมพิวเตอร์จากกลุ่มผู้ป่วยมะเร็งศีรษะและคอรายใหม่

จำนวน 15 ราย ตรวจความถูกต้องของขอบเขตอวัยวะที่มีความเสี่ยง โดยการคำนวณค่า DSC, average symmetric surface distance (ASSD) และหาประสิทธิภาพของเครื่องมือการกำหนดขอบเขตอวัยวะที่มีความเสี่ยง จากการพิจารณาเวลาที่ใช้กำหนดขอบเขตอวัยวะที่มีความเสี่ยงจากวิธีการวาดด้วยมือเปรียบเทียบกับเวลาที่ใช้เพื่อปรับแก้ขอบเขตอวัยวะที่มีความเสี่ยงที่ถูกกำหนดแบบอัตโนมัติด้วย CNN ประเมินความแปรปรวนของการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงของแพทย์ 2 ราย โดยการคำนวณจากค่า DSC และ ASSD จากวิธีการวาดด้วยมือและวิธีการกำหนดขอบเขตอัตโนมัติแล้วถูกปรับแก้โดยแพทย์รังสีรักษา ผลการศึกษาพบว่า ระยะเวลาที่ใช้สำหรับการปรับแก้ขอบเขตที่ถูกกำหนดแบบอัตโนมัติโดยเฉลี่ยน้อยกว่าระยะเวลาจากวิธีการวาดด้วยมือ 33% (23 กับ 34 นาที) ($p < 10^{-6}$) ค่า interobserver variability (IOV) จากวิธีการปรับแก้ขอบเขตที่ถูกกำหนดแบบอัตโนมัติสำหรับ OARs เกือบทั้งหมดดีขึ้นอย่างมีนัยสำคัญ ($p < 0.05$) ผลค่าเฉลี่ย ASSD จาก OARs ทั้งหมดลดลงจาก 1.9 เป็น 1.2 มม. ผลค่าเฉลี่ย DSC ของขอบเขตที่ถูกทำนายแบบอัตโนมัติเปรียบเทียบกับขอบเขตที่ผ่านการปรับแก้ของอวัยวะที่มีความเสี่ยงทั้งหมดเท่ากับ 90% และ 84% สำหรับแพทย์รังสีรักษาท่านที่ 1 และ 2 ตามลำดับ

บทที่ 3 วิธีดำเนินงานวิจัย

ประชากรและกลุ่มตัวอย่าง

การศึกษานี้ใช้ชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษา ซึ่งประกอบด้วยภาพเอกซเรย์คอมพิวเตอร์ และข้อมูลการกำหนดขอบเขต จากฐานข้อมูลสาธารณะออนไลน์ The cancer image achieve ในชุดข้อมูล Head-Neck-PET-CT (31) จาก 4 สถาบันที่แตกต่างกันใน Québec ประเทศแคนาดา ซึ่งมีจำนวนชุดข้อมูลผู้ป่วย 101 ราย และชุดข้อมูล TCGA-HNSC (32) จำนวนชุดข้อมูลผู้ป่วย 1 ราย ภาพมีขนาดเมตริกเท่ากับ 512×512 ขนาดพิกเซลอยู่ในช่วง 0.605 mm² ถึง 1.355 mm² รายละเอียดข้อมูลคุณลักษณะผู้ป่วยทั้ง 102 ราย แสดงดังตาราง 3 ข้อมูลภาพถูกนำมาเตรียมเป็นชุดข้อมูลภาพ CT และ mask อ้างอิง เพื่อใช้สำหรับการสอนการตรวจสอบ และการทดสอบ โมเดลการเรียนรู้เชิงลึกที่พัฒนาขึ้น

ตาราง 3 ข้อมูลคุณลักษณะผู้ป่วยของชุดข้อมูลภาพ 102 ชุดข้อมูล

| characteristics | no. cases |
|-----------------|-----------|
| Tumors site | |
| Hypopharynx | 4 |
| Age (year) | 65 - 83 |
| Sex | no. cases |
| Male | 1 |
| Female | 3 |
| TNM stage | no. cases |
| stage II | 2 |
| stage III | 1 |
| stage IVA | 1 |
| Larynx | 14 |
| Age (year) | 51 - 88 |
| Sex | no. cases |
| Male | 12 |
| Female | 2 |

| characteristics | no. cases |
|------------------------|------------------|
| TNM stage | no. cases |
| stage I | 1 |
| stage III | 8 |
| stage IVA | 5 |
| Nasopharynx | 15 |
| Age (year) | 37 - 83 |
| Sex | no. cases |
| Male | 12 |
| Female | 3 |
| TNM stage | no. cases |
| stage II | 1 |
| stage IIB | 1 |
| stage III | 6 |
| stage IVA | 6 |
| stage IVB | 1 |
| Oropharynx | 67 |
| Age (year) | 41 - 80 |
| Sex | no. cases |
| Male | 49 |
| Female | 18 |
| TNM stage | no. cases |
| stage II | 6 |
| stage III | 6 |
| stage IVA | 49 |
| stage IVB | 6 |
| Unknown primary | 2 |
| Age (year) | 53, N/A |
| Sex | no. cases |
| Male | 1 |
| N/A | 1 |

| characteristics | | no. cases |
|-----------------|-----------|-----------|
| | TNM stage | no. cases |
| | stage IV | 1 |
| | N/A | 1 |
| SUM | | 102 |

หมายเหตุ N/A = not available, no. cases = number of cases

เครื่องมือที่ใช้ในการศึกษา

1. Kaggle virtual machines ซึ่งเป็น Jupyter notebooks สำหรับใช้เขียนโปรแกรม
2. Segmentation model library (44) บนพื้นฐานโครงสร้าง Keras
3. DICOM-contour library (45) สำหรับการแปลงภาพ DICOM และ contour
4. Library ในการทำงานของการศึกษา ประกอบด้วย
 - 4.1 Open CV python 4.1.1
 - 4.2 NumPy 1.16.4
 - 4.3 Pandas 0.25.1
 - 4.4 Scikit-learn 0.21.3
 - 4.5 Keras 2.3.0
 - 4.6. Matplotlib 3.0.3
 - 4.7 SciPy 1.5.4
 - 4.8 Time 0.7
5. ภาษา Python เวอร์ชัน 3.6.6

วิธีดำเนินการศึกษา

การศึกษานี้ทำการพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ กำหนดศึกษาใน 7 อวัยวะที่มีความเสี่ยง ได้แก่ brain stem, left eye globe, right eye globe, mandible, left parotid gland, right parotid gland และ spinal cord การพัฒนาโมเดลดำเนินการบนพื้นฐานรูปแบบสถาปัตยกรรมโครงข่าย U-Net และ FPN ขั้นตอนการศึกษาประกอบด้วย การเตรียมข้อมูล การพัฒนาและการสอนโมเดล การทดสอบและประเมินผลการทำงานของโมเดลที่พัฒนาขึ้น

1. การเตรียมข้อมูล image และ mask

1.1 การเตรียมข้อมูล CT image

เนื่องจากภาพ CT เป็นภาพ DICOM ที่ข้อมูลของแต่ละ pixel อยู่ในรูปแบบ Hounsfield Unit (HU) ซึ่งเป็นค่าข้อมูลที่มีช่วงค่อนข้างกว้าง (-1000, 1000 HU) โดยค่าข้อมูล HU แต่ละอวัยวะที่มีความเสี่ยงที่สนใจของการศึกษานี้มีความแตกต่างกับค่าข้อมูล HU ของเนื้อเยื่อบริเวณโดยรอบน้อย การศึกษานี้จึงได้กำหนดข้อมูลภาพ CT นำเข้าในรูปแบบของ window ที่เหมาะสม เพื่อให้สามารถแยกขอบเขตอวัยวะที่มีความเสี่ยงกับบริเวณโดยรอบได้ชัดเจนขึ้น โดยมีลำดับขั้นตอนดังแสดงในภาพที่ 15 เริ่มการทำงานของโปรแกรมจากการนำเข้าภาพ CT ลักษณะชนิดไฟล์ DICOM ทำการอ่านค่าข้อมูลจากไฟล์นำเข้า จากนั้นแปลงค่าข้อมูลให้อยู่ในรูป pixel array และทำการสร้างภาพใหม่ด้วยการกำหนดขอบเขต window level และ window width ที่สัมพันธ์กับการแยกขอบเขตอวัยวะที่มีความเสี่ยงที่สนใจศึกษา โดยการศึกษานี้เลือกใช้ bone window, brain window และ soft tissue window ขั้นตอนสุดท้ายทำการรวมภาพแต่ละ window เป็น 1 ภาพนำเข้ามีขนาด 512, 512, 3 โดยแต่ละ channel คือภาพ 1 window

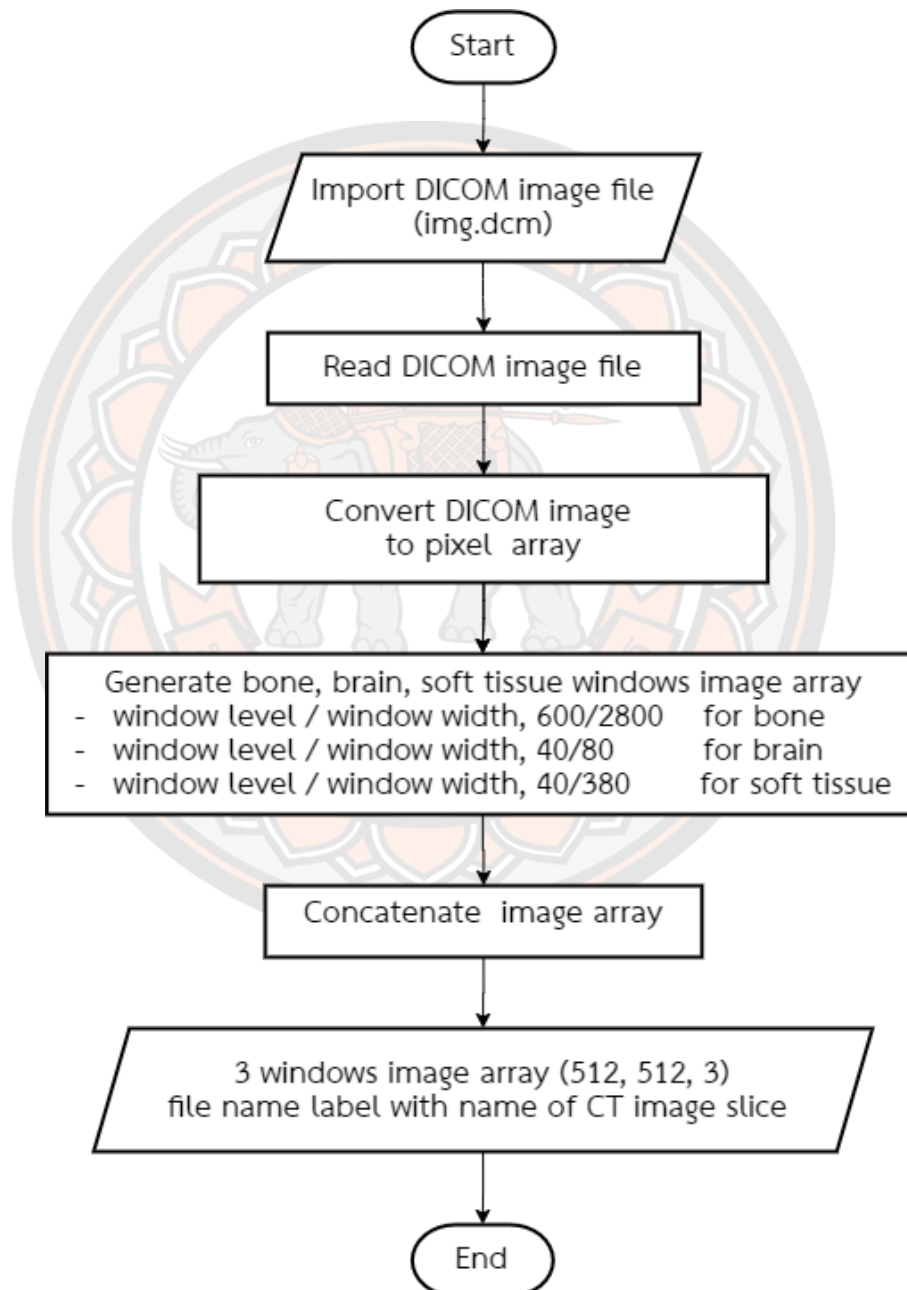
1.2 การเตรียมข้อมูล mask

ข้อมูล mask ถูกเตรียมไว้สำหรับใช้เป็นคำตอบในขั้นตอนการสอนโมเดล และถูกใช้อ้างอิงในขั้นตอนการทดสอบและประเมินผล การเตรียมข้อมูล mask จากไฟล์ข้อมูล DICOM RT structure ซึ่งถูกกำหนดขอบเขตด้วยการวาดมือจากแพทย์รังสีรักษา ถูกแยกข้อมูลโดยใช้ DICOM-contour library ด้วย 2 ขั้นตอน ดังนี้

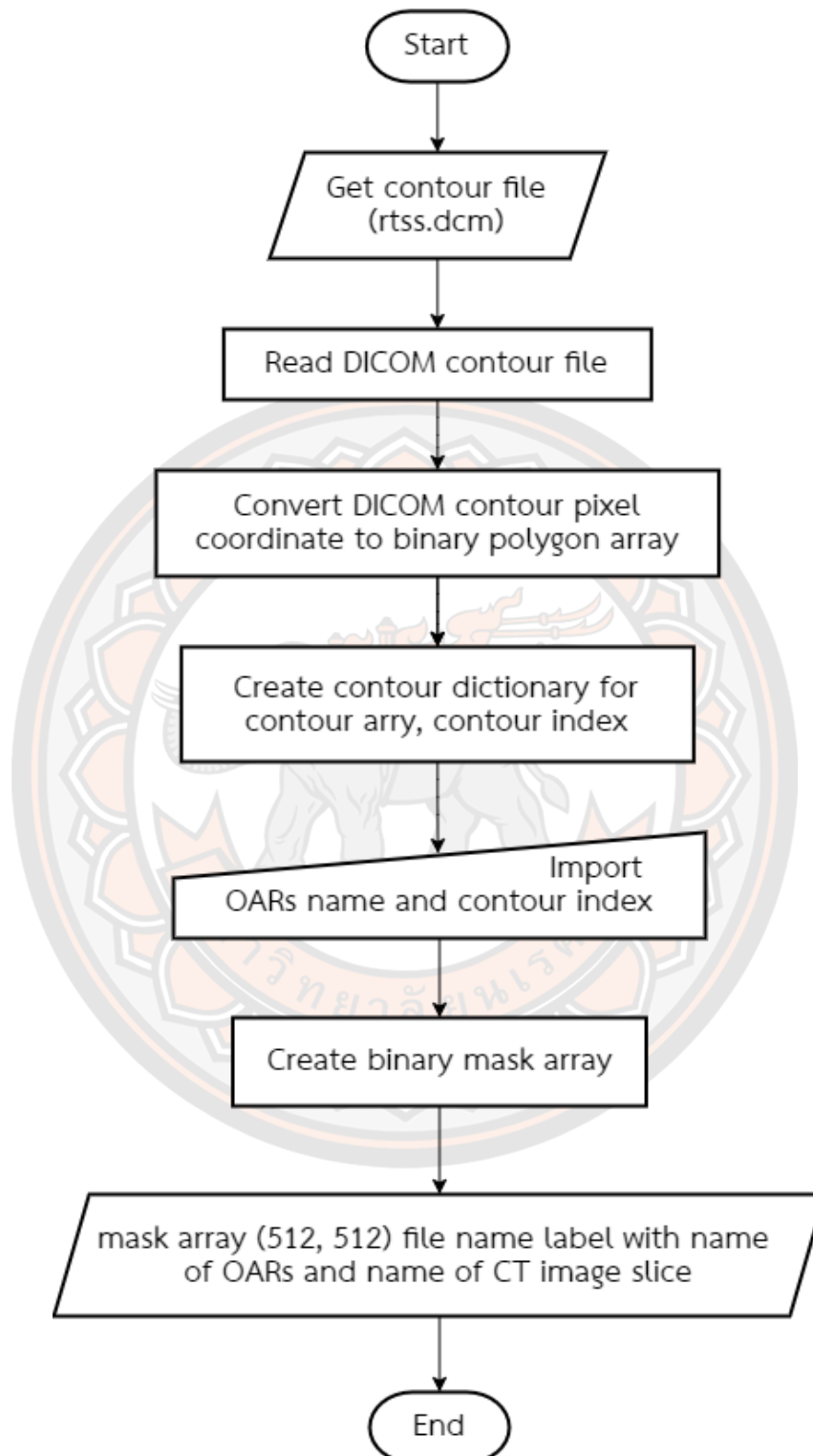
1.2.1 การแปลงข้อมูล DICOM RT structure เป็นข้อมูล binary mask array การทำงานของโปรแกรมแสดงดังภาพที่ 16 เริ่มจากการนำเข้าไฟล์ DICOM RT structure ลำดับต่อมาทำการอ่านข้อมูลการกำหนดขอบเขต (contour pixel coordinate) จากนั้นสร้างภาพ binary mask array ของทุกค่าข้อมูลการกำหนดขอบเขต ดำเนินการเก็บข้อมูลภาพ binary mask array ที่ถูกสร้างขึ้นในรูปแบบ Dictionary กำหนด key ด้วยค่า contour index เพื่อเข้าถึงข้อมูลที่สร้างขึ้น จากนั้นสร้าง binary mask array ของอวัยวะที่มีความเสี่ยงที่สนใจศึกษาโดยการป้อนค่า contour index ของอวัยวะที่สนใจลงในโปรแกรม โปรแกรมจะประมวลผลสร้างไฟล์ภาพ binary mask array ขึ้นโดยชื่อไฟล์จะถูก label ด้วยชื่อของอวัยวะที่มีความเสี่ยง และชื่อภาพ CT สไลซ์ที่สัมพันธ์กับขอบเขตอวัยวะนั้น

1.2.2 การเตรียมไฟล์ mask นำเข้า แสดงดังภาพที่ 17 เริ่มจากการนำเข้าไฟล์ binary mask array จากข้อ 1.2.1 และชื่อภาพ CT ขั้นตอนต่อมาโปรแกรมประมวลผลจัดกลุ่มไฟล์ binary mask array ของอวัยวะที่มีความเสี่ยงที่สนใจศึกษาทั้ง 7 อวัยวะ จากชื่อไฟล์

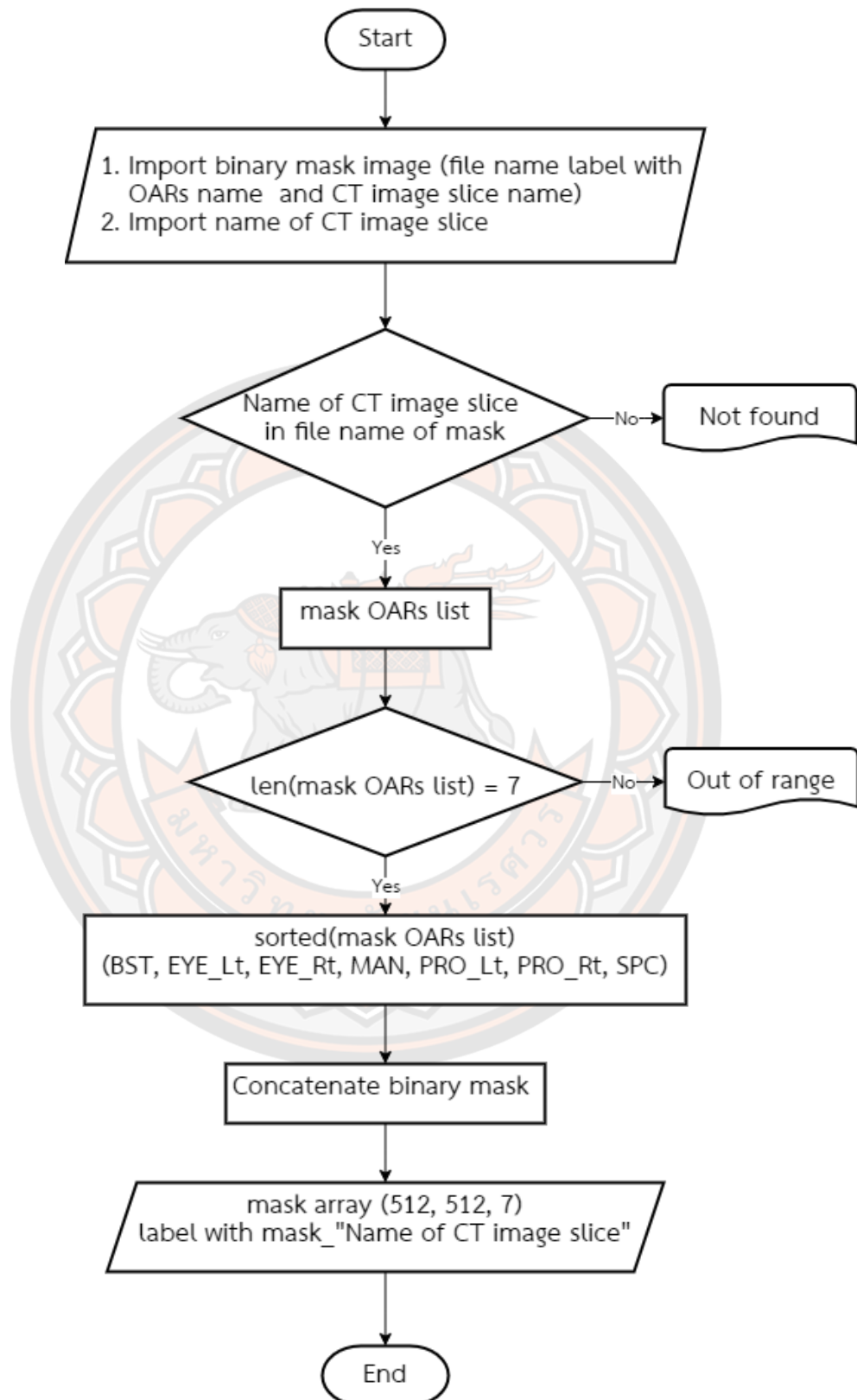
ซึ่งสัมพันธ์กับชื่อภาพ CT เดียวกัน จากนั้นโปรแกรมประมวลผลตรวจสอบจำนวนไฟล์ binary mask array ในแต่ละกลุ่ม หากมีจำนวนครบ 7 อวัยวะไฟล์ที่ถูกประมวลผลนี้จะถูกนำไปจัดเรียงลำดับคลาสและรวมเป็นไฟล์ mask array นำเข้า มีขนาด 512, 512, 7 โดยกำหนดให้แต่ละ channel เป็นภาพ binary mask array ของแต่ละอวัยวะที่มีความเสี่ยง



ภาพ 15 แผนภาพการเตรียมข้อมูล CT image



ภาพ 16 แผนภาพการแปลงข้อมูล DICOM RT structure เป็นข้อมูล binary mask array



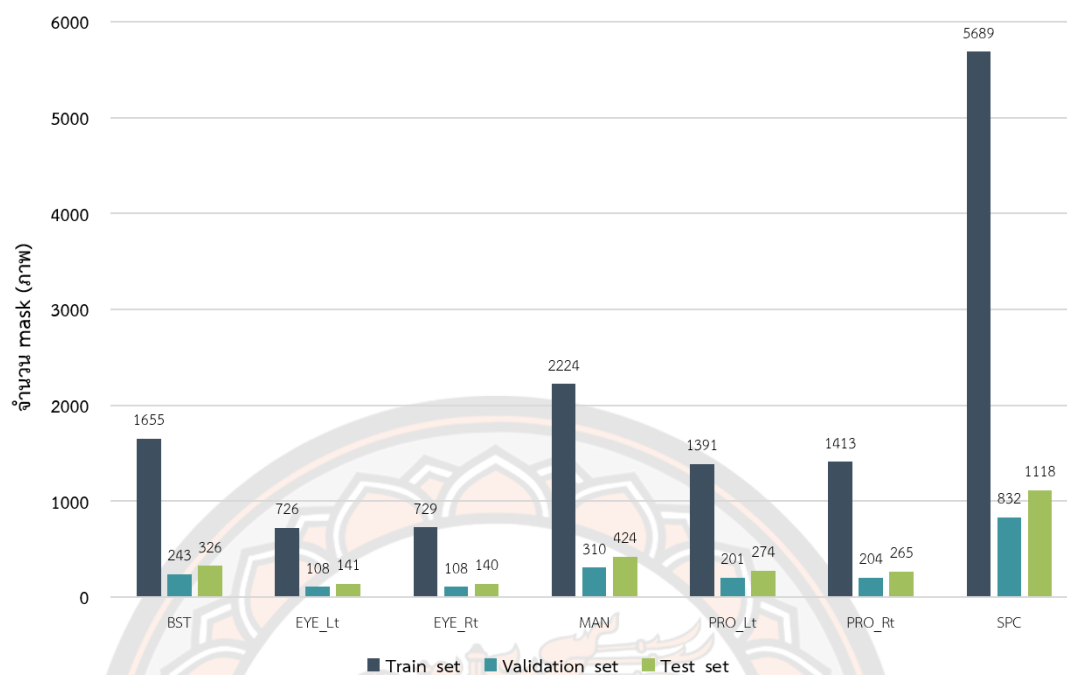
ภาพ 17 แผนภาพการเตรียมไฟล์ mask นำเข้า

1.3 การแบ่งชุดข้อมูล

ข้อมูลที่เตรียมจากข้อ 1.1 และ 1.2 จะถูกแบ่งเป็น 3 ชุดข้อมูล ได้แก่ 1. ชุดข้อมูลสำหรับการสอนโมเดล จำนวน 75 ราย (ภาพ CT และภาพ mask จำนวนอย่างละ 7683 ภาพ) 2. ชุดข้อมูลสำหรับการตรวจสอบโมเดล 12 ราย (ภาพ CT และภาพ mask จำนวนอย่างละ 1136 ภาพ) และ 3. ชุดข้อมูลสำหรับการทดสอบโมเดล 15 ราย (ภาพ CT และภาพ mask อย่างละ 1504 ภาพ) การกระจายของชุดข้อมูล แสดงดังตารางที่ 4 และข้อมูลแสดงการกระจายความถี่ mask ของข้อมูลแต่ละอวัยวะที่มีความเสี่ยง แสดงดังภาพที่ 18

ตาราง 4 การกระจายชุดข้อมูลสอน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ

| Tumors site | No. case | Train set | Validation set | Test set |
|-------------|----------|-----------|----------------|----------|
| Hypopharynx | 4 | 2 | 1 | 1 |
| Larynx | 14 | 10 | 2 | 2 |
| Nasopharynx | 15 | 10 | 2 | 3 |
| Oropharynx | 67 | 52 | 7 | 9 |
| Unknown | 21 | 2 | - | - |
| SUM | 102 | 75 | 12 | 15 |



ภาพ 18 การกระจายความถี่ mask อวัยวะที่มีความเสี่ยงของชุดข้อมูลสอน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ

2. การพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยง

การพัฒนาโมเดลการเรียนรู้เชิงลึกของการศึกษานี้ พัฒนาระบบพื้นฐานสถาปัตยกรรมโครงข่าย U-Net และ FPN โดยใช้ Segmentation model library เพื่อสร้างสถาปัตยกรรมโครงข่ายร่วมกับการเลือกใช้โมเดล backbone ชนิด VGG19, ResNet50, SeResNet50 และ InceptionV3 สำหรับทั้งสองรูปแบบสถาปัตยกรรมโครงข่าย ดังนี้

2.1 สถาปัตยกรรมโครงข่าย U-Net

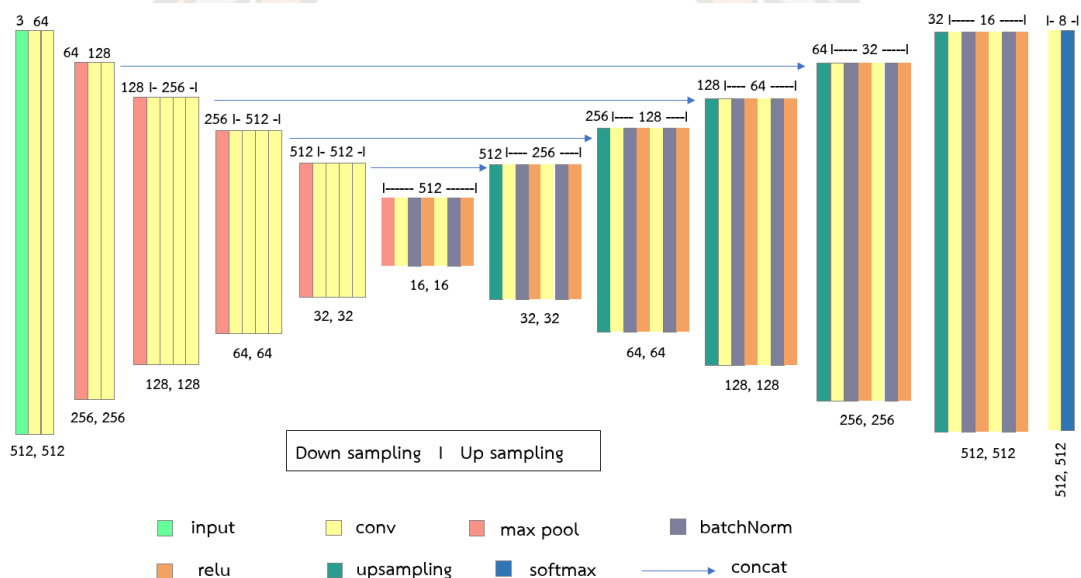
สถาปัตยกรรมโครงข่าย U-Net ประกอบด้วย 2 ส่วน ได้แก่ ส่วน down sampling และส่วน up sampling แสดงดังภาพที่ 19 ถึง 22 มีการปรับรูปแบบส่วนการเข้ารหัส (Encoder path) ในส่วน down sampling ด้วยโมเดล backbone ดังนี้

2.1.1 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ VGG19

ส่วน down sampling ใช้โมเดล backbone แบบ VGG19 โดยผลลัพธ์จากชั้น max pooling สุดท้ายของโมเดล backbone แบบ VGG19 ถูกเชื่อมต่อกับชั้น convolution—batch normalization—activation (Relu) จำนวน 2 ชุด แล้วเชื่อม

เข้ากับชั้น up sampling ของ block ที่ 1 ส่วน up sampling เรียกการเชื่อมต่อของชั้นดำเนินการ convolution ส่วนดังกล่าวนี้ว่า center block

ส่วน up sampling ประกอบด้วย 5 block ที่สัมพันธ์กับส่วน down sampling โดยแต่ละ block ประกอบด้วยชั้น up sampling ในชั้นแรก ตามด้วยชุดดำเนินการ convolution 2 ชุด (เช่นเดียวกับส่วน center block) สำหรับแต่ละ block และมีการเชื่อมต่อด้วยชั้น convolution และชั้น activation (soft max) เพิ่มใน 2 ชั้นสุดท้ายของ block 5 การเชื่อมกันของส่วน down sampling และส่วน up sampling ใช้วิธี concatenation เพื่อเชื่อมต่อ feature map ระหว่าง block ที่ 2 ถึง 5 ส่วน down sampling กับชั้น up sampling ของ block ที่ 4 ถึง 1 ส่วน up sampling ตามลำดับการเชื่อมต่อแสดงดังภาพ 19



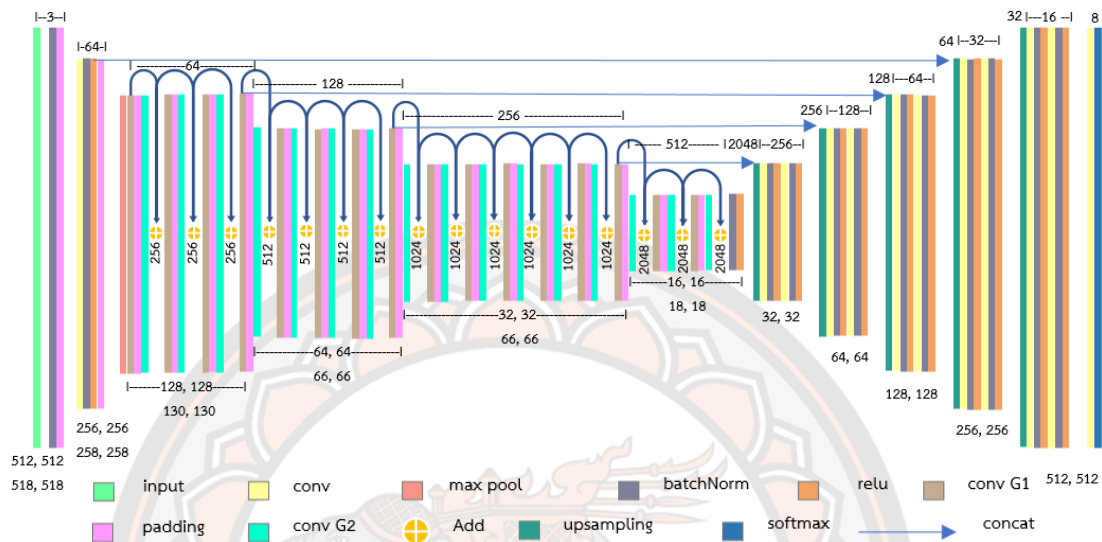
ภาพ 19 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ VGG19

2.1.2 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ ResNet50

ส่วน down sampling ใช้โมเดล backbone แบบ ResNet50 โดยผลลัพธ์สุดท้ายจากการรวม (add) ของ block 6 ถูกเชื่อมต่อเข้ากับชั้น batch normalization—activation ก่อนเชื่อมต่อเข้ากับส่วน up sampling

ส่วน up sampling ประกอบด้วย 5 block ที่มีลักษณะการเชื่อมต่อของชั้นต่าง ๆ เช่นเดียวกับส่วน up sampling ของสถาปัตยกรรม U-Net โมเดล backbone แบบ VGG19 โดยลักษณะการเชื่อมของส่วน down sampling และส่วน up sampling ใช้วิธี

concatenation เพื่อเชื่อมต่อ feature map ระหว่างชั้น activation แรกของ block ที่ 2 4 5 และ 6 ส่วน down sampling กับชั้น up sampling ของ block ที่ 4 ถึง 1 ส่วน up sampling ตามลำดับ แสดงดังภาพ 20

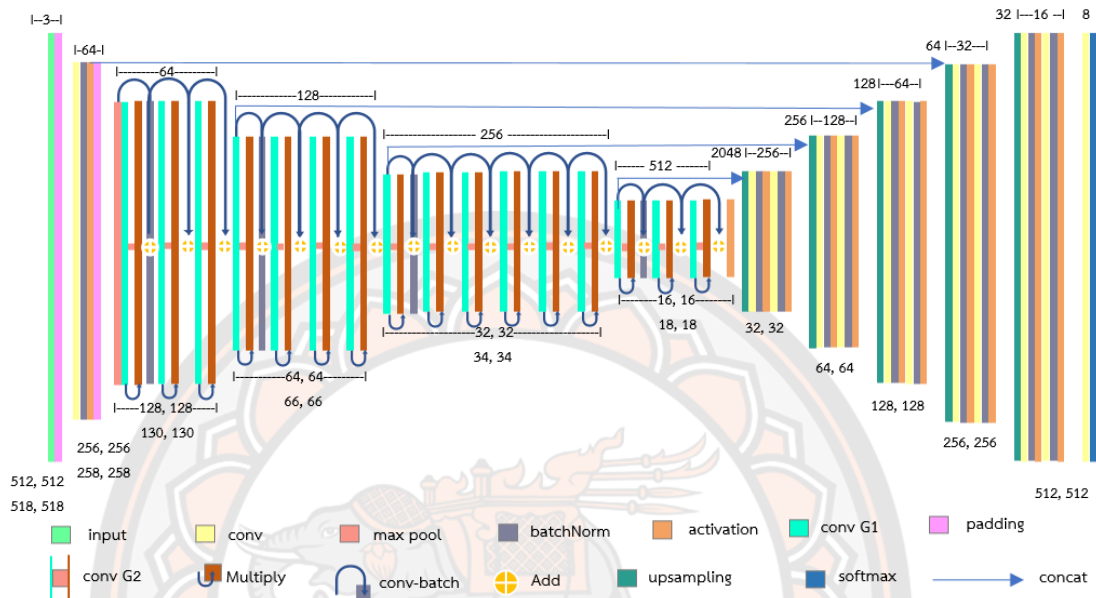


ภาพ 20 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ ResNet50

2.1.3 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ SE-ResNet50

ส่วน down sampling ใช้โมเดล backbone แบบ SE-ResNet50 โดยผลลัพธ์สุดท้ายจากการรวม (add) ของ block 6 ถูกเชื่อมต่อเข้ากับชั้น activation ก่อนเชื่อมต่อเข้ากับส่วน up sampling

ส่วน up sampling ประกอบด้วย 5 block ที่มีลักษณะการเชื่อมต่อของชั้นต่าง ๆ เช่นเดียวกับส่วน up sampling ของสถาปัตยกรรม U-Net โมเดล backbone VGG19 โดยลักษณะการเชื่อมของส่วน down sampling และส่วน up sampling ใช้วิธี concatenation เพื่อเชื่อมต่อ feature map ระหว่างชั้น activation แรกของ block ที่ 2 4 5 และ 6 ส่วน down sampling กับชั้น up sampling ของ block ที่ 4 ถึง 1 ส่วน up sampling ตามลำดับ แสดงดังภาพ 21

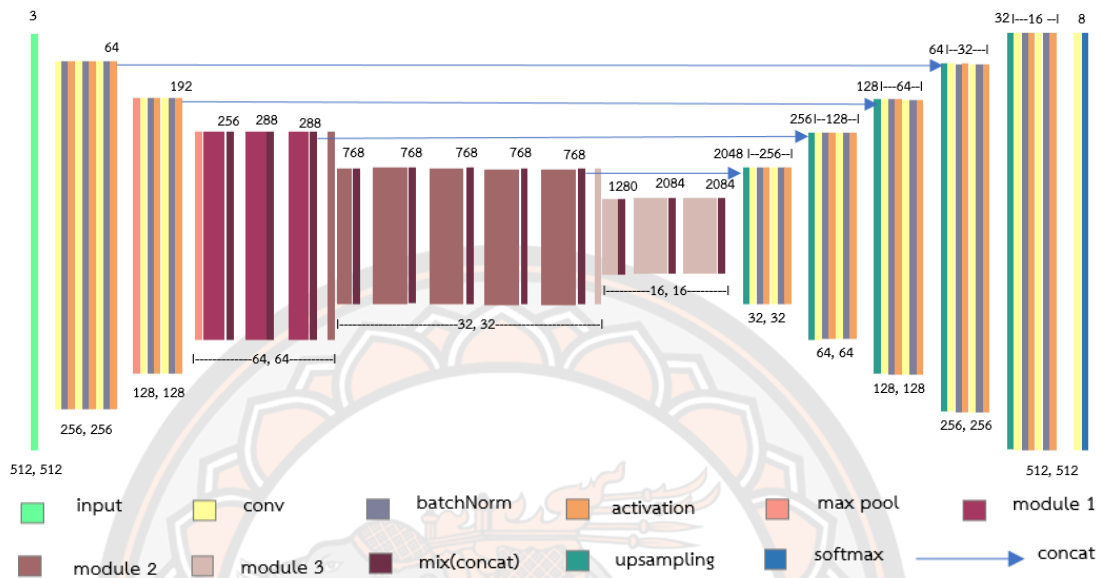


ภาพ 21 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ SE-ResNet50

2.1.4 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ InceptionV3

ส่วน down sampling ใช้โมเดล backbone แบบ InceptionV3 โดยผลลัพธ์สุดท้ายจากการ mix(concat) ของ block 6 ถูกเชื่อมต่อเข้ากับส่วน up sampling

ส่วน up sampling ประกอบด้วย 5 block ที่มีลักษณะการเชื่อมต่อของชั้นต่าง ๆ เช่นเดียวกับส่วน up sampling ของสถาปัตยกรรม U-Net โมเดล backbone แบบ VGG19 โดยลักษณะการเชื่อมของส่วน down sampling และส่วน up sampling ใช้วิธี concatenation เพื่อเชื่อมต่อ feature map ระหว่างชั้น activation ท้ายสุดของ block 2 และ 3 และชั้น concatenate ท้ายสุดของ block 4 และ 5 ส่วน down sampling กับชั้น up sampling ของ block ที่ 4 ถึง 1 ส่วน up sampling ตามลำดับ แสดงการเชื่อมต่อ ดังภาพ 22



ภาพ 22 สถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ InceptionV3

2.2 สถาปัตยกรรมโครงข่าย FPN

สถาปัตยกรรมโครงข่าย FPN ประกอบด้วย การเชื่อมต่อของขั้นตอนการ 3 แนว ได้แก่ bottom-up (BU), top-down (TD) และแนว lateral I, II (L-I, L-II) มีรายละเอียดดังนี้

2.2.1 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ VGG19

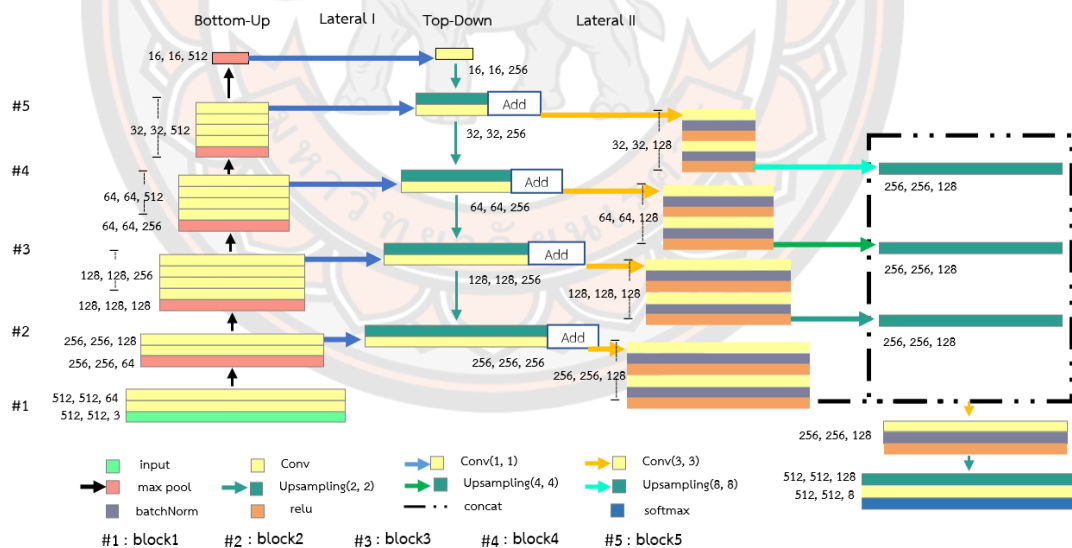
การเชื่อมต่อของแนว BU ใช้โมเดล backbone แบบ VGG19 เช่นเดียวกับส่วน down sampling ของสถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ VGG19

การเชื่อมต่อในแนว L-I ใช้ผลลัพธ์จากชั้นคอนโวลูชันสุดท้ายของ block 2 BU ถึง block 5 BU และผลลัพธ์จากชั้น max pooling ของ block 5 BU ถูกนำมาคอนโวลูชันด้วย kernel ขนาด (1, 1) แสดงเป็นผลลัพธ์ของ block 2 L-I ถึง block 5 L-I สำหรับการคอนโวลูชัน block 2 BU ถึง block 5 BU ตามลำดับ และผลลัพธ์จากการคอนโวลูชันชั้น max pooling ใน block 5 BU ถูกใช้เป็นข้อมูลเริ่มต้นสำหรับการเชื่อมต่อในแนว TD

การเชื่อมต่อในแนว TD ใช้ผลลัพธ์จากการเชื่อมต่อในแนว L-I ของชั้น max pooling ใน block 5 BU เป็นข้อมูลเริ่มต้นเพื่อทำการ up sampling ผลลัพธ์จากการ up sampling ของชั้นนี้ถูกรวมกับผลลัพธ์ block 5 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 5 TD จากนั้นผลลัพธ์นี้จะถูก up sampling และรวมกับผลของ block

4 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 4 TD จากนั้นใช้รูปแบบการเชื่อมต่อลักษณะเดียวกันนี้กับ block 3 L-I กับผล up sampling ของ block 4 TD แสดงเป็นผลลัพธ์ของ block 3 TD และ block 2 L-I กับผล up sampling ของ block 3 TD แสดงเป็นผลลัพธ์ของ block 2 TD ตามลำดับ

การเชื่อมต่อในแนว L-II ใช้ผลลัพธ์ของ block 2 TD ถึง block 5 TD ถูกนำมาดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) สำหรับแต่ละ block TD ละ 2 ชุด ผลลัพธ์สุดท้ายที่ได้แสดงเป็นผลของ block 2 L-II ถึง block 5 L-II โดยผลลัพธ์ block 3 L-II ถึง block 5 L-II ถูก up sampling ด้วย kernel ขนาด (2, 2), (4, 4), (8, 8) ตามลำดับ ผลลัพธ์จากการ up sampling ถูกเชื่อมผสานเข้ากับผลลัพธ์ block 2 L-II ด้วยวิธีการ concatenation ผลลัพธ์จากการ concatenation ถูกนำมาดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) จากนั้นทำการ up sampling ตามด้วยการคอนโวลูชัน และใช้ชั้น activation (soft max) เพื่อทำนายผลลัพธ์ การเชื่อมต่อของสถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ VGG19 แสดงดังภาพที่ 23



ภาพ 23 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ VGG19

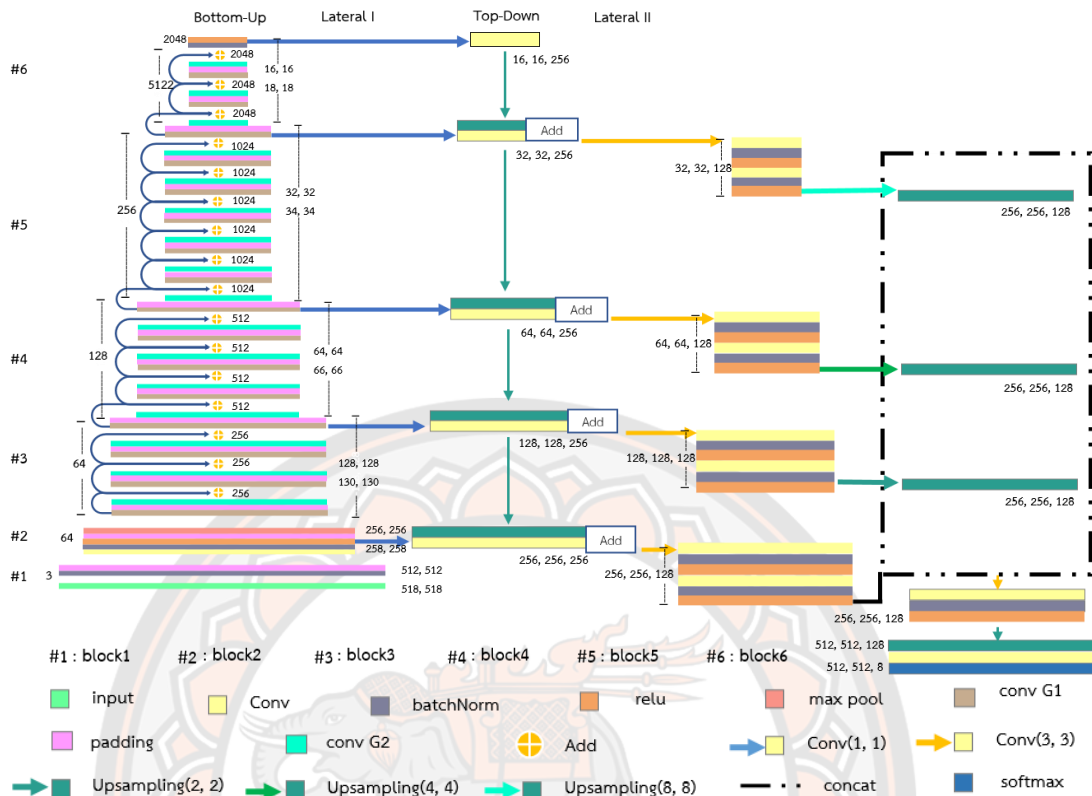
2.2.2 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ Res-Net50

การเชื่อมต่อของแนว BU ใช้โมเดล backbone แบบ Res-Net50 เช่นเดียวกับส่วน down sampling ของสถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ Res-Net50

การเชื่อมต่อในแนว L-I ใช้ผลลัพธ์จากชั้น activation แรกของ block 2 BU, 4 BU, 5 BU และ block 6 BU และผลลัพธ์จากชั้น activation สุดท้ายของ block 6 BU ถูกนำมาคอนโวลูชันด้วย kernel ขนาด (1, 1) แสดงเป็นผลลัพธ์ของ block 2 L-I ถึง block 6 L-I สำหรับการคอนโวลูชัน block 2 BU, 4 BU, 5 BU และ block 6 BU ตามลำดับ และผลลัพธ์จากการคอนโวลูชันชั้น activation สุดท้ายใน block 6 BU ถูกใช้เป็นข้อมูลเริ่มต้นสำหรับการเชื่อมต่อในแนว TD

การเชื่อมต่อในแนว TD ใช้ผลลัพธ์จากการเชื่อมต่อในแนว L-I ของชั้น activation สุดท้าย ใน block 6 BU เป็นข้อมูลเริ่มต้นเพื่อทำการ up sampling ผลลัพธ์จากการ up sampling ในชั้นนี้ถูกรวมกับผลลัพธ์ block 6 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 6 TD จากนั้นผลลัพธ์นี้จะถูก up sampling และรวมกับผลของ block 5 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 5 TD จากนั้นใช้รูปแบบการเชื่อมต่อลักษณะเดียวกันนี้กับ block 4 L-I กับผล up sampling ของ block 5 TD แสดงเป็นผลลัพธ์ของ block 4 TD และ block 2 L-I กับผล up sampling ของ block 4 TD แสดงเป็นผลลัพธ์ของ block 2 TD ตามลำดับ

การเชื่อมต่อในแนว L-II ใช้ผลลัพธ์ของ block 2 TD, 4 TD, 5 TD และ block 6 TD ถูกดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) สำหรับแต่ละ block TD ละ 2 ชุด ผลลัพธ์สุดท้ายที่ได้แสดงเป็นผลของ block 2 L-II, 4 L-II, 5 L-II และ block 6 L-II โดยผลลัพธ์ block 4 L-II ถึง block 6 L-II ถูก up sampling ด้วย kernel ขนาด (2, 2), (4, 4), (8, 8) ตามลำดับ ผลลัพธ์จากการ up sampling ถูกเชื่อมผสานกับผลลัพธ์ block 2 L-II ด้วยวิธีการ concatenation ผลลัพธ์จากการ concatenation ถูกนำมาดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) จากนั้นทำการ up sampling ตามด้วยการคอนโวลูชัน และใช้ชั้น activation (soft max) เพื่อทำนายผลลัพธ์ การเชื่อมต่อของสถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ Res-Net50 แสดงดังภาพที่ 24



ภาพ 24 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ ResNet50

2.2.3 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ SE-ResNet50

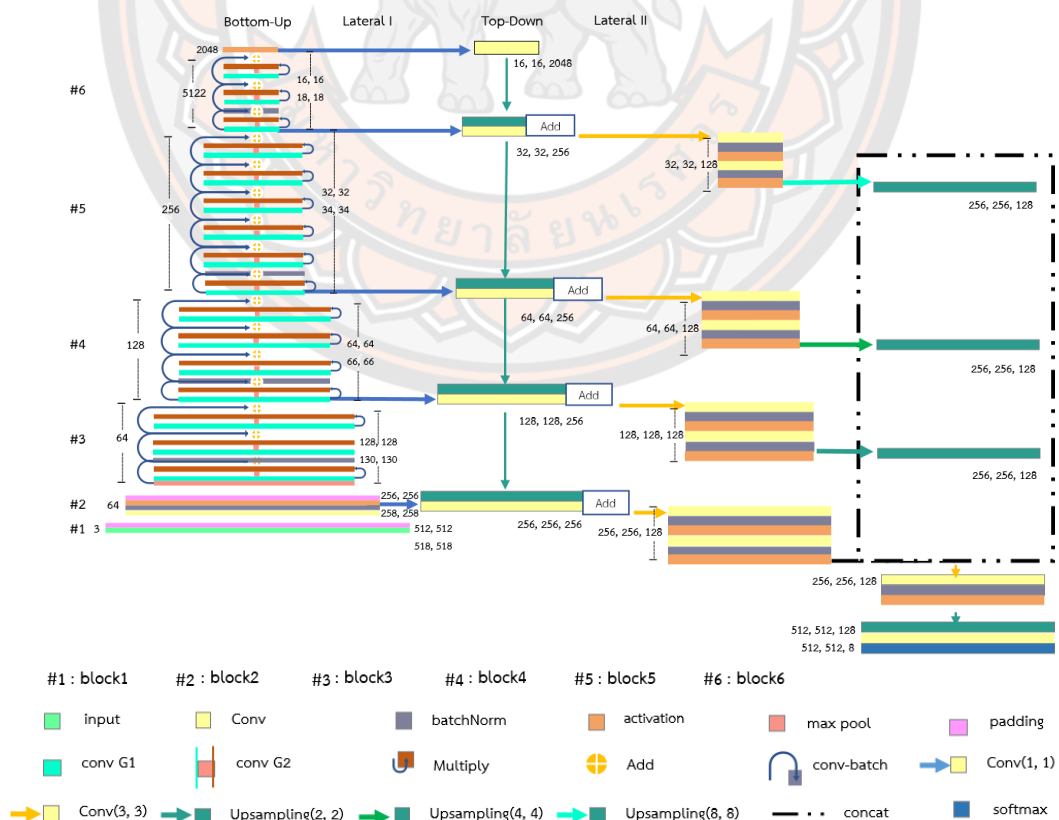
การเชื่อมต่อของแนว BU ใช้โมเดล backbone แบบ SE-ResNet50 เช่นเดียวกับ ส่วน down sampling ของสถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ SE-ResNet50

การเชื่อมต่อในแนว L-I ใช้ผลลัพธ์จากชั้น activation แรกของ block 2 BU, 4 BU, 5 BU และ block 6 BU และผลลัพธ์จากชั้น activation สุดท้ายของ block 6 BU ถูกนำมาคอนโวลูชันด้วย kernel ขนาด (1, 1) แสดงเป็นผลลัพธ์ของ block 2 L-I ถึง block 6 L-I สำหรับการคอนโวลูชัน block 2 BU, 4 BU, 5 BU และ block 6 BU ตามลำดับ และผลลัพธ์จากการคอนโวลูชันชั้น activation สุดท้ายใน block 6 BU ถูกใช้เป็นข้อมูลเริ่มต้นสำหรับการเชื่อมต่อในแนว TD

การเชื่อมต่อในแนว TD ใช้ผลลัพธ์จากการเชื่อมต่อในแนว L-I ของชั้น activation สุดท้าย ใน block 6 BU เป็นข้อมูลเริ่มต้นเพื่อทำการ up sampling ผลลัพธ์จากการ up sampling ของชั้นนี้ถูกรวมกับผลลัพธ์ block 6 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 6 TD จากนั้นผลลัพธ์นี้จะถูก up sampling และรวมกับผลของ block

5 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 5 TD จากนั้นใช้รูปแบบการเชื่อมต่อลักษณะเดียวกันนี้กับ block 4 L-I กับผล up sampling ของ block 5 TD แสดงเป็นผลลัพธ์ของ block 4 TD และ block 2 L-I กับผล up sampling ของ block 4 TD แสดงเป็นผลลัพธ์ของ block 2 TD ตามลำดับ

การเชื่อมต่อในแนว L-II ใช้ผลลัพธ์ของ block 2 TD, 4 TD, 5 TD และ block 6 TD ถูกดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) สำหรับแต่ละ block TD ละ 2 ชุด ผลลัพธ์สุดท้ายที่ได้แสดงเป็นผลของ block 2 L-II, 4 L-II, 5 L-II และ block 6 L-II โดยผลลัพธ์ block 4 L-II ถึง block 6 L-II ถูก up sampling ด้วย kernel ขนาด (2, 2), (4, 4), (8, 8) ตามลำดับ ผลลัพธ์จากการ up sampling ถูกเชื่อมผสานเข้ากับผลลัพธ์ของ block 2 L-II ด้วยวิธีการ concatenation ผลลัพธ์จากการ concatenation ถูกนำมาดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) จากนั้นทำการ up sampling ตามด้วยการคอนโวลูชัน และใช้ชั้น activation (soft max) เพื่อทำนายผลลัพธ์ การเชื่อมต่อของสถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ SE-ResNet50 แสดงดังภาพที่ 25



ภาพ 25 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ SE-ResNet50

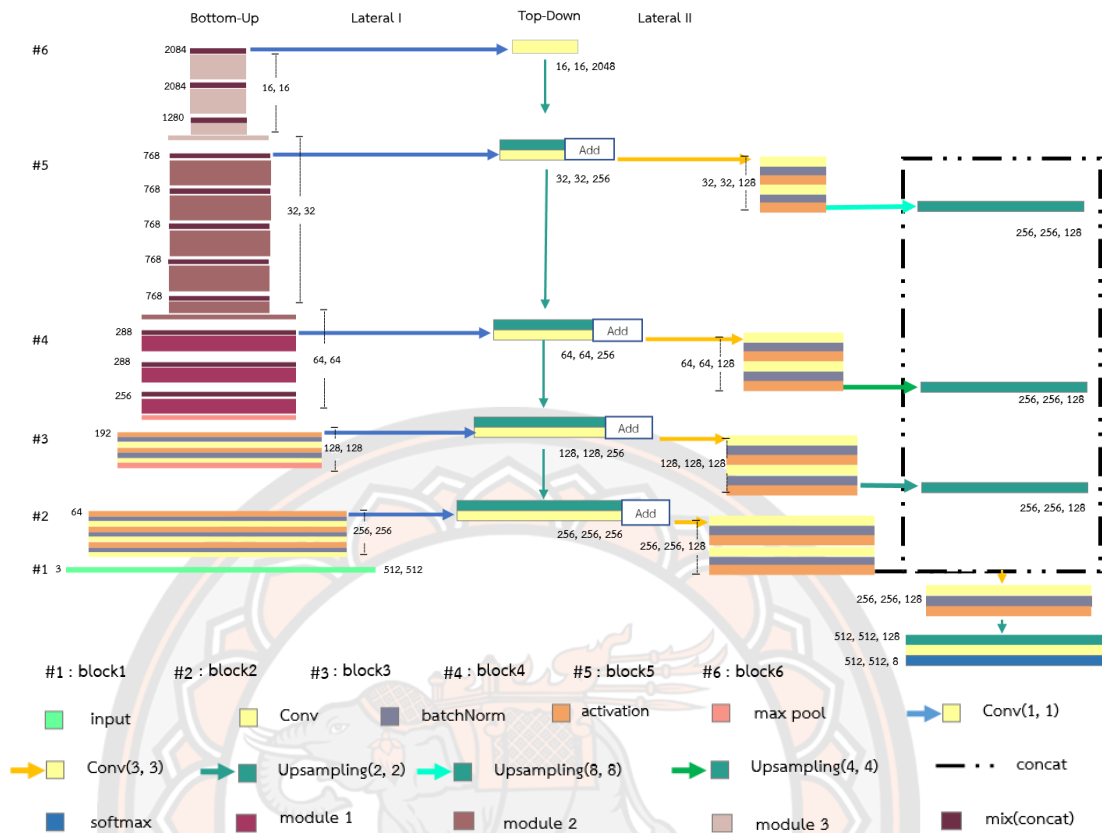
2.2.4 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ InceptionV3

การเชื่อมต่อของแนว BU ใช้โมเดล backbone แบบ InceptionV3 เช่นเดียวกับ ส่วน down sampling ของสถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ InceptionV3

การเชื่อมต่อในแนว L-I ใช้ผลลัพธ์จากชั้น activation สุดท้ายของ block 2 BU และ block 3 BU และชั้น mix concatenate ของ block 4 BU ถึง block 6 BU ถูกนำมา คอนโวลูชันด้วย kernel ขนาด (1, 1) แสดงเป็นผลลัพธ์ของ block 2 L-I ถึง block 6 L-I สำหรับการคอนโวลูชัน block 2 BU ถึง block 6 BU ตามลำดับ โดยผลลัพธ์ของ block 6 L-I ถูกใช้เป็นข้อมูลเริ่มต้นสำหรับการเชื่อมต่อในแนว TD

การเชื่อมต่อในแนว TD ใช้ผลลัพธ์จากการเชื่อมต่อในแนว L-I ของ block 6 L-I เป็นข้อมูลเริ่มต้นเพื่อทำการ up sampling ผลลัพธ์จากการ up sampling ของชั้นนี้ถูกรวม กับผลลัพธ์ block 5 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 5 TD จากนั้นผลลัพธ์นี้จะถูก up sampling และรวมกับผลลัพธ์ของ block 4 L-I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 4 TD จากนั้นใช้รูปแบบการเชื่อมต่อลักษณะ เดียวกันนี้กับ block 3 L-I กับผล up sampling ของ block 4 TD แสดงเป็นผลลัพธ์ของ block 3 TD และ block 2 L-I กับผล up sampling ของ block 3 TD แสดงเป็นผลลัพธ์ ของ block 2 TD ตามลำดับ

การเชื่อมต่อในแนว L-II ใช้ผลลัพธ์ของ block 2 TD ถึง block 5 TD ถูกดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) สำหรับแต่ละ block TD ละ 2 ชุด ผลลัพธ์สุดท้ายที่ได้แสดงเป็นผลของ block 2 L-II, 3 L-II, 4 L-II และ block 5 L-II โดยผลลัพธ์ block 3 L-II ถึง block 5 L-II ถูก up sampling ด้วย kernel ขนาด (2, 2), (4, 4), (8, 8) ตามลำดับ ผลลัพธ์จากการ up sampling ถูกเชื่อมผสานเข้ากับผลลัพธ์ block 2 L-II ด้วยวิธีการ concatenation ผลลัพธ์จากการ concatenation ถูกนำมาดำเนินการด้วยชั้น convolution — batch normalization — activation (Relu) จากนั้นทำการ up sampling ตามด้วยการ คอนโวลูชัน และใช้ชั้น activation (soft max) เพื่อทำนายผลลัพธ์ การเชื่อมต่อของ สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ InceptionV3 แสดงดังภาพที่ 26



ภาพ 26 สถาปัตยกรรมโครงข่าย FPN โมเดล backbone แบบ InceptionV3

2.3 การสอนโมเดลการเรียนรู้เชิงลึก

การสอนโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบริเวณศีรษะและลำคอ แบ่งวิธีการสอนออกเป็น 2 วิธี สำหรับแต่ละรูปแบบสถาปัตยกรรมโครงข่าย วิธีที่ 1 สอนโมเดลด้วยวิธีปรับจูน (fine tune) เริ่มต้นการเรียนรู้ด้วยค่าถ่วงน้ำหนัก ImageNet โดยค่าพารามิเตอร์ของโมเดลส่วนที่สัมพันธ์กับโมเดล backbone ซึ่งมีค่า weight ที่ถูกสอนด้วยชุดข้อมูล ImageNet 2012 ILSVRC (46) ถูกใช้เป็นค่าพารามิเตอร์เริ่มต้นในขั้นตอนการสอนโมเดล แล้วถูกปรับจูนค่าพารามิเตอร์ด้วยชุดข้อมูลสอนของนักศึกษา และวิธีที่ 2 สอนโมเดลด้วยวิธีถ่ายทอดการเรียนรู้ (transfer learning) ค่าพารามิเตอร์ของโมเดลที่สัมพันธ์กับส่วนโมเดล backbone ถูกตรึง (freeze) เป็นพารามิเตอร์ของโมเดลในส่วนการเข้ารหัส (encoder) ได้โมเดลสำหรับการศึกษานี้จำนวน 16 โมเดล ค่าพารามิเตอร์ของแต่ละโมเดลแสดงดังตารางที่ 5

ตาราง 5 จำนวนพารามิเตอร์ของโมเดลการเรียนรู้เชิงลึก

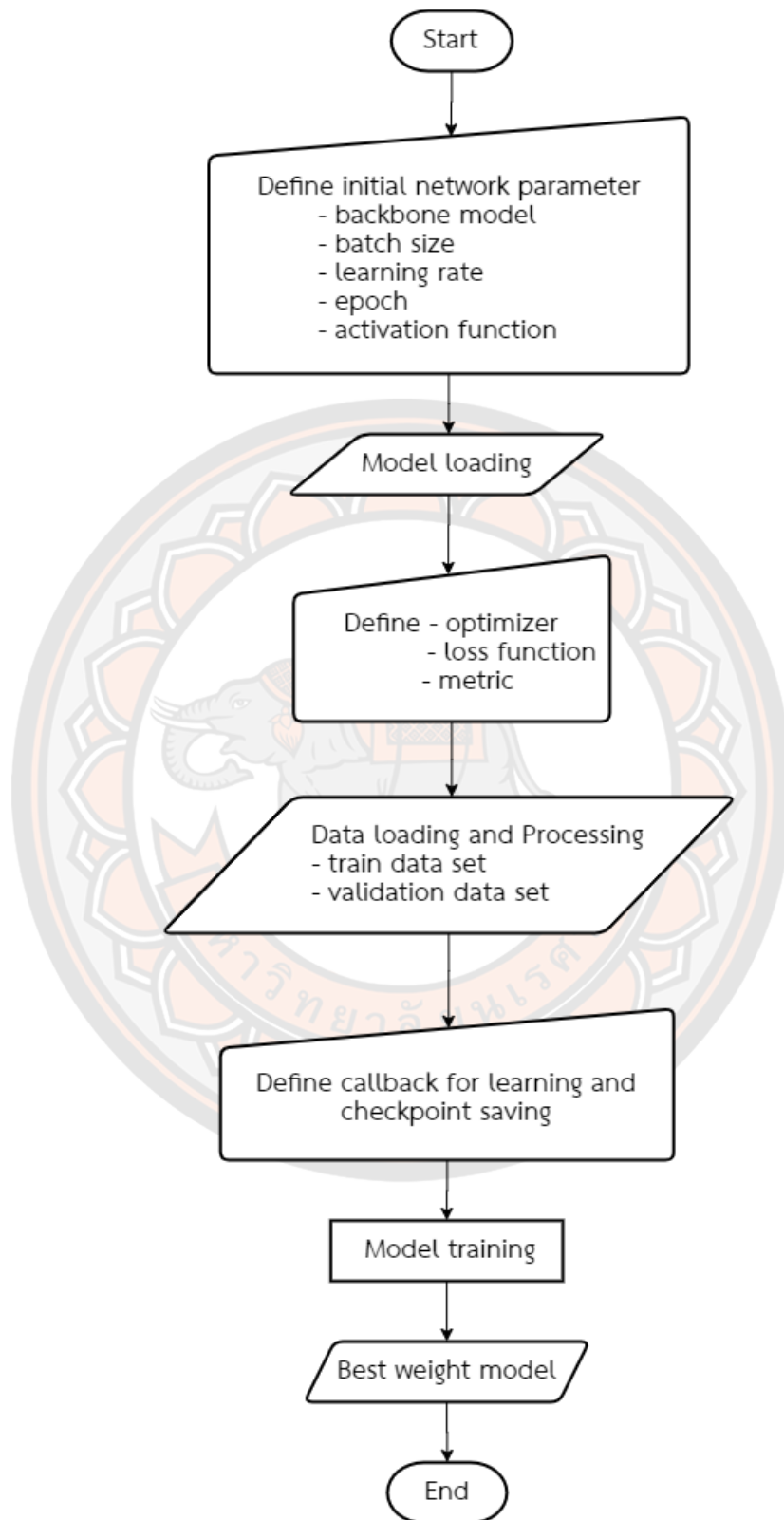
| Models | Total parameters | Trainable parameters | Non-trainable parameters |
|---------------|---------------------|-------------------------|-----------------------------|
| FUVGG19 | 29,062,984 | 29,058,952 | 4,032 |
| TUVGG19 | 29,062,984 | 9,034,568 | 20,028,416 |
| FFVGG19 | 22,890,312 | 22,888,008 | 2,304 |
| TFVGG19 | 22,890,312 | 2,863,624 | 20,026,688 |
| FUResNet50 | 32,562,129 | 32,514,571 | 47,558 |
| TUResNet50 | 32,562,129 | 9,059,659 | 23,502,470 |
| FFResNet50 | 26,919,889 | 26,872,011 | 47,878 |
| TFResNet50 | 26,919,889 | 3,417,099 | 23,502,790 |
| FUSE-ResNet50 | 35,108,216 | 35,053,112 | 55,104 |
| TUSE-ResNet50 | 35,108,216 | 9,067,208 | 26,041,008 |
| FFSE-ResNet50 | 29,465,976 | 29,410,552 | 55,424 |
| TFSE-ResNet50 | 29,465,976 | 3,424,648 | 26,041,328 |
| FUIncV3 | 29,934,120 | 29,897,704 | 36,416 |
| TUIncV3 | 29,934,120 | 8,146,568 | 21,787,552 |
| FFIncV3 | 25,037,352 | 25,000,616 | 36,736 |
| TFIncV3 | 25,037,352 | 3,249,480 | 21,787,872 |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3

ขั้นตอนการสอนโมเดล เริ่มต้นการทำงานของโปรแกรมผ่านการกำหนดค่าพารามิเตอร์ เริ่มต้นโครงข่ายการเรียนรู้เชิงลึก โดยกำหนดค่าพารามิเตอร์ดังนี้

- 1.) ชนิดของโมเดล backbone (VGG19, Res-Net50, SE-ResNet50, InceptionV3)
- 2.) กำหนด batch size (ขั้นตอนการสอน (train) batch size = 4 ขั้นตอนการตรวจสอบ (validation) batch size = 1)
- 3.) อัตราการเรียนรู้ (0.0001)
- 4.) กำหนดจำนวน epoch ที่ใช้ในการสอนโมเดล 20 epoch
- 5.) กำหนดฟังก์ชันกระตุ้น (Relu, softmax activation function)

เมื่อกำหนดค่าพารามิเตอร์เริ่มต้นโครงข่ายแล้วเสร็จ ทำการเรียกใช้โมเดลตามรูปแบบสถาปัตยกรรมที่กำหนด จากนั้นทำการกำหนด optimizer ฟังก์ชันการสูญเสีย (loss function) และ metric (สำหรับการศึกษานี้เลือกใช้ optimizer = Adam optimizer, ฟังก์ชันการสูญเสีย = Dice Loss + Categorical Focal Loss, Metrics = IOU score, F1 score, Precision score, Recall score) หลังจากกำหนดค่าแล้วเสร็จ ทำการเรียกข้อมูลนำเข้าเพื่อประมวลผลข้อมูล ในขั้นตอนนี้มีการเพิ่มคลาสให้ข้อมูล Mask จากเดิมมีขนาด 512, 512, 7 เพิ่มคลาส background (background = 1 - mask.sum(axis=-1, keepdims=True)) ในชุดข้อมูลสอน และชุดข้อมูลตรวจสอบ ทำให้ข้อมูลนำเข้ามีขนาด 512, 512, 8 จากนั้นทำการกำหนดค่าเรียกกลับสำหรับขั้นตอนการสอน กำหนดการบันทึกค่าถ่วงน้ำหนักจากการสอน และกำหนดการหยุดสอนเมื่อค่า validation loss ไม่มีเปลี่ยนแปลงต่อเนื่องกัน 3 ครั้ง เริ่มทำการสอนโมเดล ระหว่างกระบวนการสอนมีการจับเวลาสำหรับแต่ละ epoch และบันทึกเวลา บันทึกค่าถ่วงน้ำหนักที่ดี ภาพขั้นตอนการทำงานของโปรแกรมการสอนโมเดล แสดงดังภาพที่ 27



ภาพ 27 แผนภาพการทำงานของโปรแกรมในขั้นตอนการสอนโมเดล

3. การทดสอบและประเมินผล

3.1 การทดสอบการทำงานของโมเดล

การทดสอบโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงที่พัฒนาขึ้นจากการศึกษานี้ ถูกทดสอบด้วยชุดข้อมูลทดสอบที่ถูกเตรียมไว้ในข้อ 1.3 ขั้นตอนการทำงานของโปรแกรมทำนายข้อมูล mask สำหรับภาพ CT นำเข้าแต่ละภาพ แสดงดังภาพที่ 28 มีลำดับการดำเนินการดังนี้ เริ่มการทดสอบด้วยการนำเข้าโมเดลที่ต้องการทดสอบ จากนั้นเรียกการทำงานของโมเดล นำเข้าชุดข้อมูลทดสอบและนำเข้าค่าถ่วงน้ำหนักที่ดีสำหรับแต่ละโมเดล ซึ่งได้จากขั้นตอนการสอนโมเดล เมื่อโมเดลพร้อมใช้งานทำการป้อนภาพ CT ที่ต้องการทำนาย mask เข้าสู่โมเดล โมเดลจะประมวลผลทำนาย mask และบันทึก mask ในขั้นตอนการประมวลผลทำนาย Mask ทำการจับเวลาการประมวลผลของโมเดลด้วยคำสั่งการทำงานจาก Time library ซึ่งใช้หลักการบันทึกเวลาเริ่มการทำงานของโมเดล และเวลาสิ้นสุดการบันทึกผลการทำนาย จากนั้นคำนวณหาผลต่างของระยะเวลา ผลลัพธ์การทำนายของโมเดลได้ข้อมูล mask ขนาด 512, 512, 8 ที่ถูก label ด้วยชื่อไฟล์ภาพ CT แต่ละสไลซ์

3.2 การประเมินผลการทดสอบ

ดำเนินการประเมินผลความเหมือนของภาพ mask array ซึ่งทำนายจากโมเดลการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบนภาพ CT ศีรษะและลำคอที่พัฒนาขึ้นกับ mask array อ้างอิงซึ่งถูกกำหนดขอบเขตอวัยวะที่มีความเสี่ยงด้วยการวาดมือในการวางแผนการรักษาด้วยภาพลักษณะ 2 มิติ (2D CT) บริเวณศีรษะและลำคอ โดยใช้ชุดข้อมูลทดสอบตามที่กำหนดไว้ในข้อ 1.3

การเปรียบเทียบความเหมือนเชิงรูปร่างถูกคำนวณค่า Dice Similarly Coefficient (DSC) (47) แสดงการคำนวณดังสมการที่ 16 และการเปรียบเทียบเพื่อหาความสัมพันธ์ของ mask array ทั้งสองในรูปแบบระยะห่างของขอบ mask ถูกคำนวณด้วยค่า 95 Percentile Hausdorff Distance (95%HD) (47) แสดงการคำนวณดังสมการที่ 17 และ 18

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (16)$$

เมื่อ $|A|$ และ $|B|$ เป็นพื้นที่ของ mask ที่ได้จากการทำนาย และ mask อ้างอิง ตามลำดับ และ $|A \cap B|$ เป็นพื้นที่ที่ซ้อนทับกันของ mask จากการทำนาย และ mask อ้างอิง โดยค่า DSC เท่ากับ 1 หาก mask ทั้ง 2 ชุด ซ้อนทับกันอย่างสมบูรณ์ และเท่ากับ 0 หาก mask ทั้ง 2 ชุด ไม่ซ้อนทับกัน

$$\vec{d}_{H, r}(x, y) = K_r \left\{ \min_{y \in Y} d(x, y) \right\}, \forall x \in X \quad (17)$$

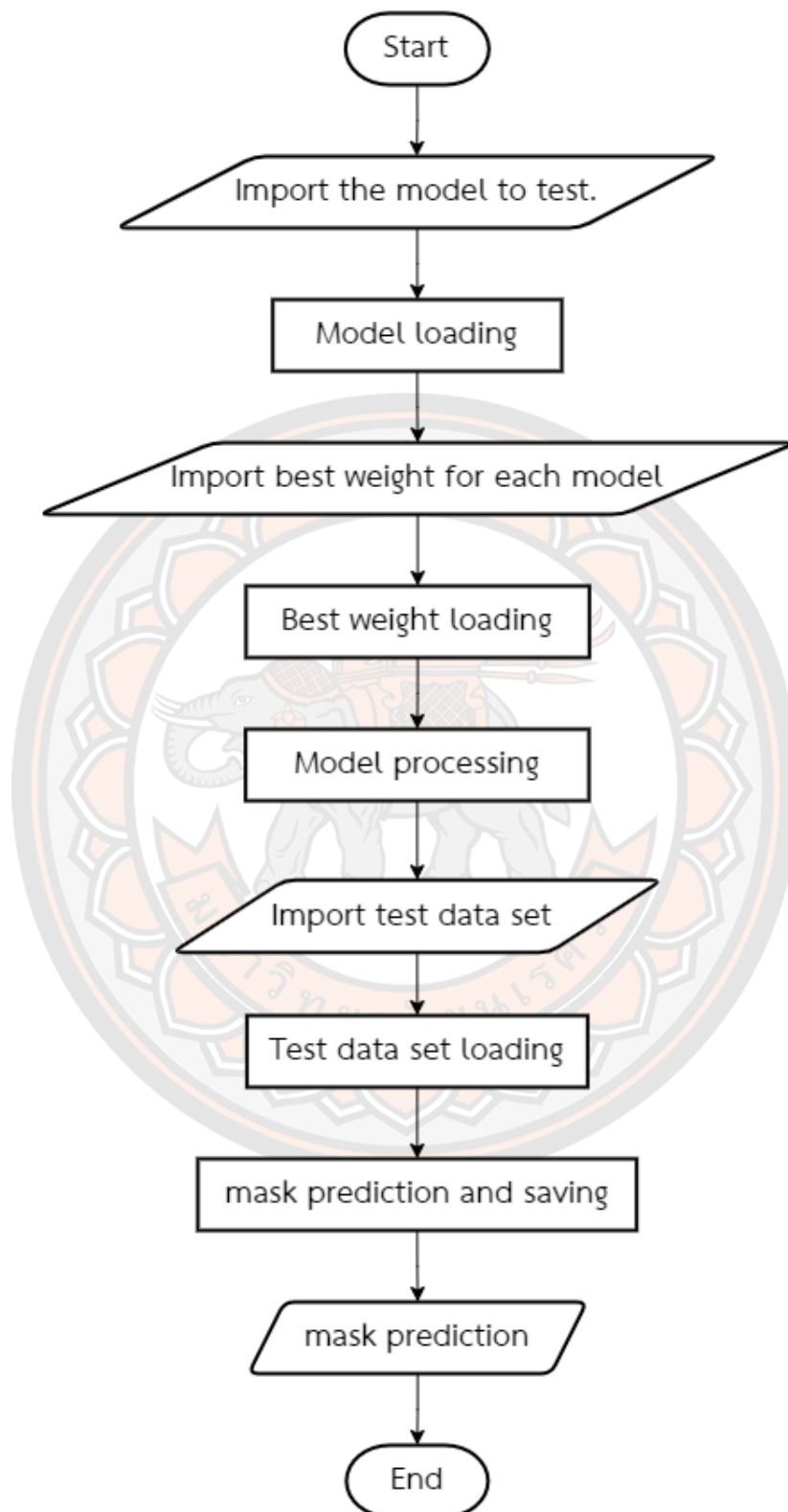
$$d_{H, r}(x, y) = \frac{\vec{d}_{H, r}(x, y) + \vec{d}_{H, r}(y, x)}{2} \quad (18)$$

เมื่อ K_r คือ ค่าเปอร์เซ็นต์ไทล์ที่ใช้คำนวณ (95 Percentile) และ $d(x, y)$ คือ ระยะทางระหว่าง จุด x บนเส้นแสดงรูปร่างขอบเขต Mask จากการทำนาย และจุด y บนเส้นแสดงรูปร่างขอบเขต mask อ้างอิง

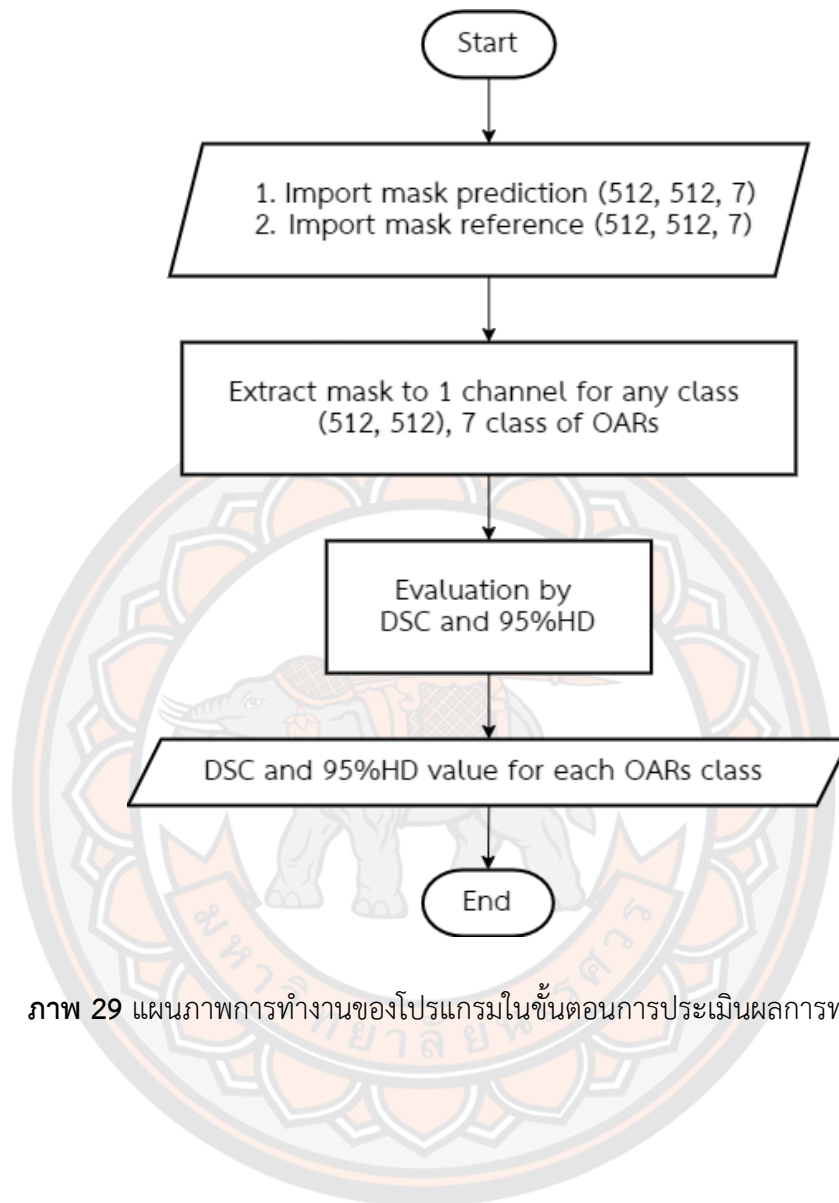
การประเมินผลเริ่มจากการนำเข้าสู่ข้อมูล mask array prediction และ mask array reference จากนั้นทำการแยกไฟล์ภาพให้มีขนาด 512, 512 สำหรับแต่ละคลาส สำหรับข้อมูล mask array prediction จะเลือกใช้ 7 คลาส (ยกเว้น background class) เท่านั้นในขั้นตอนการประเมินผล ข้อมูล mask ในชุดทำนายจะถูกจับคู่กับข้อมูล mask อ้างอิงในคลาสที่ตรงกันสำหรับแต่ละสไลซ์ภาพ CT เดียวกัน เพื่อประเมินหาค่า DSC และค่า 95%HD ตามสูตรการคำนวณที่แสดงข้างต้น ผลลัพธ์จากโปรแกรมจะได้ข้อมูลค่า DSC และค่า 95%HD ของแต่ละคลาสในแต่ละสไลซ์ภาพ CT ของชุดข้อมูลทดสอบ แผนภาพการทำงานของโปรแกรมแสดงดังภาพที่ 29

3.3 การประเมินเวลา

การประเมินเวลาของการศึกษานี้ใช้ข้อมูลเวลาที่บันทึกจากขั้นตอนที่ 2.2 การสอนโมเดล และข้อมูลเวลาการทำนาย mask array ของแต่ละโมเดลที่ได้จากขั้นตอนที่ 3.1 การทดสอบการทำงานของโมเดล ถูกลนำมาวิเคราะห์เปรียบเทียบระยะเวลาการทำงานของแต่ละโมเดลกับโมเดลอื่นทั้งหมด



ภาพ 28 แผนภาพการทำงานของโปรแกรมในขั้นตอนการทำนาย mask

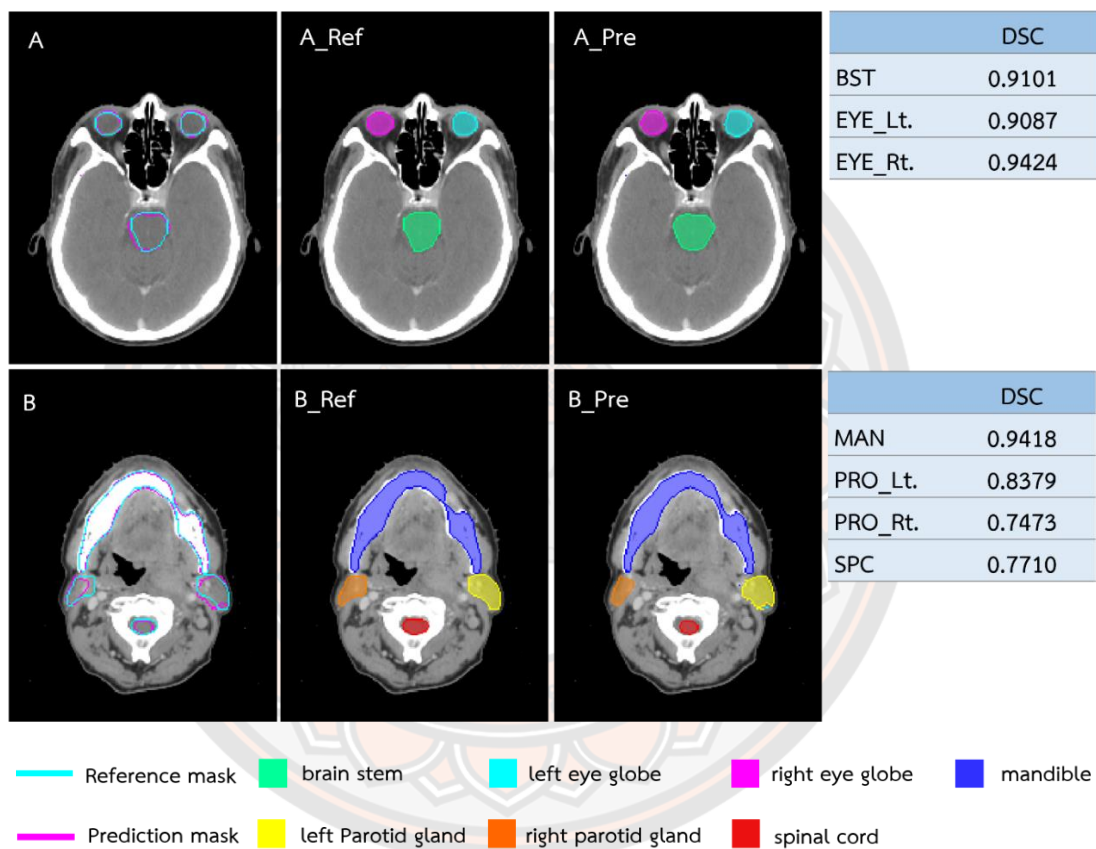


ภาพ 29 แผนภาพการทำงานของโปรแกรมในขั้นตอนการประเมินผลการทดสอบ

บทที่ 4 ผลการศึกษา

ผลการทำนาย mask

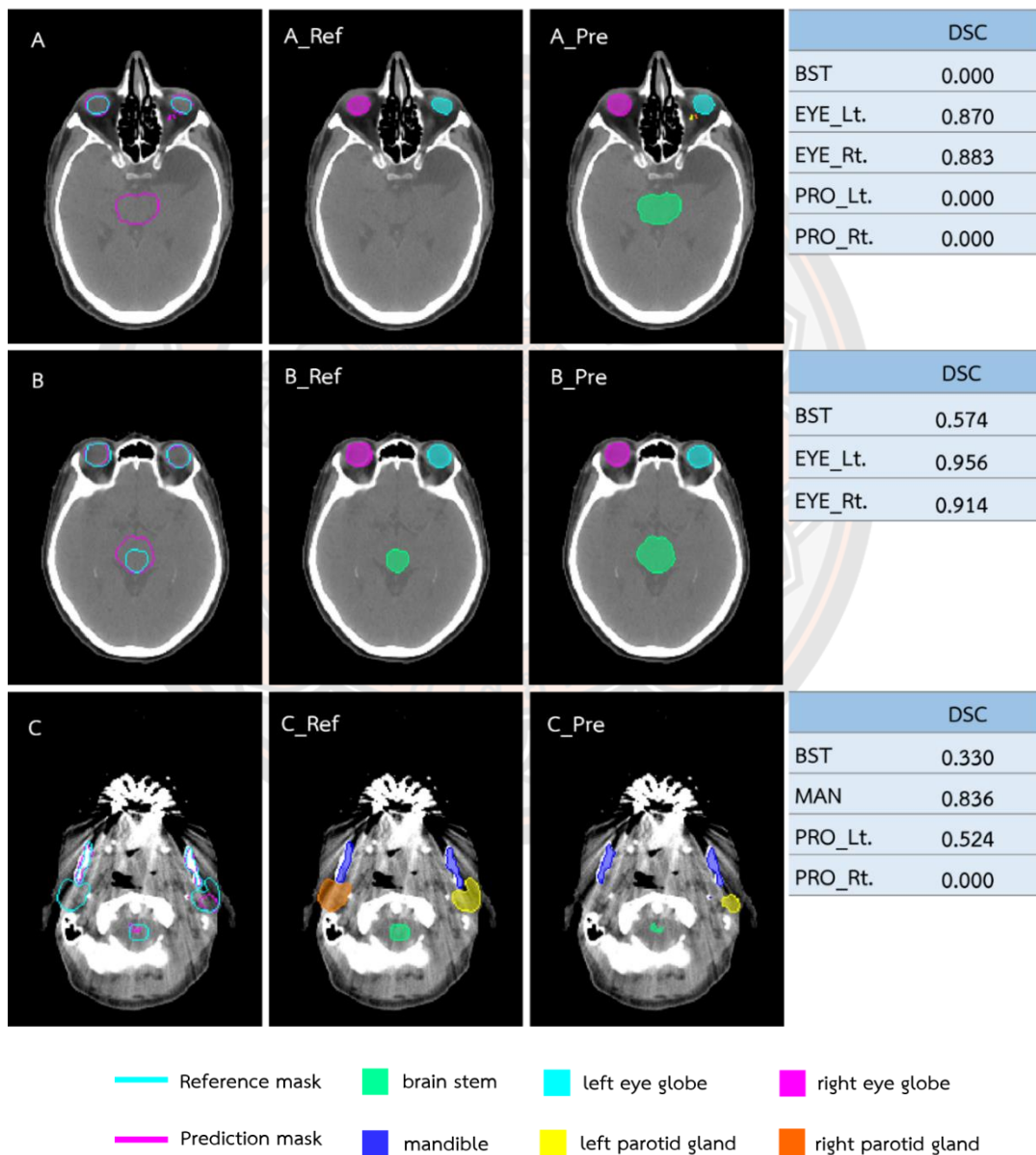
ผลการทำนาย mask จากโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงของมะเร็งศีรษะและลำคอที่พัฒนาขึ้น แสดงตัวอย่างผลของการทำนายดังภาพที่ 30



ภาพ 30 ผลการทำนายขอบเขตของ 7 อวัยวะที่มีความเสี่ยงจากโมเดล TUVGG19

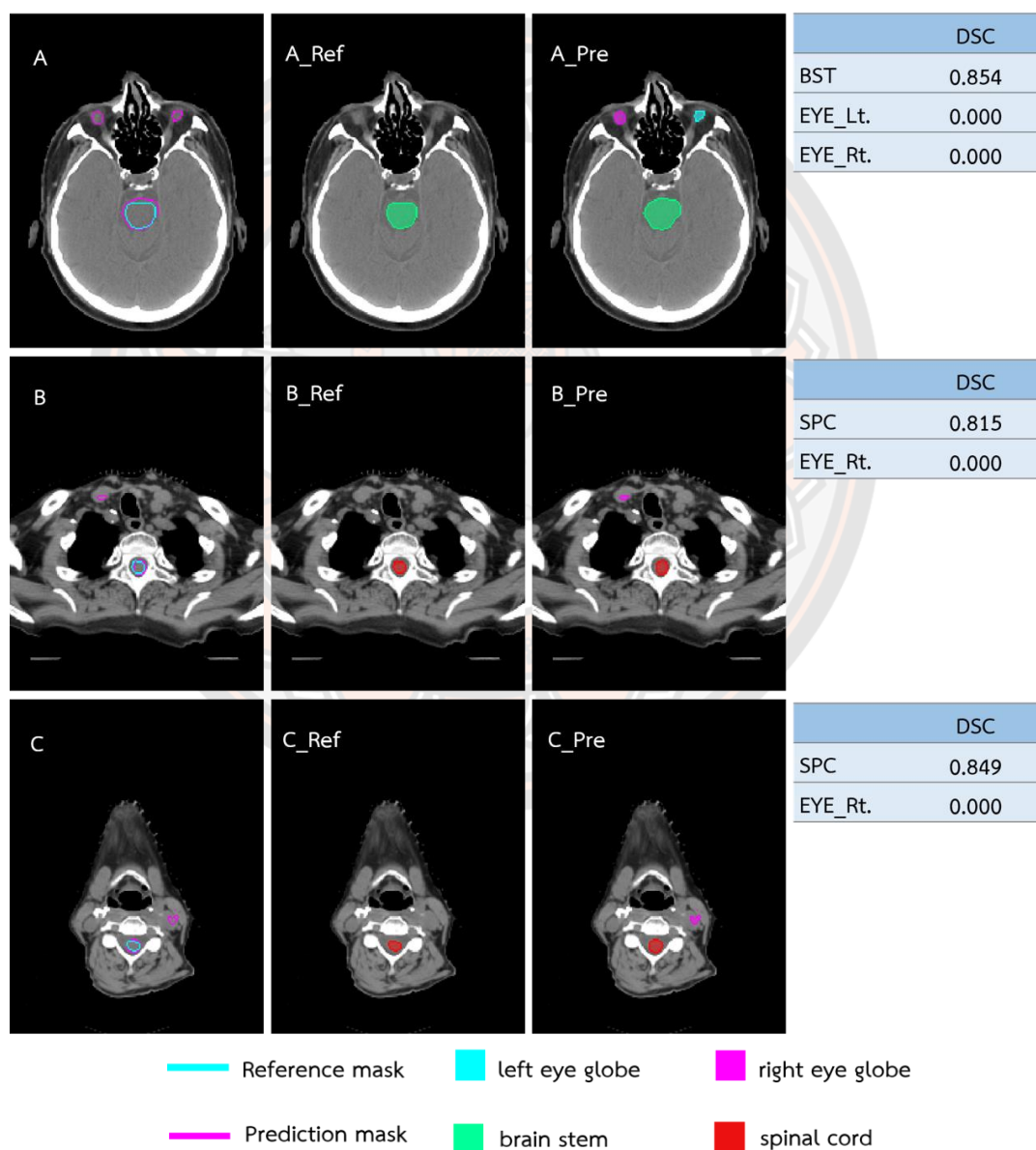
เมื่อเปรียบเทียบ mask อ้างอิงในภาพ 30A_Ref, 30B_Ref กับ mask ทำนายในภาพ 30A_Pre, 30B_Pre โมเดลสามารถทำนาย mask ในแต่ละอวัยวะที่มีความเสี่ยงได้ดี และให้ผลคล้ายกันนี้กับภาพ CT ชุดข้อมูลทดสอบส่วนมาก แต่ก็มีกรณีการทำนายผิดพลาดในบางภาพ CT สามารถจำแนกรูปแบบของการทำนายผิดพลาดตามแต่ละอวัยวะที่มีความเสี่ยงดังนี้

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง brain stem ผลการทำนายผิดของโมเดลเป็น mask ของ brain stem ในบริเวณภาพส่วนเนื้อสมองโดยที่ mask อ้างอิงไม่ได้ถูก label เป็น brain stem แสดงการเปรียบเทียบ mask ทำนาย กับ mask อ้างอิงดังภาพที่ 31A_Ref กับภาพ 31A_Pre และแสดงผลการทำนาย mask ของ brain stem ที่มีขนาดใหญ่กว่าและเล็กกว่าขนาดของ mask อ้างอิง แสดงการเปรียบเทียบดังภาพที่ 31B_Ref, 31C_Ref กับภาพ 31B_Pre, 31C_Pre ตามลำดับ



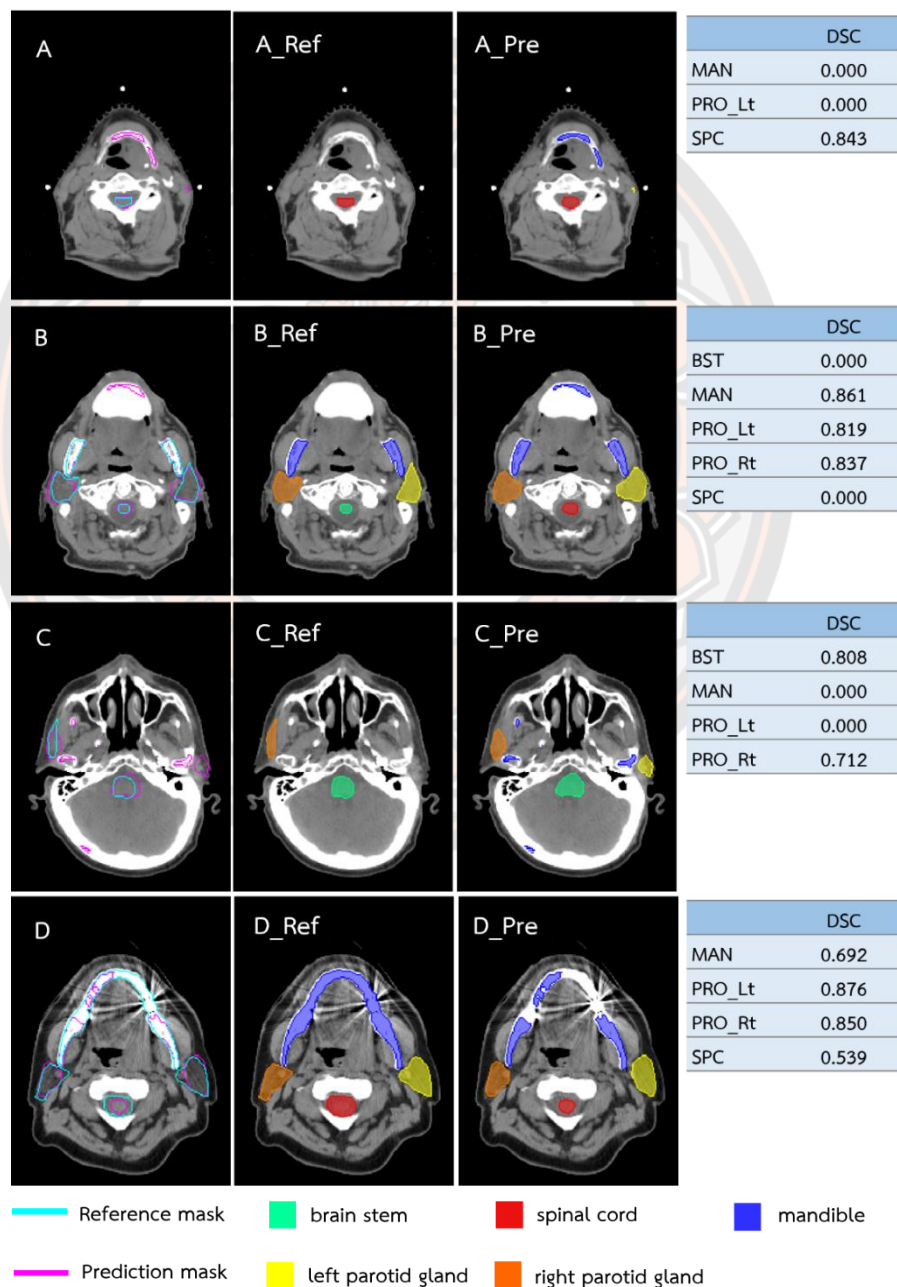
ภาพ 31 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง brain stem

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right eye globes ผลการทำนายผิดพลาดของโมเดลเป็น mask ของ eye globes บริเวณเนื้อเยื่อที่ปรากฏบนภาพ CT ในสไลซ์ส่วนขอบบนหรือขอบล่างของ eye globes จริงโดยที่ภาพ mask อ้างอิงไม่ได้ label เป็น eye globes แสดงการเปรียบเทียบดังภาพที่ 32A_Ref กับภาพที่ 32A_Pre และผลการทำนายผิดพลาดเป็น mask ของ eye globes ในบริเวณเนื้อเยื่ออื่นที่ตำแหน่งไกลจาก eye globes จริง เช่น เนื้อเยื่อที่ปรากฏบนภาพ CT สไลซ์ส่วนทรวงอก ส่วนคอ แสดงการเปรียบเทียบดังภาพที่ 32B_Ref, 32C_Ref กับภาพที่ 32B_Pre, 32C_Pre ตามลำดับ



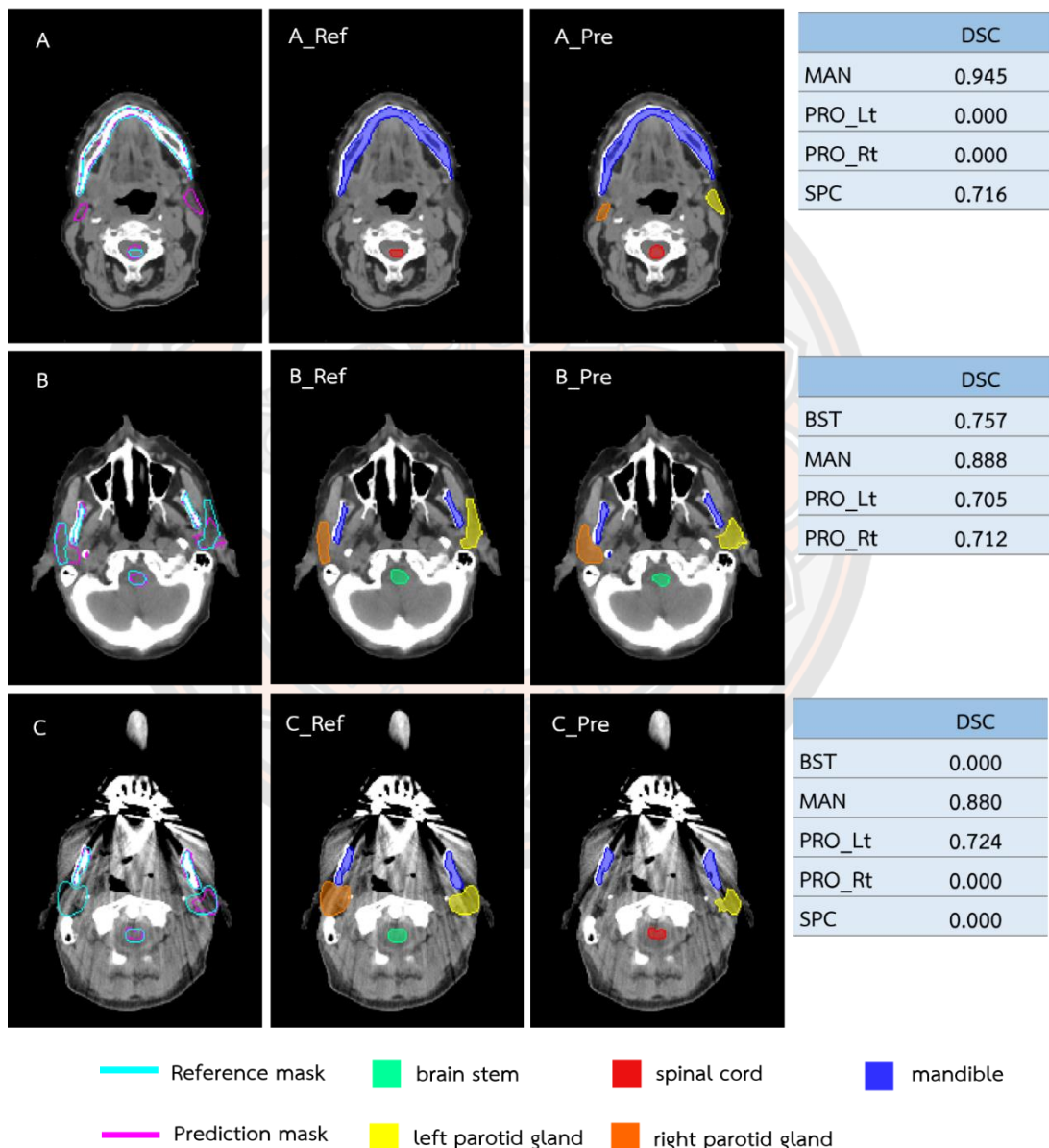
ภาพ 32 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง eye globes

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง mandible ผลการทำนายผิดของโมเดล เป็น mask ของ mandible ที่ตำแหน่งกระดูก Hyoid แสดงการเปรียบเทียบดังภาพที่ 33A_Ref กับ ภาพที่ 33A_Pre ผลการทำนายผิดเป็น mask ของ mandible ที่ตำแหน่งอุปกรณ์ยึดตรึงช่องปากที่ ปราบกฏบนภาพ CT แสดงการเปรียบเทียบดังภาพที่ 33B_Ref กับภาพที่ 33B_Pre ผลการทำนาย mask ของ mandible ขาดหรือเกินจาก mask อ้างอิงที่ตำแหน่ง TM joint และบางส่วนของกระดูก mandible แสดงการเปรียบเทียบดังภาพ 33C_Ref, 33D_Ref กับภาพ 33C_Pre, 33D_Pre ตามลำดับ



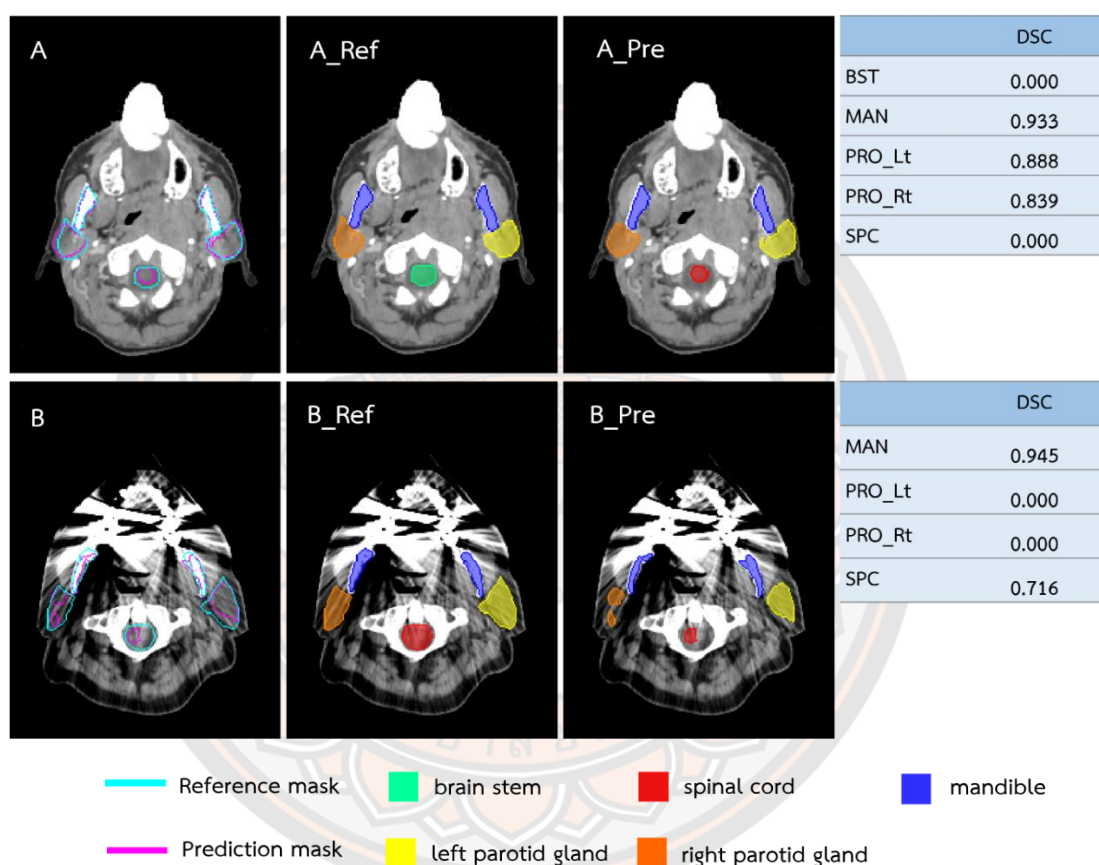
ภาพ 33 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง mandible

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right parotid glands ผลการทำนายผิดของโมเดลเป็น mask ของ parotid glands บริเวณเนื้อเยื่อที่ปรากฏบนภาพ CT ในสไลซ์ส่วนขอบบนหรือขอบล่างของ parotid glands จริง โดยที่ภาพ mask อ้างอิงไม่ได้ label เป็น parotid glands แสดงการเปรียบเทียบดังภาพที่ 34A_Ref กับภาพที่ 34A_Pre ผลการทำนาย mask parotid glands คลาดเคลื่อนที่บริเวณเนื้อเยื่ออื่นโดยรอบ parotid glands จริงแสดงการเปรียบเทียบดังภาพที่ 34B_Ref, 34C_Ref กับภาพที่ 34B_Pre, 34C_Pre ตามลำดับ



ภาพ 34 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง parotid glands

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง spinal cord ผลการทำนายผิดของโมเดลเป็น mask ของ spinal cord แต่ภาพ mask อ้างอิงถูก label เป็น brain stem แสดงการเปรียบเทียบดังภาพที่ 35A_Ref กับภาพที่ 35A_Pre โมเดลทำนาย mask ของ spinal cord คลาดเคลื่อนจากขอบเขตของ spinal cord ตามภาพอ้างอิง แสดงการเปรียบเทียบดังภาพที่ 35B_Ref กับภาพที่ 35B_Pre



ภาพ 35 ผลการทำนายผิดพลาดของโมเดล TUVGG19 ในอวัยวะที่มีความเสี่ยง spinal cord

จากผลการทำนาย mask ทั้ง 16 โมเดลบนชุดข้อมูลภาพทดสอบจำนวน 15 ชุด พบว่ามี 1 ชุดข้อมูล ที่ไม่สามารถให้ผลการทำนาย mask ได้ ดังนั้นการศึกษาในส่วนของค่า DSC และค่า 95%HD ของการศึกษานี้จะคำนวณเฉพาะข้อมูลใน 14 ชุดข้อมูล ที่สามารถทำนาย mask ได้เท่านั้น

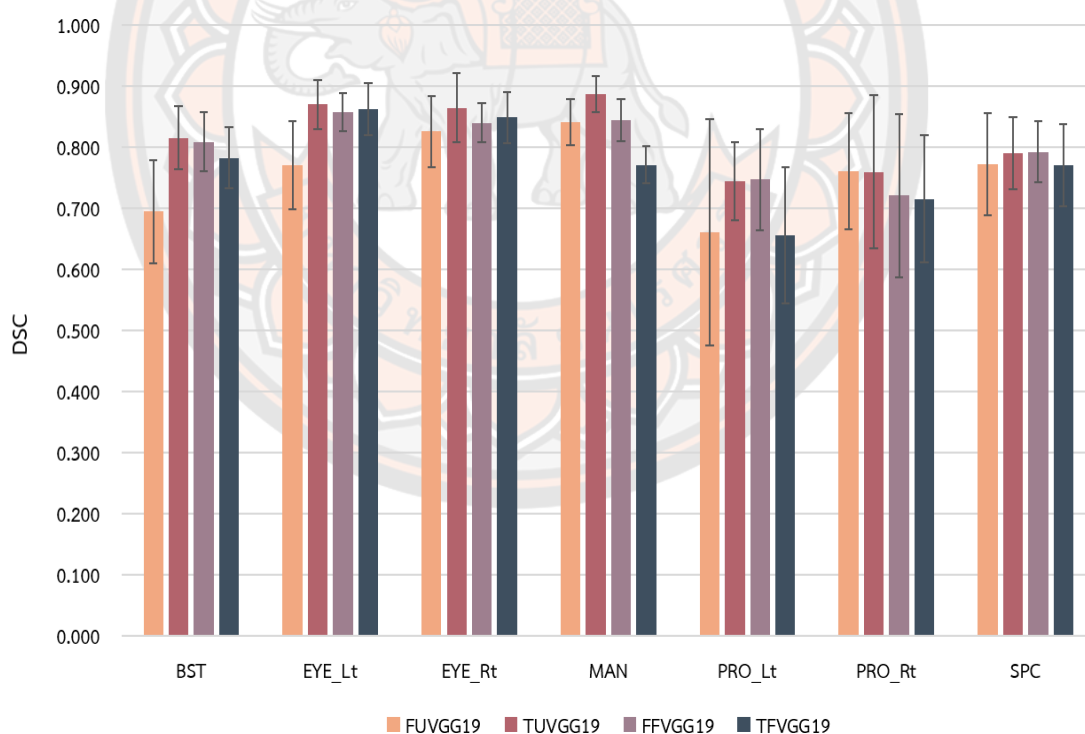
ผลการประเมินค่าสัมประสิทธิ์ความเหมือน (DSC)

ผลการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ VGG19 ถูกนำมาคำนวณหาค่า DSC ของภาพทำนาย 7 อวัยวะที่มีความเสี่ยง เปรียบเทียบกับภาพอ้างอิง ได้ผลค่าเฉลี่ย DSC ดังแสดงในภาพที่ 36

พบว่าโมเดล TUVGG19 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง brain stem, left eye globe, right eye globe และ mandible ค่าเฉลี่ย DSC เท่ากับ 0.815 ± 0.052 , 0.869 ± 0.040 , 0.864 ± 0.056 และ 0.887 ± 0.030 ตามลำดับ

โมเดล FFVGG19 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง left parotid gland และ spinal cord ค่าเฉลี่ย DSC เท่ากับ 0.746 ± 0.082 และ 0.792 ± 0.050 ตามลำดับ

โมเดล FUVGG19 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง right parotid gland ค่าเฉลี่ย DSC เท่ากับ 0.760 ± 0.095

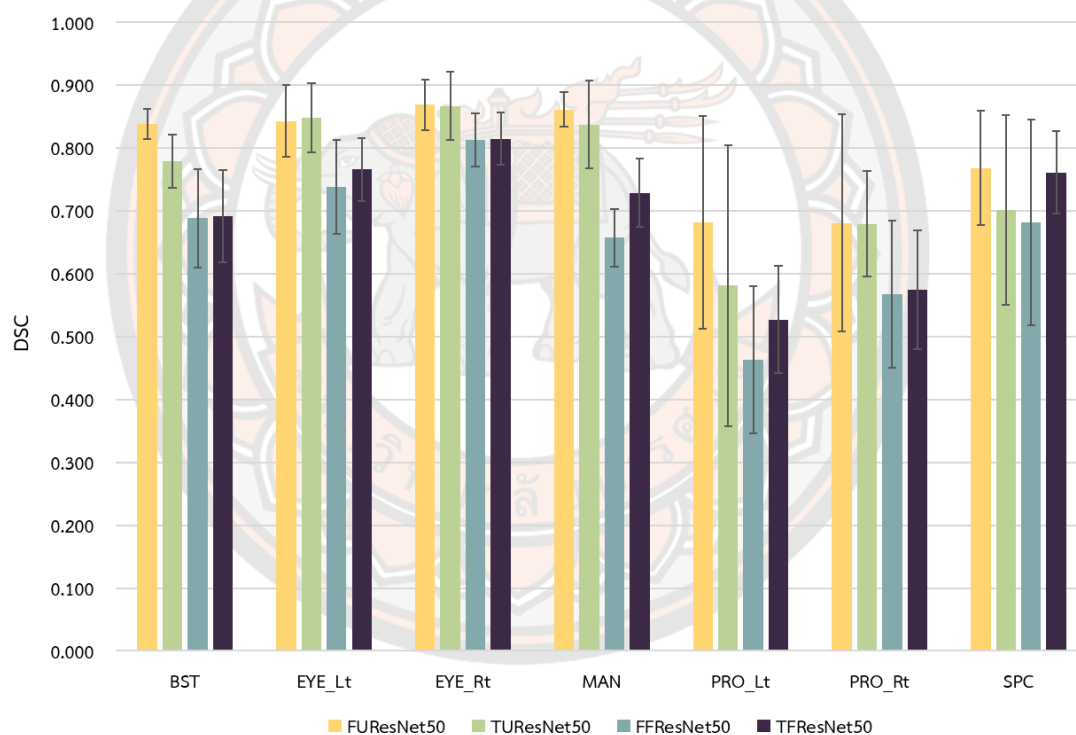


ภาพ 36 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ VGG19

ผลการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ ResNet50 ถูกนำมาคำนวณหาค่า DSC ของภาพทำนาย 7 อวัยวะที่มีความเสี่ยง เปรียบเทียบกับภาพอ้างอิง ได้ผลค่าเฉลี่ย DSC ดังแสดงในภาพที่ 37

พบว่าโมเดล FUResNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง brain stem, right eye globe, mandible, left parotid gland, right parotid gland และ spinal cord ค่าเฉลี่ย DSC เท่ากับ 0.839 ± 0.024 , 0.869 ± 0.040 , 0.861 ± 0.027 , 0.682 ± 0.169 , 0.681 ± 0.172 และ 0.768 ± 0.091 ตามลำดับ

โมเดล TURESNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง left eye globe ค่าเฉลี่ย DSC เท่ากับ 0.848 ± 0.055



ภาพ 37 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ ResNet50

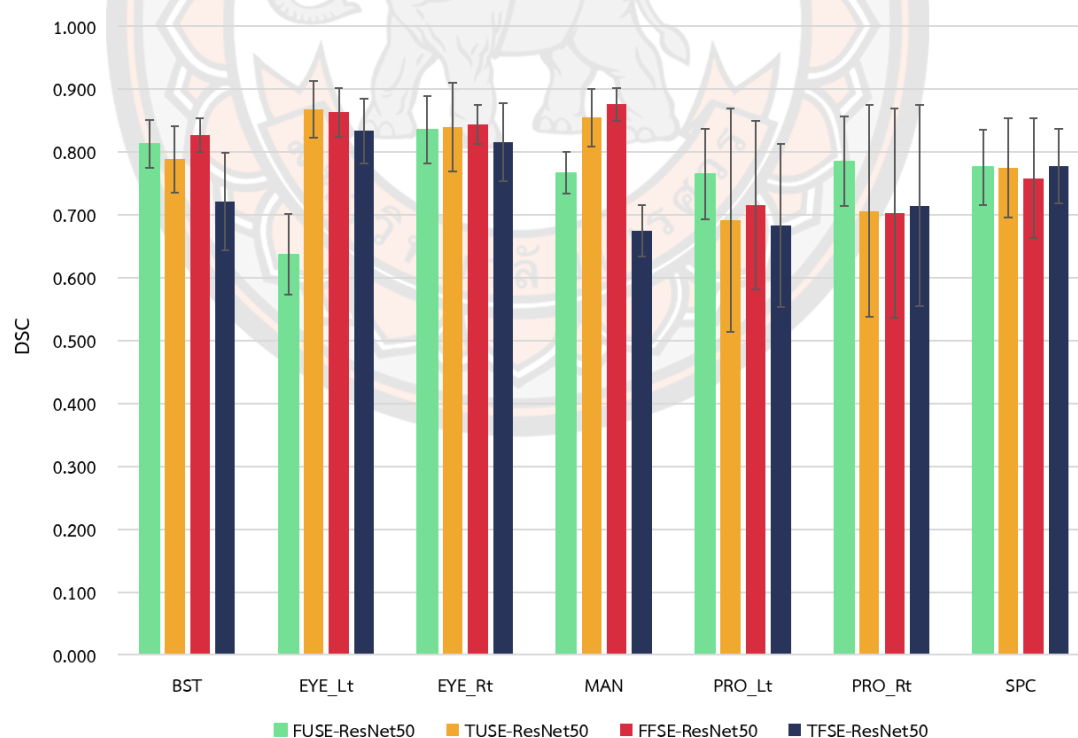
ผลการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบโดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ SE-ResNet50 ถูกนำมาคำนวณหาค่า DSC ของภาพทำนาย 7 อวัยวะที่มีความเสี่ยง เปรียบเทียบกับภาพอ้างอิง ได้ผลค่าเฉลี่ย DSC ดังแสดงในภาพที่ 38

พบว่าโมเดล FFSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง brain stem, right eye globe และ mandible ค่าเฉลี่ย DSC เท่ากับ 0.827 ± 0.027 , 0.844 ± 0.031 และ 0.876 ± 0.026 ตามลำดับ

โมเดล FUSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง left parotid gland และ right parotid gland ค่าเฉลี่ย DSC เท่ากับ 0.765 ± 0.072 และ 0.785 ± 0.071 ตามลำดับ

โมเดล TUSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง left eye globe ค่าเฉลี่ย DSC เท่ากับ 0.868 ± 0.045

โมเดล TFSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง spinal cord ค่าเฉลี่ย DSC เท่ากับ 0.778 ± 0.059

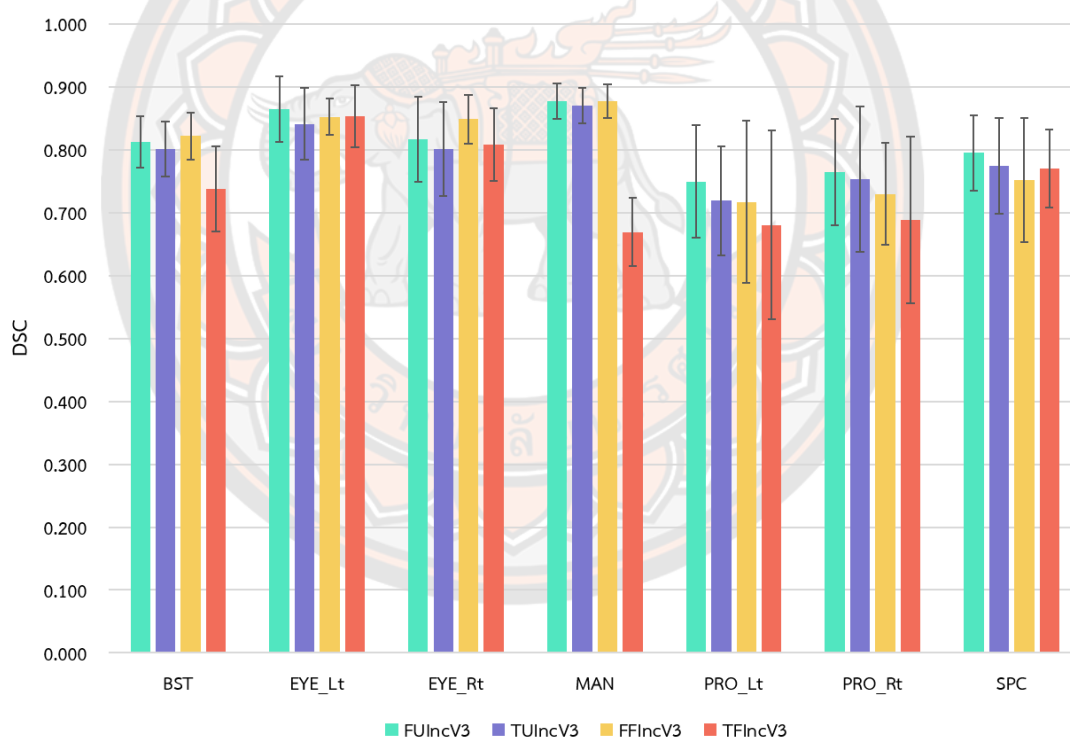


ภาพ 38 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ SE-ResNet50

ผลการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบโดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ InceptionV3 ถูกนำมาคำนวณหาค่า DSC ของภาพทำนาย 7 อวัยวะที่มีความเสี่ยง เปรียบเทียบกับภาพอ้างอิง ได้ผลค่าเฉลี่ย DSC ดังแสดงในภาพที่ 39

พบว่าโมเดล FUIncV3 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง left eye globe, mandible, left parotid gland, right parotid gland, และ spinal cord ค่าเฉลี่ย DSC เท่ากับ 0.865 ± 0.052 , 0.877 ± 0.028 , 0.750 ± 0.089 , 0.765 ± 0.084 และ 0.795 ± 0.060 ตามลำดับ

โมเดล FFIncV3 ทำนาย mask ที่มีค่าเฉลี่ย DSC สูงสุดในอวัยวะที่มีความเสี่ยง brain stem, right eye globe และ mandible ค่าเฉลี่ย DSC เท่ากับ 0.822 ± 0.037 , 0.849 ± 0.039 และ 0.877 ± 0.027 ตามลำดับ



ภาพ 39 ค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ InceptionV3

ผลการประเมินค่าระยะห่าง 95 เปอร์เซ็นไทล์ฮาวดอร์ฟ (95%HD)

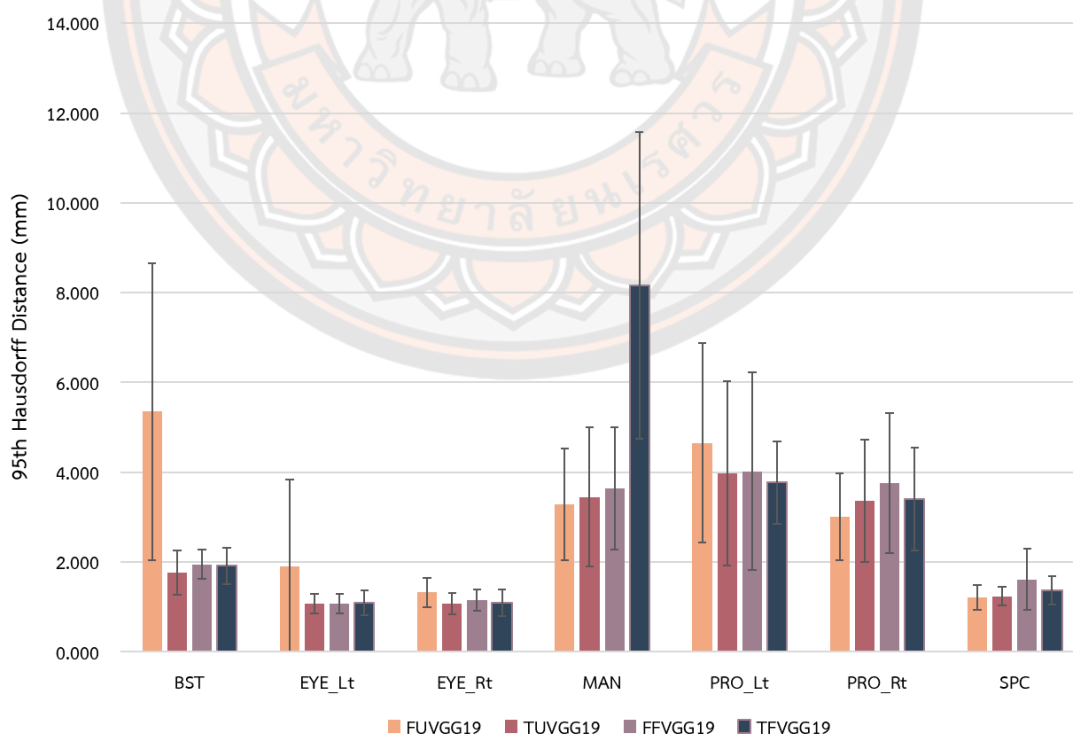
ผลการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ VGG19 ผลการคำนวณ ค่าเฉลี่ย 95%HD แสดงดังภาพที่ 40

ผลของโมเดล FUVGG19 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง mandible, right parotid gland และ spinal cord ค่าเฉลี่ย 95%HD เท่ากับ 3.288 ± 1.241 , 3.008 ± 0.960 และ 1.209 ± 0.281 มิลลิเมตร ตามลำดับ

โมเดล TUVGG19 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง brain stem และ right eye globe ค่าเฉลี่ย 95%HD เท่ากับ 1.764 ± 0.486 และ 1.079 ± 0.242 มิลลิเมตร ตามลำดับ

โมเดล FFVGG19 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง left eye globe ค่าเฉลี่ย 95%HD เท่ากับ 1.075 ± 0.212 มิลลิเมตร

โมเดล TFVGG19 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง left parotid gland ค่าเฉลี่ย 95%HD เท่ากับ 3.770 ± 0.919 มิลลิเมตร

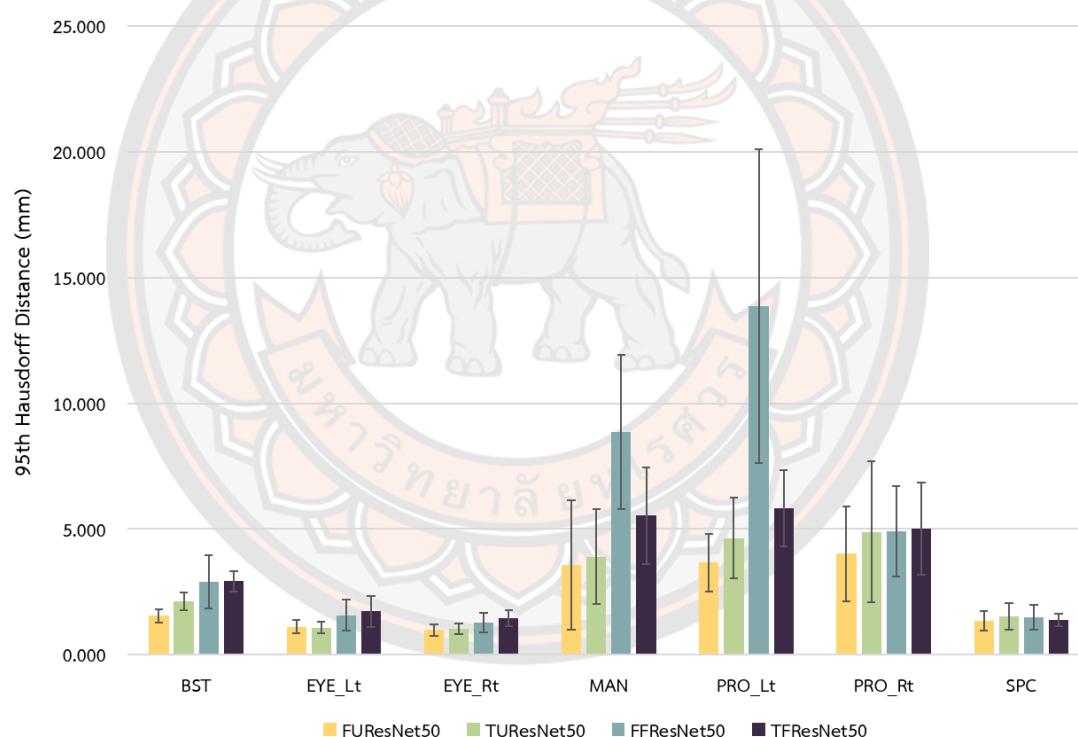


ภาพ 40 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ VGG19

ผลการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ ResNet50 ผลการคำนวณค่าเฉลี่ย 95%HD แสดงดังภาพที่ 41

ผลของโมเดล FUResNet50 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง brain stem, right eye globe, mandible, left parotid gland, right parotid gland และ spinal cord ค่าเฉลี่ย 95%HD เท่ากับ 1.544 ± 0.274 , 0.979 ± 0.217 , 3.563 ± 2.574 , 3.659 ± 1.155 , 4.012 ± 1.880 และ 1.347 ± 0.377 มิลลิเมตร ตามลำดับ

โมเดล TUResNet50 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง left eye globe ค่าเฉลี่ย 95%HD เท่ากับ 1.068 ± 0.226 มิลลิเมตร



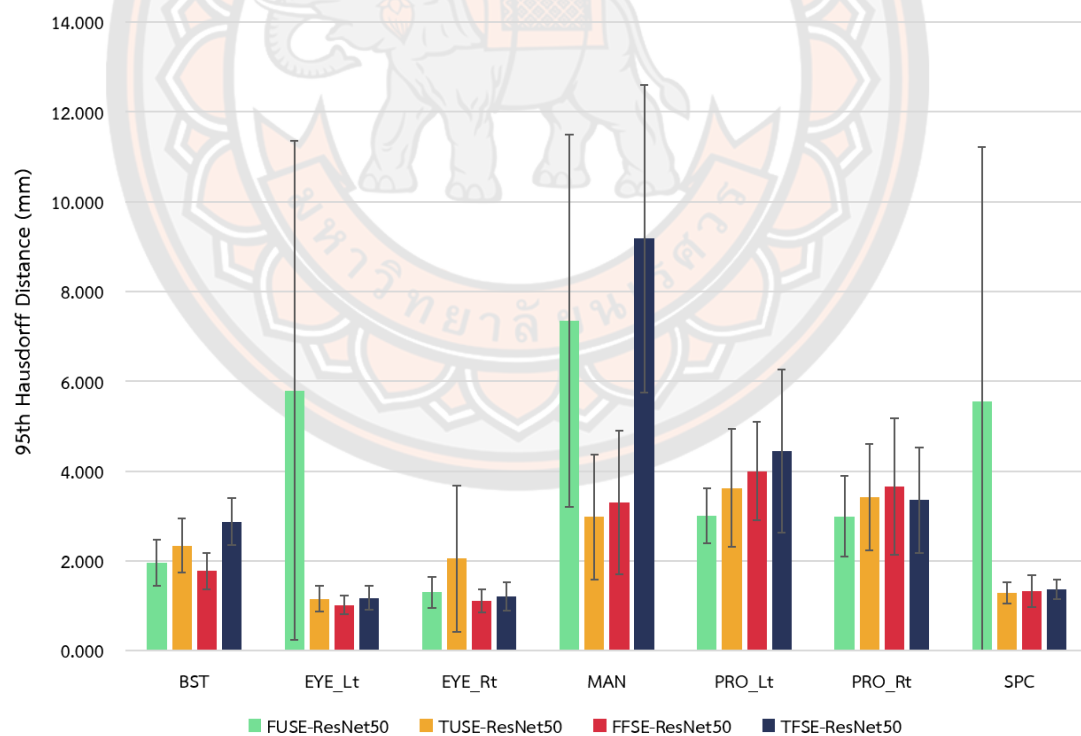
ภาพ 41 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ ResNet50

ผลการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ SE-ResNet50 ผลการคำนวณค่าเฉลี่ย 95%HD แสดงดังภาพที่ 42

ผลของโมเดล FFSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง brain stem, left eye globe และ right eye globe ค่าเฉลี่ย 95%HD เท่ากับ 1.775 ± 0.402 , 1.016 ± 0.206 และ 1.110 ± 0.247 มิลลิเมตร ตามลำดับ

โมเดล TUSE-ResNet50 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง mandible และ spinal cord ค่าเฉลี่ย 95%HD เท่ากับ 2.978 ± 1.396 และ 1.286 ± 0.238 มิลลิเมตร ตามลำดับ

โมเดล FUSE-ResNet50 มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง left parotid gland และ right parotid gland ค่าเฉลี่ย 95%HD เท่ากับ 3.006 ± 0.607 และ 2.992 ± 0.901 มิลลิเมตร ตามลำดับ



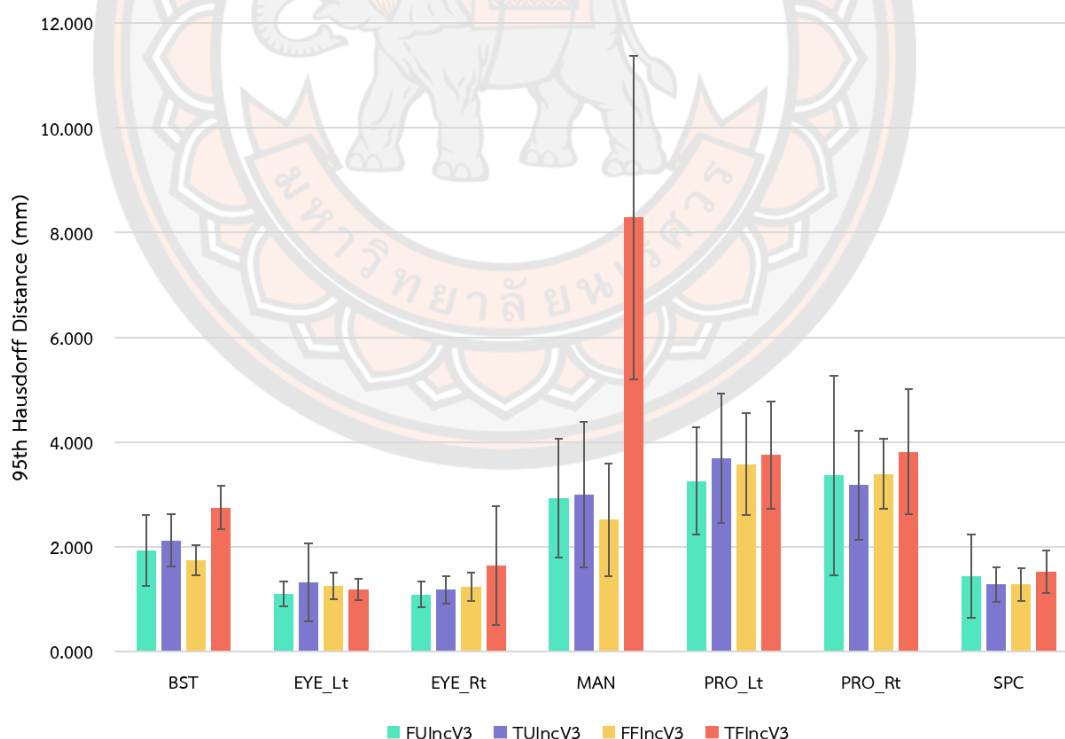
ภาพ 42 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ SE-ResNet50

ผลการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ โดยใช้โมเดลที่มีรูปแบบโมเดล backbone แบบ InceptionV3 ผลการคำนวณค่าเฉลี่ย 95%HD แสดงดังภาพ 43

ผลของโมเดล FUIncV3 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง right eye globe, left eye globe และ left parotid gland ค่าเฉลี่ย 95%HD เท่ากับ 1.101 ± 0.234 , 1.095 ± 0.239 และ 3.256 ± 1.025 มิลลิเมตร ตามลำดับ

โมเดล FFIncV3 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง brain stem, mandible และ spinal cord ค่าเฉลี่ย 95%HD เท่ากับ 1.745 ± 0.282 , 2.522 ± 1.076 และ 1.283 ± 0.313 มิลลิเมตร ตามลำดับ

โมเดล TUIncV3 ทำนาย mask ที่มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะที่มีความเสี่ยง right parotid gland และ spinal cord ค่าเฉลี่ย 95%HD เท่ากับ 3.183 ± 1.041 , 1.283 ± 0.334 มิลลิเมตร ตามลำดับ



ภาพ 43 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึกที่ใช้โมเดล backbone แบบ InceptionV3

ผลการประเมินเวลา

ผลการประเมินเวลาการทำงานของโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยง ข้อมูลเวลาสอนโมเดลและข้อมูลเวลาทำนาย mask ของแต่ละโมเดล แสดงดังตาราง 6 พบว่าโมเดลที่สอนโดยใช้วิธี transfer learning ใช้เวลาน้อยกว่าการสอนโมเดลด้วยวิธี fine tune ทั้ง 2 รูปแบบสถาปัตยกรรมโครงข่าย (U-Net และ FPN) เวลาลดลง 257.200, 296.000, 108.667, 75.000, 110.333, 104.500, 48.067, 106.000 วินาทีต่อ epoch สำหรับแต่ละสถาปัตยกรรม และโมเดล backbone UVGG19, FVGG19, UResNet50, FResNet50, USE-ResNet50, FSE-ResNet50, UIncV3 และ FIncV3 ตามลำดับ

เวลาการทำนาย mask ของโมเดลที่ถูกสอนด้วยวิธี fine tune และโมเดลที่สอนด้วยวิธี transfer learning มีความใกล้เคียงกันทั้ง 2 วิธี และทั้ง 2 สถาปัตยกรรมโครงข่าย โดยโมเดลที่ใช้สถาปัตยกรรมโครงข่าย U-Net มีแนวโน้มทำนาย Mask ได้เร็วกว่าโมเดลที่ใช้ FPN ยกเว้นโมเดลที่มีรูปแบบโมเดล backbone แบบ SE-ResNet50 ที่สถาปัตยกรรมโครงข่าย FPN ทำนาย mask ได้เร็วกว่าโมเดลที่ใช้ U-Net

ตาราง 6 ผลการประเมินเวลาสอนโมเดลและเวลาทำนาย mask ของโมเดล

| Models | Time train (second per epoch) | Time Prediction (second per image) |
|---------------|----------------------------------|---------------------------------------|
| FUVGG19 | 737.200 | 1.307 |
| TUVGG19 | 480.000 | 1.286 |
| FFVGG19 | 1056.000 | 1.534 |
| TFVGG19 | 760.000 | 1.430 |
| FUResNet50 | 781.167 | 1.348 |
| TUResNet50 | 672.500 | 1.415 |
| FFResNet50 | 978.500 | 1.435 |
| TFResNet50 | 903.500 | 1.471 |
| FUSE-ResNet50 | 792.000 | 1.452 |
| TUSE-ResNet50 | 681.667 | 1.415 |
| FFSE-ResNet50 | 1062.000 | 1.370 |
| TFSE-ResNet50 | 957.500 | 1.378 |
| FUIncV3 | 710.400 | 1.403 |
| TUIncV3 | 662.333 | 1.461 |
| FFIncV3 | 1016.500 | 1.425 |
| TFIncV3 | 910.500 | 1.429 |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3

บทที่ 5

บทสรุป

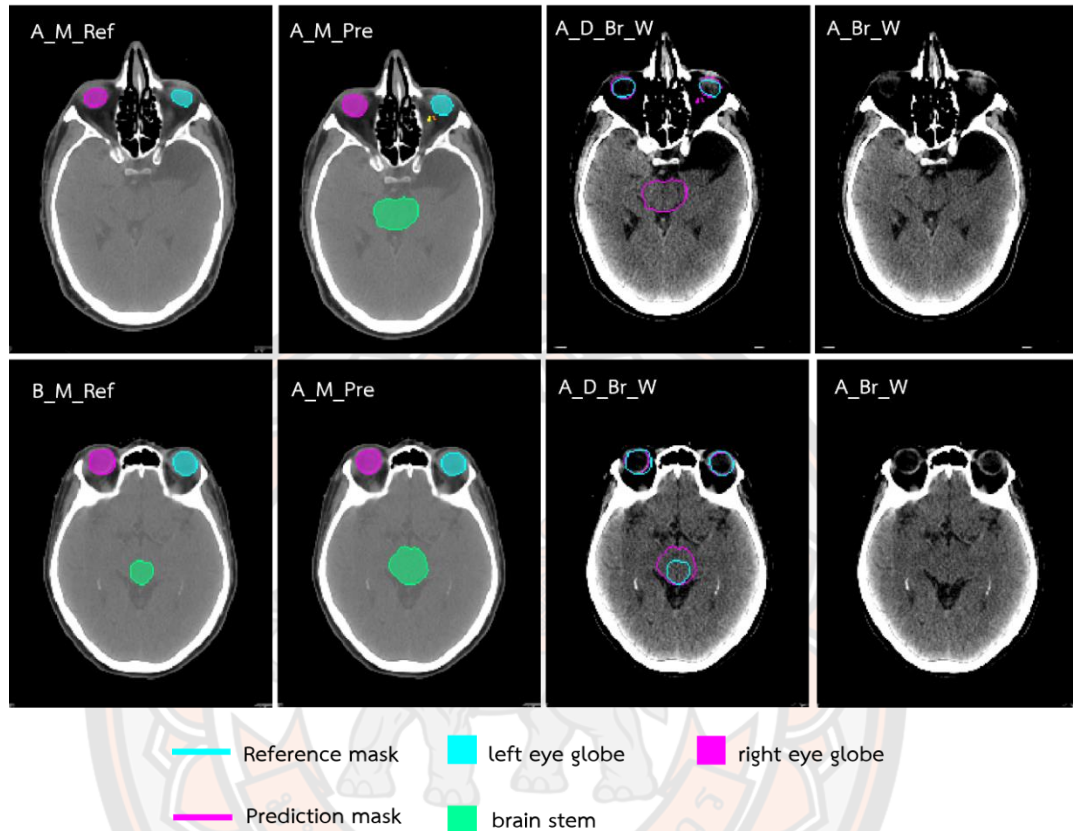
อภิปรายผลการทำนาย mask

การศึกษานี้พัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอัตโนมัติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษา โดยใช้รูปแบบสถาปัตยกรรมโครงข่าย 2 รูปแบบ ได้แก่ U-Net และ FPN ร่วมกับการใช้โมเดล backbone 4 โมเดล ได้แก่ โมเดล VGG19, ResNet50, SE-ResNet50 และ InceptionV3 และใช้วิธีการสอนโมเดล 2 วิธี ได้แก่ วิธี fine tune และวิธี transfer learning ด้วยค่าถ่วงน้ำหนัก ImageNet

จากผลการศึกษาพบว่าโมเดลที่พัฒนาขึ้นสามารถทำนาย mask ได้ในระดับดีในชุดภาพ CT ทดสอบส่วนมาก แต่ยังพบการทำนายผิดหรือคลาดเคลื่อนจาก mask อ้างอิงในบางสไลซ์ภาพ CT เมื่อพิจารณารูปแบบความคลาดเคลื่อนจำแนกตามประเภทอวัยวะที่มีความเสี่ยง พบว่ารูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง brain stem โมเดลทำนาย mask ผิดเป็นส่วนของ brain stem ในบริเวณภาพส่วนเนื้อสมองโดยที่ mask อ้างอิงไม่ถูก label เป็น brain stem แสดงการเปรียบเทียบดังภาพที่ 44A_M_Pre และ 44A_M_Ref พบความสัมพันธ์ของรูปแบบการทำนายผิดดังกล่าวกับภาพนำเข้าสู่ส่วน channel ภาพ brain window โดยภาพ brain window ทำให้ขอบเขตของเนื้อสมองส่วนต่าง ๆ ชัดเจนขึ้น ส่งผลให้โมเดลมีการตรวจจับคุณลักษณะภาพบริเวณที่มีรูปร่างกลมรีคล้าย brain stem และมีตำแหน่งอยู่บริเวณส่วนกลางของภาพกะโหลกศีรษะ ทำให้โมเดลทำนายบริเวณดังกล่าวเป็น mask อวัยวะที่มีความเสี่ยง brain stem แสดงดังภาพที่ 44A_D_Br_W และ 44A_Br_W กรณีเดียวกันนี้ยังทำให้มีการทำนายขอบเขตของ brain stem คลาดเคลื่อนโดยทำนายขอบเขตของ brain stem ใหญ่กว่าขอบเขตในภาพอ้างอิง แสดงการเปรียบเทียบดังภาพที่ 44B_M_Pre กับ 44B_M_Ref สำหรับผล mask ทำนายและ mask อ้างอิงบนภาพ soft tissue window ตามลำดับ ภาพ 44B_D_Br_W แสดงความแตกต่างผล mask ทำนาย และ mask อ้างอิงบนภาพ brain window และภาพ 44B_Br_W แสดงภาพ brain window

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right eye globes โมเดลทำนายผิดเป็น mask ของ eye globes บริเวณของเนื้อเยื่อที่ปรากฏบนภาพ CT สไลซ์ส่วนขอบบนหรือขอบล่างของ eye globes จริง และบริเวณเนื้อเยื่ออื่นที่อยู่ในไกลจากตำแหน่งของ eye globes จริง เช่น เนื้อเยื่อที่ปรากฏบนภาพ CT สไลซ์ส่วนทรวงอก ส่วนคอ โดยที่ภาพ mask อ้างอิงไม่ถูก label เป็น eye globes พบว่าบริเวณที่โมเดลทำนายเป็น mask ของ eye globes มีรูปร่างคล้ายกับ

ส่วนของ eye globes จริง และมีตำแหน่งอยู่ส่วนขอบของร่างกายในภาพ CT คล้ายกับตำแหน่งของ eye globes ส่งผลให้โมเดลทำนายบริเวณดังกล่าวเป็น mask ของ eye globes



ภาพ 44 เปรียบเทียบผลการทำนาย mask ผิดหรือคลาดเคลื่อน บนภาพ soft tissue window กับภาพ brain window

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง mandible โมเดลทำนายผิดเป็น mask ของ mandible ที่ตำแหน่งกระดูก Hyoid และที่ตำแหน่งอุปกรณ์ยึดตรึงช่องปากที่ปรากฏบนภาพ CT พบว่ามีความสัมพันธ์กันของรูปร่างกระดูก Hyoid และอุปกรณ์ยึดตรึงช่องปากที่ปรากฏบนภาพ CT กับรูปร่างกระดูก mandible และยังมีความสัมพันธ์ด้านความหนาแน่นของกระดูก mandible กับกระดูก Hyoid ทำให้โมเดลที่เรียนรู้จากคุณลักษณะภาพไม่สามารถแยกความแตกต่างนี้ได้ จึงทำนาย mask ผิด แม้ว่าภาพนำเข้าสู่ส่วน channel ที่เป็นภาพ bone window จะแสดงถึงความแตกต่างของอุปกรณ์ยึดตรึงช่องปากกับกระดูก แต่โมเดลยังคงทำนายบางส่วนของอุปกรณ์ยึดตรึงช่องปากที่รูปร่างโค้งเว้าคล้าย mandible เป็น mask ของ mandible นอกจากนี้พบว่าโมเดลทำนาย mask ของ mandible คลาดเคลื่อนที่บริเวณ TM joint แม้ว่าจะใช้ภาพ bone window เพื่อช่วยให้ขอบเขตของกระดูกชัดเจนขึ้น แต่ในบริเวณดังกล่าวนี้มีการเชื่อมกันของกระดูก

หลายส่วน โดยที่ขั้นตอนการสแกนโมเดลได้กำหนด label เฉพาะส่วนของกระดูก mandible เท่านั้น ทำให้การทำนาย mask ของโมเดลมีความคลาดเคลื่อน ทำนายกระดูกชนิดอื่นที่มีรูปร่างใกล้เคียงกับ mandible ในบริเวณดังกล่าวเป็น mask ของ Mandible

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right parotid glands โมเดลทำนายผิดเป็น mask ของ parotid glands บริเวณเนื้อเยื่อที่ปรากฏบนภาพ CT ในสไลด์ส่วนขอบบนหรือขอบล่างของ parotid glands จริง โดยที่ภาพ mask อ้างอิงไม่ได้ label เป็น parotid glands พบว่าการทำนายผิดของโมเดลในรูปแบบนี้มีความสัมพันธ์กับรูปร่าง และตำแหน่งของเนื้อเยื่อดังกล่าวมีความใกล้เคียงกับอวัยวะที่มีความเสี่ยง parotid glands จึงส่งผลให้โมเดลทำนายผิด นอกจากนี้พบว่าการทำนาย mask ของ parotid glands ยังคงมีการทำนายขอบเขตคลาดเคลื่อน ขนาด mask ทำนายไม่สัมพันธ์กับ mask อ้างอิง ถึงแม้ว่าจะใช้ภาพ soft tissue window เป็นส่วนหนึ่งของชุดภาพ CT นำเข้า แต่ขอบเขตของ Parotid gland ยังไม่ชัดเจนขึ้นเท่าที่ควร เนื่องจากบริเวณโดยรอบ parotid glands ประกอบด้วยอวัยวะที่มีลักษณะเป็นเนื้อเยื่อ ซึ่งมีความแตกต่างของ contrast บนภาพ CT น้อยจึงส่งผลต่อการทำนายของโมเดล

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง spinal cord โมเดลทำนาย mask ผิดจาก mask อ้างอิงที่ถูกกำหนดเป็น brain stem เป็นอวัยวะที่มีความเสี่ยง spinal cord ในกรณีนี้พบว่ามักเกิดในบริเวณรอยต่อขอบบนของ spinal cord กับขอบล่างของ brain stem มีสาเหตุมาจากชุดข้อมูลสอนซึ่งถูกเก็บมาจากหลายแหล่งข้อมูล ซึ่งกำหนดขอบเขตส่วนขอบบนและขอบล่างของ spinal cord และ brain stem ไม่ชัดเจน แม้ว่าการปฏิบัติทางคลินิกมีการแนะนำตามคู่มือการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงอย่างชัดเจน ที่ระบุว่าให้กำหนดขอบบนของ spinal cord ที่ตำแหน่ง Tip of dens ของ C2 ซึ่งพบว่ามีมีความคลาดเคลื่อนของขอบเขตนี้เช่นเดียวกับการศึกษาของ van der Veen J และคณะ (28)

นอกจากการทำนายผิดหรือคลาดเคลื่อนที่เกิดจากลักษณะจำเพาะของแต่ละอวัยวะที่มีความเสี่ยงและลักษณะของอวัยวะอื่นโดยรอบแล้ว ยังมีปัจจัยอื่นที่ส่งผลให้การทำนาย mask ของโมเดลเกิดความคลาดเคลื่อนกล่าวคือ ปัจจัยด้านคุณภาพของภาพ จากผลการทำนายแสดงดังภาพที่ 31C, 33D, 34C และ 35B พบว่าสิ่งแปลกปลอม (Artifact) จากอุปกรณ์ทางทันตกรรมส่งผลให้โมเดลทำนาย mask ของ brain stem, mandible, parotid glands และ spinal cord คลาดเคลื่อนจากขอบเขตของ mask อ้างอิง ซึ่งสิ่งแปลกปลอม (Artifact) จากอุปกรณ์ทางทันตกรรมนี้ ไม่เพียงส่งผลต่อโมเดลการเรียนรู้เชิงลึกเท่านั้น ยังส่งผลต่อความมั่นใจของแพทย์และความถูกต้องของขอบเขตอวัยวะจากการกำหนดขอบเขตอวัยวะบนภาพ CT ที่มีสิ่งแปลกปลอมโดยแพทย์ เช่น การศึกษาของ Kovacs DG และคณะ (48) ที่นำเสนอการลดสิ่งแปลกปลอมด้วยอัลกอริทึมลดสิ่งแปลกปลอม หรือการใช้ dual energy computed tomography ซึ่งสามารถช่วยเพิ่มความถูกต้องของการกำหนด

ขอบเขตอวัยวะและเพิ่มความมั่นใจของแพทย์ได้ ดังนั้นการใช้วิธีการลดสิ่งแปลกปลอมดังกล่าวนี้เป็นแนวทางหนึ่งเพื่อใช้เตรียมภาพ CT นำเข้าสำหรับการทดสอบโมเดล อาจช่วยเพิ่มความถูกต้องของผลการทำนาย mask จากโมเดล และเสนอให้มีการเพิ่มข้อมูลภาพ CT ที่มีสิ่งแปลกปลอมสำหรับการสอนโมเดลที่มากขึ้น อาจช่วยปรับปรุงการเรียนรู้โมเดลและทำให้ผลการทำนาย mask ของโมเดลดีขึ้น

จากผลทำนาย mask ทั้ง 16 โมเดล พบว่ามี 1 ชุดข้อมูลการทดสอบจากทั้งหมด 15 ชุดข้อมูลที่ไม่สามารถทำนาย mask ได้ เมื่อพิจารณาความแตกต่างของข้อมูลกลุ่มทดสอบทั้ง 15 ชุด พบว่าข้อมูลชุดที่ไม่สามารถทำนาย mask ได้ มีขนาด field of view (FOV) เท่ากับ 320×320 มิลลิเมตร ต่างจากชุดข้อมูลอีก 14 ชุด ที่มีขนาด FOV อยู่ในช่วง 522×522 มิลลิเมตร ถึง 650×650 มิลลิเมตร ซึ่งมีความใกล้เคียงกันและใกล้เคียงกับขนาด FOV ของชุดข้อมูลสอนโมเดล ด้วยเหตุผลดังกล่าวนี้ แสดงให้เห็นว่าขนาดของ FOV เป็นปัจจัยหนึ่งที่ส่งผลต่อการทำนายของโมเดลที่พัฒนาขึ้น โดยแนะนำให้ใช้ทำนายภาพที่มีขนาด FOV ตั้งแต่ 500 ถึง 600 มิลลิเมตร จึงจะทำให้โมเดลสามารถทำนายผล mask ได้

อภิปรายผลค่าเฉลี่ย DSC และค่าเฉลี่ย 95%HD

จากผลการศึกษาส่วนของการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบกับภาพ mask อ้างอิง ทั้ง 14 ชุดข้อมูล ผลการประเมินค่าเฉลี่ย DSC พบว่าโมเดลที่สามารถทำนาย mask แล้วให้ค่าเฉลี่ย DSC สูงในแต่ละรูปแบบโมเดล backbone มีดังนี้

โมเดล backbone แบบ VGG19 ได้แก่ โมเดล TUVGG19 และ FFVGG19 ค่าเฉลี่ย DSC มากกว่า 0.80 ยกเว้นใน parotid glands ทั้ง 2 ข้าง และ spinal cord ค่าเฉลี่ย DSC มากกว่า 0.72

โมเดล backbone แบบ ResNet50 ได้แก่ โมเดล FUResNet50 ค่าเฉลี่ย DSC มากกว่า 0.839 ยกเว้นใน parotid glands ทั้ง 2 ข้าง ค่าเฉลี่ย DSC มากกว่า 0.680 และ spinal cord ค่าเฉลี่ย DSC เท่ากับ 0.768

โมเดล backbone แบบ SE-ResNet50 ได้แก่ โมเดล FFSE-ResNet50 ค่าเฉลี่ย DSC มากกว่า 0.827 ยกเว้นใน parotid glands ทั้ง 2 ข้าง และ spinal cord ค่าเฉลี่ย DSC มากกว่า 0.703

โมเดล backbone แบบ InceptionV3 ได้แก่ โมเดล FUIncV3, TUIncV3 และโมเดล FFIncV3 ค่าเฉลี่ย DSC มากกว่า 0.801 ยกเว้นใน parotid glands ทั้ง 2 ข้าง และ spinal cord ค่าเฉลี่ย DSC มากกว่า 0.717

จากค่าเฉลี่ย DSC ของโมเดลที่สามารถทำนาย mask ได้ดีในแต่ละรูปแบบโมเดล backbone ผลค่าเฉลี่ย DSC มีแนวโน้มใกล้เคียงกับผลค่าเฉลี่ย DSC จากการศึกษาอื่นก่อนหน้านี้ แสดงการเปรียบเทียบดังตาราง 7

จากผลการศึกษาร่วมการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ พบว่าโมเดลที่สามารถทำนาย mask ได้ดีในแต่ละรูปแบบโมเดล backbone ให้ค่าเฉลี่ย 95%HD ต่ำ ซึ่งสัมพันธ์แบบแปรผกผันกับผลค่าเฉลี่ย DSC (สูง) โดยค่าเฉลี่ย 95%HD จากโมเดล TUVGG19, FUResNet50, FFSE-ResNet50, และโมเดล FUIncV3 มีค่าใกล้เคียงกับการศึกษาอื่นก่อนหน้านี้แสดงการเปรียบเทียบดังตาราง 8

จากผลค่าเฉลี่ย DSC และค่าเฉลี่ย 95%HD พบรูปแบบการทำนายที่สอดคล้องกันของโมเดลที่ให้ผลการทำนาย mask ที่ดี (ค่าเฉลี่ย DSC สูง และค่าเฉลี่ย 95%HD ต่ำ) กล่าวคือ สามารถทำนายขอบเขตต่อวัยวะที่มีความแตกต่างของเนื้อเยื่อสูงได้ดีกว่าการทำนายขอบเขตต่อวัยวะที่มีความแตกต่างของเนื้อเยื่อต่ำ สอดคล้องกับผลค่าเฉลี่ย DSC และค่าเฉลี่ย 95%HD จากการศึกษาของ Ibragimov B และ Xing L.(13) และ Wu X.และคณะ (15) แม้ว่าวิธีการเตรียมข้อมูลภาพของการศึกษานี้ กำหนดค่า windows ในชุดข้อมูลภาพ CT นำเข้า โดยเลือกใช้ soft tissue windows เป็นส่วนหนึ่งของชุดภาพนำเข้า เพื่อเพิ่มความแตกต่างขอบเขตของ parotid glands กับเนื้อเยื่อโดยรอบมากขึ้น แต่จากผลการประเมินพบว่า การเลือกใช้ค่า windows นี้ยังไม่สามารถเพิ่มประสิทธิภาพการแยกขอบเขตให้โมเดลที่พัฒนาขึ้นได้ ส่วนการกำหนดค่าข้อมูลภาพ CT นำเข้าด้วย bone windows และ brain windows โมเดลที่พัฒนาขึ้นสามารถแยกขอบเขตต่อวัยวะที่มีความเสี่ยงในบริเวณที่สัมพันธ์กับทั้ง 2 windows นี้ ให้ผลค่อนข้างดีกับภาพ CT ส่วนมากในชุดข้อมูลทดสอบ ยกเว้นบางภาพ CT ดังที่อภิปรายก่อนหน้านี้ที่ยังเป็นข้อจำกัดสำหรับการทำนาย mask ของโมเดล

เมื่อพิจารณาความสัมพันธ์ของผลค่าเฉลี่ย DSC สูง และผลค่าเฉลี่ย 95%HD ต่ำ กับ รูปแบบสถาปัตยกรรมโครงข่าย วิธีการสอนโมเดล และรูปแบบโมเดล backbone พบว่าการเลือกใช้รูปแบบสถาปัตยกรรมโครงข่าย และวิธีการสอนที่ทำให้โมเดลมีค่าพารามิเตอร์ที่สามารถสอนได้ในจำนวนที่ไม่มากหรือน้อยเกินไปร่วมกับโมเดล backbone แบบ VGG19, SE-ResNet50 และ InceptionV3 จะส่งผลให้โมเดลมีการเรียนรู้ได้ดีและสามารถทำนาย mask ที่มีค่าเฉลี่ย DSC สูง และค่าเฉลี่ย 95%HD ต่ำ เช่น

โมเดล TUVGG19 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่าโมเดล FUVGG19 ที่มีจำนวนพารามิเตอร์ที่สอนได้มากเกินไป

โมเดล FFVGG19 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่าโมเดล TFVGG19 ที่มีจำนวนพารามิเตอร์ที่สอนได้น้อยเกินไป

โมเดล FFSE-ResNet50 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่าโมเดล TFSE-ResNet5 ที่มีจำนวนพารามิเตอร์ที่สอนได้น้อยเกินไป

โมเดล FUIncV3, TUIncV3, FFIncV3 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่าโมเดล TFIncV3 ที่มีจำนวนพารามิเตอร์ที่สอนได้น้อยเกินไป

โดยแตกต่างจากรูปแบบโมเดล backbone แบบ ResNet50 ที่โมเดล FUResNet50 มีการเรียนรู้ได้ดีและสามารถทำนาย mask ที่มีค่าเฉลี่ย DSC สูง และค่าเฉลี่ย 95%HD ต่ำ มีจำนวนพารามิเตอร์ที่สอนได้มากกว่าโมเดลอื่นที่ใช้รูปแบบโมเดล backbone แบบ ResNet50

จากโมเดลที่ให้ผลค่าเฉลี่ย DSC สูงของแต่ละรูปแบบโมเดล backbone ดังกล่าวข้างต้นสังเกตได้ว่าเป็นโมเดลที่ใช้รูปแบบสถาปัตยกรรมโครงข่าย U-Net มากกว่าสถาปัตยกรรมโครงข่าย FPN และใช้วิธีสอนโมเดลด้วยวิธี fine tune มากกว่าวิธี transfer learning แสดงให้เห็นว่ารูปแบบสถาปัตยกรรมโครงข่าย U-Net สามารถใช้ร่วมกับโมเดล backbone ได้หลายโมเดลมากกว่าสถาปัตยกรรมโครงข่าย FPN เนื่องจากรูปแบบสถาปัตยกรรมโครงข่าย U-Net ส่วน decoder มีจำนวนพารามิเตอร์มากกว่าสถาปัตยกรรมโครงข่าย FPN ทำให้เมื่อใช้การสอนโมเดลด้วยวิธี transfer learning ซึ่งมีการตรึงพารามิเตอร์ในส่วน encoder สถาปัตยกรรมโครงข่าย U-Net จะมีจำนวนพารามิเตอร์ที่สอนได้มากกว่าสถาปัตยกรรมโครงข่าย FPN จึงส่งผลให้สถาปัตยกรรม U-Net ที่สอนโมเดลด้วยวิธี transfer learning สามารถเรียนรู้รูปแบบของข้อมูลที่ใช้สอนได้ดีกว่าโมเดลที่ใช้สถาปัตยกรรมโครงข่าย FPN ร่วมกับการสอนโมเดลด้วยวิธี transfer learning และด้วยปัจจัยจำนวนพารามิเตอร์ที่สามารถสอนได้ดังที่กล่าวมานี้ ทำให้เมื่อใช้วิธีการสอนโมเดลด้วยวิธี fine tune ซึ่งใช้ค่าถ่วงน้ำหนักของโมเดล backbone เป็นค่าถ่วงน้ำหนักในการเริ่มต้นการเรียนรู้ของโมเดล โดยที่ยังสามารถปรับปรุงค่าถ่วงน้ำหนักได้ จึงส่งผลให้โมเดลที่ใช้วิธี fine tune ในขั้นตอนการสอนโมเดลสามารถเรียนรู้ และปรับพารามิเตอร์ให้เข้ากับข้อมูลที่ใช้สอนได้ดีกว่าโมเดลที่สอนด้วยวิธี transfer learning แต่ผลการศึกษาที่อภิปรายนี้อยู่ภายใต้เงื่อนไขการทดลองโดยใช้วิธี transfer learning ด้วยการตรึงค่าพารามิเตอร์ส่วน encoder ทั้งหมด หากมีการปรับวิธีการศึกษาโดยการเลือกตรึงค่าพารามิเตอร์บางส่วนของ encoder ในจำนวนที่เหมาะสม อาจได้ผลการศึกษาที่แตกต่างออกไปจากผลการศึกษานี้

ตาราง 7 ผลค่าเฉลี่ย DSC จากโมเดลที่ให้ผลการทำนายที่ตรงการศึกษานี้เปรียบเทียบกับค่าเฉลี่ย DSC จากการศึกษาอื่นก่อนหน้านี้ (ค่าเฉลี่ย DSC \pm SD, หน่วย: ร้อยละ)

| | TUVGG19 | FUResNet50 | FFSE-ResNet50 | FUIncV3 | Men K. (18) | Ibragimov B. (13) | Zhu W. (17) | Tong N (16) | Wang Y. (14) | Chen A. (49) | Wu X. (15) |
|---------------|-------------------|-------------------|-------------------|------------------|------------------|-------------------|------------------|------------------|------------------|--------------|------------|
| BST | 81.50 \pm 5.19 | 83.90 \pm 2.40 | 82.69 \pm 2.68 | 81.32 \pm 4.09 | 90.00 \pm 2.00 | NA | 86.65 \pm 2.00 | 82.97 \pm 2.95 | 87.50 \pm 2.20 | 80.30 | N/A |
| EYE_Lt | 86.93 \pm 4.01 | 84.30 \pm 5.71 | 86.32 \pm 3.87 | 86.46 \pm 5.19 | 93.00 \pm 1.00 | 88.40 \pm 2.70 | N/A | N/A | N/A | N/A | N/A |
| EYE_Rt | 86.43 \pm 5.61 | 86.90 \pm 4.04 | 84.43 \pm 3.12 | 81.73 \pm 6.77 | 92.00 \pm 2.00 | 87.70 \pm 3.70 | N/A | N/A | N/A | N/A | N/A |
| MAN | 88.70 \pm 2.96 | 86.11 \pm 2.74 | 87.58 \pm 2.56 | 87.72 \pm 2.84 | 92.00 \pm 2.00 | 89.50 \pm 3.60 | 92.51 \pm 2.00 | 93.67 \pm 1.12 | 93.00 \pm 1.90 | 91.70 | 89.00 |
| PRO_Lt | 74.35 \pm 6.40 | 68.22 \pm 16.90 | 71.62 \pm 13.41 | 74.95 \pm 8.93 | 86.00 \pm 3.00 | 76.60 \pm 6.10 | 88.07 \pm 2.00 | 83.49 \pm 2.29 | 86.40 \pm 2.60 | 81.10 | 77.00 |
| PRO_Rt | 75.95 \pm 12.53 | 68.10 \pm 17.24 | 70.29 \pm 16.58 | 76.52 \pm 8.44 | 86.00 \pm 3.00 | 77.90 \pm 5.40 | 87.35 \pm 4.00 | 83.18 \pm 1.45 | 84.80 \pm 7.00 | 81.40 | 76.00 |
| SPC | 78.97 \pm 5.87 | 76.79 \pm 9.10 | 75.86 \pm 9.48 | 79.54 \pm 5.97 | 91.00 \pm 1.00 | 87.00 \pm 3.20 | N/A | N/A | N/A | N/A | N/A |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3, N/A = not available

ตาราง 8 ผลค่าเฉลี่ย 95%HD จากโมเดลที่ให้ผลการทำนายที่ตั้งของการศึกษาที่เปรียบเทียบกับค่าเฉลี่ย 95%HD จากการศึกษาอื่นก่อนหน้า (ค่าเฉลี่ย 95%HD \pm SD, หน่วย: มิลลิเมตร)

| | TUVGG19 | FUResNet50 | FFSE-ResNet50 | FUIncV3 | Zhu W. (17) | Tong N. (16) | Wang Y. (14) | Chen A. (49) | Wu X. (15) |
|---------------|-----------------|-----------------|-----------------|-----------------|------------------|-----------------|-----------------|-----------------|---------------|
| BST | 1.76 \pm 0.49 | 1.54 \pm 0.27 | 1.78 \pm 0.40 | 1.93 \pm 0.67 | 6.42 \pm 2.38 | 4.01 \pm 0.93 | 2.01 \pm 0.33 | 5.15 | NA |
| EYE_Lt | 1.08 \pm 0.22 | 1.12 \pm 0.25 | 1.02 \pm 0.21 | 1.10 \pm 0.23 | N/A | N/A | N/A | N/A | N/A |
| EYE_Rt | 1.08 \pm 0.24 | 0.98 \pm 0.22 | 1.11 \pm 0.25 | 1.10 \pm 0.24 | N/A | N/A | N/A | N/A | N/A |
| MAN | 3.45 \pm 1.56 | 3.56 \pm 2.57 | 3.30 \pm 1.60 | 2.93 \pm 1.13 | 6.28 \pm 2.21 | 1.50 \pm 0.32 | 1.26 \pm 0.50 | 2.49 | 1.60 |
| PRO_Lt | 3.98 \pm 2.05 | 3.66 \pm 1.16 | 4.00 \pm 1.10 | 3.26 \pm 1.03 | 9.31 \pm 3.32 | 5.35 \pm 3.30 | 2.41 \pm 0.54 | 6.97 | 3.25 |
| PRO_Rt | 3.37 \pm 1.36 | 4.01 \pm 1.88 | 3.66 \pm 1.52 | 3.37 \pm 1.91 | 10.08 \pm 5.09 | 4.20 \pm 1.27 | 2.93 \pm 1.48 | 6.43 | 3.23 |
| SPC | 1.24 \pm 0.20 | 1.35 \pm 0.38 | 1.33 \pm 0.35 | 1.44 \pm 0.80 | N/A | N/A | N/A | N/A | N/A |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3, N/A = not available

อภิปรายผลการประเมินเวลา

จากผลการประเมินเวลาการทำงานของโมเดล เมื่อเปรียบเทียบโมเดลที่ใช้รูปแบบโมเดล backbone เดียวกัน พบว่าโมเดลที่ใช้สถาปัตยกรรมโครงข่าย U-Net ใช้เวลาในการสอนโมเดล (second per epoch) น้อยกว่าสถาปัตยกรรมโครงข่าย FPN แม้ว่าจะมีจำนวนพารามิเตอร์มากกว่า และพบว่าการสอนโมเดลด้วยวิธี transfer learning ใช้เวลาในการสอน (second per epoch) น้อยกว่าวิธี fine tune ทั้ง 2 รูปแบบสถาปัตยกรรมโครงข่าย แสดงให้เห็นว่ารูปแบบสถาปัตยกรรมโครงข่ายเป็นปัจจัยที่มีผลต่อระยะเวลาการสอนโมเดลมากที่สุด และจำนวนพารามิเตอร์ที่สามารถสอนได้ของแต่ละสถาปัตยกรรมเป็นปัจจัยรองที่ส่งผลกระทบต่อระยะเวลาการสอนโมเดล

เมื่อเปรียบเทียบระยะเวลาสอนโมเดลจากวิธีการสอน และรูปแบบสถาปัตยกรรมโครงข่าย เดียวกันที่โมเดล backbone ต่างกัน พบว่ารูปแบบโมเดล backbone ที่ใช้เวลาในการสอน (second per epoch) เรียงลำดับจากน้อยไปมากแสดงดังตารางที่ 9

ตาราง 9 การเปรียบเทียบระยะเวลาสอนโมเดลจากวิธีการสอนและรูปแบบสถาปัตยกรรมโครงข่าย เดียวกันที่โมเดล backbone ต่างกัน

| Training methods | Network architecture | Training time (second per epoch) of each backbone models in ascending order. |
|-------------------|----------------------|--|
| fine tune | U-Net | InceptionV3 < VGG19 < ResNet50 < SE-ResNet50 |
| transfer learning | U-Net | VGG19 < InceptionV3 < ResNet50 < SE-ResNet50 |
| fine tune | FPN | ResNet50 < InceptionV3 < VGG19 < SE-ResNet50 |
| transfer learning | FPN | VGG19 < ResNet50 < InceptionV3 < SE-ResNet50 |

จะเห็นได้ว่าการเลือกใช้โมเดล backbone ต่างกันมีผลต่อระยะเวลาการสอนโมเดล (second per epoch) โดยขึ้นอยู่กับความเหมาะสมของโมเดล backbone กับรูปแบบสถาปัตยกรรมโครงข่ายหลัก และจำนวนพารามิเตอร์โดยรวมทั้งหมด (พารามิเตอร์ที่สอนได้ของแต่ละรูปแบบสถาปัตยกรรม และค่าถ่วงน้ำหนักที่ถูกสอนมาก่อนหน้าจากชุดข้อมูล ImageNet ซึ่งถูกตรึงในวิธีการสอนแบบ transfer learning และถูกปรับจนได้เมื่อใช้วิธีการสอนแบบ fine tune)

การประเมินเวลาการทำนาย mask ของโมเดลที่ถูกสอนด้วยวิธี fine tune และโมเดลที่สอนด้วยวิธี transfer learning ใกล้เคียงกันทั้งสองวิธี ในทั้ง 2 สถาปัตยกรรมโครงข่าย และแต่ละโมเดล backbone โดยพบว่าโมเดลที่ใช้สถาปัตยกรรมโครงข่าย U-Net สามารถทำนาย Mask ได้เร็วกว่าโมเดลที่ใช้สถาปัตยกรรมโครงข่าย FPN เล็กน้อย ยกเว้นโมเดลที่ใช้รูปแบบโมเดล backbone แบบ SE-ResNet50 พบว่าโมเดลที่ใช้สถาปัตยกรรมโครงข่าย FPN ทำนาย Mask ได้เร็วกว่าโมเดลที่

ใช้สถาปัตยกรรม U-Net แต่เนื่องจากการศึกษานี้ทำการศึกษาผ่าน Virtual Machines บนระบบออนไลน์ ดังนั้นอาจมีความคลาดเคลื่อนของการประเมินเวลาจากการประมวลผลผ่านเครือข่ายอินเทอร์เน็ตได้

ข้อจำกัด

การศึกษานี้ทำการศึกษาจากชุดข้อมูลภาพเอกซเรย์คอมพิวเตอร์และข้อมูลการกำหนดขอบเขตอวัยวะที่มีความเสี่ยงจากฐานข้อมูลสาธารณะออนไลน์ the cancer image archive ซึ่งเป็นชุดข้อมูลของกลุ่มประชากรในประเทศแคนาดา ซึ่งมีลักษณะทางกายวิภาคศาสตร์ที่แตกต่างจากทวีปอื่น เช่น เอเชีย แอฟริกา เป็นต้น ดังนั้นจึงมีข้อจำกัดสำหรับการใช้งานโมเดลนี้เพื่อทำนาย mask บนชุดข้อมูลภาพ CT จากผู้ป่วยในทวีปอื่น ๆ อาจทำให้ผลการทำนายแตกต่างออกไปได้

โมเดลจากการศึกษานี้เป็นโมเดลที่ถูกพัฒนาบนพื้นฐานข้อมูลภาพ 2 มิติ ทำให้ภาพการทำนาย mask จะได้ผลการทำนายในลักษณะ 2 มิติ

สภาพแวดล้อมของการดำเนินการศึกษา ทำการศึกษาผ่านการใช้งาน Kaggle virtual machine ซึ่งทำงาน และประมวลผลผ่านเครือข่ายอินเทอร์เน็ต ดังนั้นความแม่นยำของการประเมินผลด้านเวลาการทำงานของโมเดล อาจเป็นข้อจำกัดของการศึกษานี้

ข้อเสนอแนะ

ข้อเสนอแนะแนวทางการปรับปรุงประสิทธิภาพการทำนายของโมเดลเพื่อลดการทำนายผิดพลาดหรือคลาดเคลื่อน รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง brain stem เสนอแนะให้เพิ่ม label ส่วนโดยรอบ brain stem เช่น การ label ส่วนเนื้อสมอง เพื่อให้โมเดลสามารถเรียนรู้คุณลักษณะภาพที่ใกล้เคียง brain stem แต่ไม่ใช่ส่วนของ brain stem

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right eye globes เสนอแนะให้เพิ่มชุดข้อมูลสอนสำหรับส่วน eye globes จะสามารถช่วยลดความคลาดเคลื่อนได้ เพราะในการศึกษานี้มีชุดข้อมูลสอนจำกัด ทำให้มีความถี่ของข้อมูล mask ของ eye globes ค่อนข้างน้อย

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง mandible เสนอแนะให้ label กระดูกชนิดอื่นบริเวณ TM joint ในชุดข้อมูลสำหรับการสอนโมเดล เพื่อให้โมเดลเรียนรู้รูปแบบของกระดูกชนิดอื่น ๆ ในบริเวณดังกล่าวนี้ได้

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง left and right parotid glands เสนอแนะให้มีการ label อวัยวะส่วนอื่นที่มีรูปร่างและตำแหน่งใกล้เคียงกับ parotid glands เพื่อให้โมเดลเรียนรู้ความแตกต่างได้ รวมถึงการเลือกใช้เทคนิคการสร้างภาพ หรือการประมวลผลภาพอื่นที่ทำให้เห็นความแตกต่างของเนื้อเยื่อได้ชัดเจนขึ้นใช้เป็นข้อมูลภาพนำเข้าสำหรับสอนโมเดล

รูปแบบการทำนายผิดพลาดในอวัยวะที่มีความเสี่ยง spinal cord เสนอแนะให้เพิ่มชุดข้อมูลสอนสำหรับส่วนนี้รวมถึงการ label ขอบบนของ spinal cord และขอบล่างของ brain stem ที่ถูกต้องในชุดข้อมูลสอนจะช่วยลดความแปรปรวนของการกำหนดขอบเขตนี้ได้

ข้อเสนอแนะอื่น ๆ การศึกษานี้มีข้อมูลที่ค่อนข้างจำกัดสำหรับการสอนโมเดล จากข้อมูลพบว่ายังมีปัจจัยอื่นที่อาจมีผลต่อความถูกต้องของการทำนาย mask ของโมเดลกล่าวคือ โรคกระแฉะ ศีรษะและลำคอมีตำแหน่งของรอยโรคกระแฉะได้หลายตำแหน่ง เช่น โพรงหลังจมูก คอหอยส่วนปาก คอหอยส่วนล่าง กล่องเสียง เป็นต้น และโรคกระแฉะมีระยะดำเนินการของโรคหลายระยะ (ระยะ I-IV) ปัจจัยเหล่านี้อาจมีผลต่อการกำหนดขอบเขตของอวัยวะที่มีความเสี่ยงได้ หากรอยโรคอยู่ใกล้กับตำแหน่งอวัยวะที่มีความเสี่ยง จากชุดข้อมูลที่ใช้ศึกษานี้ไม่เพียงพอต่อการวิเคราะห์หาความสัมพันธ์ เสนอแนะให้มีการศึกษาความสัมพันธ์ของปัจจัยดังกล่าวนี้ต่อประสิทธิภาพการทำนายของโมเดลกับชุดข้อมูลที่เพียงพอ

การศึกษานี้พัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับชุดข้อมูลภาพ 2 มิติ เสนอแนะให้ศึกษาในรูปแบบโมเดลการเรียนรู้เชิงลึกสำหรับชุดข้อมูลภาพ 3 มิติ นอกจากนี้แล้ว วิธีการสอนโมเดลด้วยวิธี transfer learning ของการศึกษานี้เลือกตรึงพารามิเตอร์ทุกตัวในส่วน encoder เพื่อให้เห็นผลที่แตกต่างระหว่างวิธีการสอนแบบ transfer learning กับ fine tune เสนอแนะให้มีการพัฒนาโมเดลโดยใช้วิธี transfer learning ในการสอนโมเดลและเลือกปรับตรึงค่าพารามิเตอร์บางชั้นของส่วน encoder ที่เหมาะสมต่อประสิทธิภาพการทำนายของโมเดลต่อไป

สรุป

การศึกษานี้ทำการพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ กำหนดศึกษาใน 7 อวัยวะที่มีความเสี่ยงได้แก่ brain stem, left eye globe, right eye globe, mandible, left parotid gland, right parotid gland และ spinal cord โดยพัฒนาบนพื้นฐานรูปแบบสถาปัตยกรรมโครงข่าย U-Net และ FPN ร่วมกับ โมเดล backbone แบบ VGG19, ResNet50, SE-ResNet50 และ InceptionV3 ซึ่งมีค่า pre-train weight สำหรับแต่ละโมเดล backbone และทำการสอนโมเดลด้วยวิธี fine tune และ transfer learning รวมได้โมเดลทั้งหมด 16 โมเดล ประเมินผลโมเดลด้วยการใช้งานโมเดลเพื่อทำนาย mask อวัยวะที่มีความเสี่ยงบนชุดข้อมูลทดสอบ 15 ชุดข้อมูล ประเมินผลจากการคำนวณหาค่าสัมประสิทธิ์ความเหมือน (DSC) ค่าระยะห่าง 95 เปอร์เซ็นไทล์ฮาวดรอป (95%HD) เวลาการสอนโมเดล เวลาทำนายผล mask และประเมินปัจจัยอื่นที่มีผลต่อประสิทธิภาพการทำนายของโมเดลการเรียนรู้เชิงลึก ผลการประเมินพบว่าโมเดลที่ให้การทำนายที่ดีของแต่ละรูปแบบโมเดล backbone แสดงดังตารางที่ 10

ตาราง 10 สรุปผลชนิดของอวัยวะที่มีความเสี่ยงที่มีค่าเฉลี่ย DSC สูงสุด ค่าเฉลี่ย 95%HD ต่ำสุด เวลาสอน และเวลาทำนาย mask ของโมเดลที่ให้ผลการทำนายที่ดีในแต่ละรูปแบบโมเดล backbone (ค่าเฉลี่ย DSC \pm SD หน่วย: ร้อยละ, ค่าเฉลี่ย 95%HD \pm SD หน่วย: มิลลิเมตร, Training time หน่วย: second per epoch, Prediction time หน่วย: second per image)

| | TUVGG19 | FUResNet50 | FFSE-ResNet50 | FUIncV3 |
|-----------------|------------------|-------------------|-------------------|------------------|
| OARs | MAN | EYE_Rt | MAN | MAN |
| Max DSC | 88.70 \pm 2.96 | 86.90 \pm 4.04 | 87.58 \pm 2.56 | 87.72 \pm 2.84 |
| OARs | PRO_Lt | PRO_Rt | PRO_Rt | PRO_Lt |
| MIN DSC | 74.35 \pm 6.40 | 68.10 \pm 17.24 | 70.29 \pm 16.58 | 74.95 \pm 8.93 |
| OARs | EYE_Lt | EYE_Rt | EYE_Lt | EYE_Lt |
| MIN 95%HD | 1.08 \pm 0.22 | 0.98 \pm 0.22 | 1.02 \pm 0.21 | 1.10 \pm 0.23 |
| OARs | PRO_Lt | PRO_Rt | PRO_Lt | PRO_Lt |
| Max 95%HD | 3.98 \pm 2.05 | 4.01 \pm 1.88 | 4.00 \pm 1.10 | 3.26 \pm 1.03 |
| Training time | 480.00 | 781.17 | 1062.00 | 710.40 |
| Prediction time | 1.29 | 1.35 | 1.37 | 1.40 |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3, OARs = Organs at Risk

จากโมเดลที่ให้ผลการทำนายที่ดีในแต่ละรูปแบบโมเดล backbone เมื่อพิจารณาผลค่าเฉลี่ย DSC สูงสุด ค่าเฉลี่ย 95%HD ต่ำสุด ระยะเวลาการสอนโมเดล และระยะเวลาการทำนาย mask โมเดลที่ดีที่สุดของการศึกษานี้เรียงตามลำดับได้แก่ โมเดล TUVGG19 > FUInceptionV3 > FFSE-ResNet50 > FUResNet50

ปัจจัยอื่นนอกจากสถาปัตยกรรมโครงข่าย โมเดล backbone และวิธีการสอนโมเดล ที่มีผลต่อประสิทธิภาพการทำนาย mask ของโมเดลการเรียนรู้เชิงลึกที่พัฒนาขึ้น ได้แก่

1. รูปแบบของข้อมูลสอน ควรเป็นข้อมูลที่มีคุณภาพ เลือกใช้ช่วงค่า window width, window level ที่เหมาะสมจะช่วยให้ขอบเขตของอวัยวะที่มีความเสี่ยงชัดเจนขึ้น มีสิ่งแปลกปลอม (artifacts) น้อย มีขนาด FOV ที่ใกล้เคียงกัน ทั้งในชุดข้อมูลภาพที่ใช้สอน และชุดข้อมูลภาพที่ใช้ทำนาย

2. การ label mask ที่ดีขอบเขตถูกต้องและชัดเจน เช่น บริเวณขอบบนของ spinal cord และขอบล่างของ brain stem รวมถึงการ label อวัยวะอื่นที่มีรูปร่าง ตำแหน่ง ใกล้เคียงกับอวัยวะที่มีความเสี่ยงที่สนใจ เช่น กระดูกชนิดอื่นบริเวณ TM joint กระดูกที่มีลักษณะรูปร่างใกล้เคียงกับ mandible เช่น กระดูก Hyoid เป็นต้น การ label เนื้อสมองเพื่อให้แยกเนื้อสมองที่มีลักษณะกลมรี และตำแหน่งอยู่กลางกะโหลกศีรษะในภาพ CT ออกจาก brain stem ได้อย่างชัดเจน รวมถึงการ label เนื้อเยื่ออื่นที่อยู่รอบ parotid glands

3. ปริมาณข้อมูลสอนควรมีปริมาณมาก และมีการจัดการข้อมูลในแต่ละคลาสให้มีความถี่ที่ใกล้เคียงกัน

ดังนั้นแล้วหากจัดการกับปัจจัยอื่นดังที่กล่าวมานี้ได้ มีแนวโน้มที่จะเพิ่มประสิทธิภาพการทำงานของโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะที่มีความเสี่ยงบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอได้

บรรณานุกรม

1. Bhardwaj A, Di W, Wei J. Deep learning essentials: your hands-on guide to the fundamentals of deep learning and neural network modeling: Packt Publishing; 2018.
2. Nakamoto P. neural networks and deep learning: neural networks and deep learning, deep learning, big data: CreateSpace Independent Publishing Platform; 2018.
3. Adrian R. Deep learning for computer vision with python. 1 st ed: PYIMAGESEARCH; 2017.
4. Xue Y, Xu T, Zhang H, Long LR, Huang X. SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. Neuroinformatics. 2018;16(3-4):383-92.
5. Talo M, Baloglu UB, Yildirim Ö, Rajendra Acharya U. Application of deep transfer learning for automated brain abnormality classification using MR images. Cognitive Systems Research. 2019;54:176-88.
6. Fang L, Jin Y, Huang L, Guo S, Zhao G, Chen X. Iterative fusion convolutional neural networks for classification of optical coherence tomography images. Journal of Visual Communication and Image Representation. 2019;59:327-33.
7. Sarker A, Klein AZ, Mee J, Harik P, Gonzalez-Hernandez G. An interpretable natural language processing system for written medical examination assessment. Journal of biomedical informatics. 2019:103268.
8. Filice R. Deep-learning language-modeling approach for automated, personalized, and iterative radiology-pathology correlation. Journal of the American College of Radiology : JACR. 2019.
9. Liu S, Utriainen D, Chai C, Chen Y, Wang L, Sethi SK, et al. Cerebral microbleed detection using susceptibility weighted Imaging and deep learning. NeuroImage. 2019;198:271-82.
10. Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity

prediction. *Bioinformatics*. 2018;34(17):i821-i9.

11. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science*. 2018;4(1):120-31.

12. Isaksson M, Jalden J, Murphy MJ. On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Medical physics*. 2005;32(12):3801-9.

13. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Medical physics*. 2017;44(2):547-57.

14. Wang Y, Zhao L, Wang M, Song Z. Organ at risk segmentation in head and neck CT images using a two-stage segmentation framework based on 3D U-Net. *IEEE Access*. 2019;7:144591-602.

15. Wu X, Udupa JK, Tong Y, Odhner D, Pednekar GV, Simone CB, 2nd, et al. Auto-contouring via automatic anatomy recognition of organs at risk in head and neck cancer on CT images. *Proceedings of SPIE--the International Society for Optical Engineering*. 2018;10576.

16. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Medical physics*. 2018;45(10):4558-67.

17. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*. 2018;46.

18. Men K, Geng H, Cheng C, Zhong H, Huang M, Fan Y, et al. Technical Note: More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades. *Medical physics*. 2019;46(1):286-92.

19. Tseng HH, Luo Y, Cui S, Chien JT, Ten Haken RK, Naqa IE. Deep reinforcement learning for automated radiation adaptation in lung cancer. *Medical physics*.

2017;44(12):6690-705.

20. Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ. Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Scientific reports*. 2017;7(1):1648.

21. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports*. 2017;7(1):10353.

22. สาขารังสีรักษาและมะเร็งวิทยา ฝ้ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์. สถิติผู้ป่วยที่มารับบริการที่สาขารังสีรักษาและมะเร็งวิทยา 20 อันดับโรค (ราย) ปี 2562 2562 [cited 27 ตุลาคม 2563]. Available from: <https://www.chulacancer.net/service-statistics-inner.php?id=699>.

23. Merlotti A, Alterio D, Vigna-Taglianti R, Muraglia A, Lastrucci L, Manzo R, et al. Technical guidelines for head and neck cancer IMRT on behalf of the Italian association of radiation oncology - head and neck working group. *Radiation oncology*. 2014;9:264.

24. Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. *The Lancet Oncology*. 2010;11(1):85-91.

25. Thariat J, Ramus L, Maingon P, Odin G, Gregoire V, Darcourt V, et al. Dentalmaps: automatic dental delineation for radiotherapy planning in head-and-neck cancer. *International journal of radiation oncology, biology, physics*. 2012;82(5):1858-65.

26. Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. Radiation dose-volume effects of optic nerves and chiasm. *International journal of radiation oncology, biology, physics*. 2010;76(3):S28-S35.

27. Kirkpatrick JP, van der Kogel AJ, Schultheiss TE. Radiation dose-volume effects in the spinal cord *International journal of radiation oncology, biology, physics*. 2010;76(3):S42-S9.

28. van der Veen J, Willems S, Deschuymer S, Robben D, Crijns W, Maes F, et al. Benefits of deep learning for delineation of organs at risk in head and neck cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology*

and Oncology. 2019;138:68-74.

29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. arXiv e-prints [Internet]. 2015 May 01, 2015. Available from: <https://ui.adsabs.harvard.edu/abs/2015arXiv150504597R>.
30. Lin T, Dollár P, Girshick R, He K, Hariharan B, Belongie S, editors. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 21-26 July 2017.
31. Head-Neck-PET-CT [Internet]. Available from: <https://wiki.cancerimagingarchive.net/display/Public/Head-Neck-PET-CT>.
32. TCGA-HNSC [Internet]. [cited Oct 27]. Available from: <https://wiki.cancerimagingarchive.net/display/Public/TCGA-HNSC>.
33. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep learning: MIT Press; 2016.
34. Keras layers API 2020 [cited 2020 Dec 12]. Available from: <https://keras.io/api/layers/>.
35. Odena A, Dumoulin V, Olah C. deconvolution and checkerboard artifacts. Distill. 2016.
36. Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. ICLR 2015;2015.
37. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 27-30 June 2016.
38. Hu J, Shen L, Sun G, editors. squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 18-23 June 2018.
39. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the Inception Architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 27-30 June 2016.

40. Fchollet. Transfer learning & fine-tuning 2020 [updated 2020 May 12; cited 2020 Oct 29]. Available from: https://keras.io/guides/transfer_learning/.
41. The National Cancer Institute. Head and Neck cancers 2017 [updated March 29, 2017; cited 2020 Dec 12]. Available from: <https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet>.
42. Dobbs J, Barrett A. Practical radiotherapy planning [fourth edition]: Hodder Education; 2009.
43. Beyzadeoglu M, Ozyigit G, Ebruli C. Basic radiation oncology: Springer Berlin Heidelberg; 2010.
44. Yakubovskiy P. Segmentation models 2018 [cited 2020 Oct 27]. Available from: <https://segmentation-models.readthedocs.io/en/latest/>.
45. Turgutlu K. dicom-contour 2018 [cited 2020 Oct 27]. Available from: <https://github.com/KeremTurgutlu/dicom-contour>.
46. IMAGENET large scale visual recognition challenge 2012 (ILSVRC2012) [cited 2020 Oct 29]. Available from: <http://image-net.org/challenges/LSVRC/2012/index#data>.
47. Ren X, Xiang L, Nie D, Shao Y, Zhang H, Shen D, et al. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. Medical physics. 2018;45(5):2063-75.
48. Kovacs DG, Rechner LA, Appelt AL, Berthelsen AK, Costa JC, Friberg J, et al. Metal artefact reduction for accurate tumour delineation in radiotherapy. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2018;126(3):479-86.
49. Chen A, Dawant B. A multi-atlas approach for the automatic segmentation of multiple structures in head and neck CT images. The MIDAS Journal. 2016.

ภาคผนวก ก

ผลการประเมินค่า DSC

ตาราง ก1 ผลการประเมินค่าเฉลี่ย DSC ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึก 16 โมเดล (ค่าเฉลี่ย DSC \pm SD)

| Models | organs at risk | | | | | | | |
|---------------|-------------------|-------------------|--------------------|-------------------|-----------------------|------------------------|-------------------|--|
| | brain stem | left eye globe | right eye globe | mandible | left parotid gland | right parotid gland | spinal cord | |
| FUVGG19 | 0.694 \pm 0.084 | 0.770 \pm 0.072 | 0.826 \pm 0.058 | 0.841 \pm 0.037 | 0.661 \pm 0.185 | 0.760 \pm 0.095 | 0.771 \pm 0.083 | |
| TUVGG19 | 0.815 \pm 0.052 | 0.869 \pm 0.040 | 0.864 \pm 0.056 | 0.887 \pm 0.030 | 0.744 \pm 0.064 | 0.759 \pm 0.125 | 0.790 \pm 0.059 | |
| FFVGG19 | 0.809 \pm 0.048 | 0.857 \pm 0.031 | 0.839 \pm 0.032 | 0.844 \pm 0.035 | 0.746 \pm 0.082 | 0.720 \pm 0.134 | 0.792 \pm 0.050 | |
| TFVGG19 | 0.782 \pm 0.051 | 0.862 \pm 0.043 | 0.848 \pm 0.042 | 0.771 \pm 0.030 | 0.656 \pm 0.112 | 0.715 \pm 0.104 | 0.770 \pm 0.067 | |
| FUResNet50 | 0.839 \pm 0.024 | 0.843 \pm 0.057 | 0.869 \pm 0.040 | 0.861 \pm 0.027 | 0.682 \pm 0.169 | 0.681 \pm 0.172 | 0.768 \pm 0.091 | |
| TUResNet50 | 0.779 \pm 0.043 | 0.848 \pm 0.055 | 0.867 \pm 0.055 | 0.838 \pm 0.070 | 0.581 \pm 0.224 | 0.680 \pm 0.084 | 0.701 \pm 0.151 | |
| FFResNet50 | 0.688 \pm 0.078 | 0.738 \pm 0.075 | 0.813 \pm 0.042 | 0.657 \pm 0.045 | 0.464 \pm 0.117 | 0.567 \pm 0.117 | 0.682 \pm 0.164 | |
| TFResNet50 | 0.691 \pm 0.074 | 0.766 \pm 0.050 | 0.815 \pm 0.041 | 0.729 \pm 0.054 | 0.527 \pm 0.086 | 0.575 \pm 0.095 | 0.761 \pm 0.065 | |
| FUSE-ResNet50 | 0.813 \pm 0.038 | 0.637 \pm 0.064 | 0.835 \pm 0.054 | 0.767 \pm 0.033 | 0.765 \pm 0.072 | 0.785 \pm 0.071 | 0.776 \pm 0.060 | |
| TUSE-ResNet50 | 0.788 \pm 0.053 | 0.868 \pm 0.045 | 0.840 \pm 0.070 | 0.855 \pm 0.046 | 0.692 \pm 0.178 | 0.706 \pm 0.168 | 0.775 \pm 0.079 | |

| Models | organs at risk | | | | | | | |
|---------------|----------------|-------------------|--------------------|---------------|-----------------------|------------------------|---------------|--|
| | brain stem | left eye globe | right eye globe | mandible | left parotid gland | right parotid gland | spinal cord | |
| FFSE-ResNet50 | 0.827 ± 0.027 | 0.863 ± 0.039 | 0.844 ± 0.031 | 0.876 ± 0.026 | 0.716 ± 0.134 | 0.703 ± 0.166 | 0.759 ± 0.095 | |
| TFSE-ResNet50 | 0.721 ± 0.078 | 0.834 ± 0.051 | 0.816 ± 0.062 | 0.675 ± 0.041 | 0.683 ± 0.130 | 0.714 ± 0.160 | 0.778 ± 0.059 | |
| FUIncV3 | 0.813 ± 0.041 | 0.865 ± 0.052 | 0.817 ± 0.068 | 0.877 ± 0.028 | 0.750 ± 0.089 | 0.765 ± 0.084 | 0.795 ± 0.060 | |
| TUIncV3 | 0.802 ± 0.043 | 0.842 ± 0.057 | 0.801 ± 0.075 | 0.870 ± 0.028 | 0.720 ± 0.087 | 0.754 ± 0.115 | 0.775 ± 0.076 | |
| FFIncV3 | 0.822 ± 0.037 | 0.853 ± 0.029 | 0.849 ± 0.039 | 0.877 ± 0.027 | 0.717 ± 0.129 | 0.730 ± 0.081 | 0.752 ± 0.099 | |
| TFIncV3 | 0.738 ± 0.068 | 0.853 ± 0.049 | 0.808 ± 0.058 | 0.670 ± 0.054 | 0.681 ± 0.151 | 0.689 ± 0.133 | 0.770 ± 0.062 | |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3

ภาคผนวก ข

ผลการประเมินค่า 95%HD

ตาราง ข1 ผลการประเมินค่าเฉลี่ย 95%HD ของแต่ละอวัยวะที่มีความเสี่ยงจาก mask array ที่ทำนายโดยโมเดลการเรียนรู้เชิงลึก 16 โมเดล (ค่าเฉลี่ย 95%HD \pm SD, หน่วย: มิลลิเมตร)

| Models | organs at risk | | | | | | | |
|---------------|-------------------|-------------------|-------------------|-------------------|--------------------|---------------------|-------------------|--|
| | brain stem | left eye globe | right eye globe | mandible | left parotid gland | right parotid gland | spinal cord | |
| FUVGG19 | 5.355 \pm 3.306 | 1.908 \pm 1.938 | 1.329 \pm 0.325 | 3.288 \pm 1.241 | 4.656 \pm 2.215 | 3.008 \pm 0.960 | 1.209 \pm 0.281 | |
| TUVGG19 | 1.764 \pm 0.486 | 1.079 \pm 0.221 | 1.072 \pm 0.242 | 3.452 \pm 1.558 | 3.978 \pm 2.051 | 3.365 \pm 1.357 | 1.241 \pm 0.200 | |
| FFVGG19 | 1.952 \pm 0.320 | 1.075 \pm 0.212 | 1.157 \pm 0.243 | 3.646 \pm 1.362 | 4.024 \pm 2.205 | 3.766 \pm 1.557 | 1.612 \pm 0.679 | |
| TFVGG19 | 1.915 \pm 0.399 | 1.089 \pm 0.277 | 1.087 \pm 0.298 | 8.157 \pm 3.416 | 3.770 \pm 0.919 | 3.400 \pm 1.143 | 1.371 \pm 0.313 | |
| FUResNet50 | 1.544 \pm 0.274 | 1.116 \pm 0.250 | 0.979 \pm 0.217 | 3.563 \pm 2.574 | 3.659 \pm 1.155 | 4.012 \pm 1.880 | 1.347 \pm 0.377 | |
| TUResNet50 | 2.116 \pm 0.363 | 1.068 \pm 0.226 | 1.027 \pm 0.208 | 3.895 \pm 1.877 | 4.632 \pm 1.606 | 4.879 \pm 2.802 | 1.520 \pm 0.535 | |
| FFResNet50 | 2.885 \pm 1.056 | 1.568 \pm 0.620 | 1.275 \pm 0.399 | 8.854 \pm 3.068 | 13.85 \pm 6.233 | 4.903 \pm 1.796 | 1.481 \pm 0.483 | |
| TFResNet50 | 2.924 \pm 0.412 | 1.718 \pm 0.600 | 1.452 \pm 0.308 | 5.533 \pm 1.916 | 5.831 \pm 1.517 | 5.011 \pm 1.826 | 1.389 \pm 0.253 | |
| FUSE-ResNet50 | 1.959 \pm 0.507 | 5.796 \pm 5.553 | 1.302 \pm 0.348 | 7.351 \pm 4.145 | 3.006 \pm 0.607 | 2.992 \pm 0.901 | 5.545 \pm 5.681 | |
| TUSE-ResNet50 | 2.342 \pm 0.604 | 1.159 \pm 0.281 | 2.052 \pm 1.631 | 2.978 \pm 1.396 | 3.626 \pm 1.308 | 3.424 \pm 1.190 | 1.286 \pm 0.238 | |

| Models | organs at risk | | | | | | | |
|---------------|----------------|-------------------|--------------------|---------------|-----------------------|------------------------|---------------|--|
| | brain stem | left eye globe | right eye globe | mandible | left parotid gland | right parotid gland | spinal cord | |
| FFSE-ResNet50 | 1.775 ± 0.402 | 1.016 ± 0.206 | 1.110 ± 0.247 | 3.303 ± 1.603 | 4.001 ± 1.096 | 3.657 ± 1.516 | 1.325 ± 0.350 | |
| TFSE-ResNet50 | 2.874 ± 0.529 | 1.174 ± 0.265 | 1.207 ± 0.311 | 9.181 ± 3.423 | 4.448 ± 1.823 | 3.357 ± 1.173 | 1.365 ± 0.215 | |
| FUIncV3 | 1.932 ± 0.672 | 1.101 ± 0.234 | 1.095 ± 0.239 | 2.928 ± 1.130 | 3.256 ± 1.025 | 3.365 ± 1.909 | 1.441 ± 0.799 | |
| TUIncV3 | 2.127 ± 0.500 | 1.330 ± 0.742 | 1.182 ± 0.266 | 3.004 ± 1.384 | 3.696 ± 1.233 | 3.183 ± 1.041 | 1.283 ± 0.334 | |
| FFIncV3 | 1.745 ± 0.282 | 1.262 ± 0.256 | 1.234 ± 0.271 | 2.522 ± 1.076 | 3.578 ± 0.974 | 3.395 ± 0.666 | 1.283 ± 0.313 | |
| TFIncV3 | 2.753 ± 0.419 | 1.186 ± 0.206 | 1.644 ± 1.136 | 8.289 ± 3.093 | 3.753 ± 1.024 | 3.818 ± 1.187 | 1.525 ± 0.412 | |

หมายเหตุ: อักษรตัวแรกของชื่อโมเดลแทนชื่อวิธีการสอนโมเดล กำหนดให้ F แทนวิธี fine tune และ T แทนวิธี transfer learning อักษรตัวที่ 2 แทนชื่อสถาปัตยกรรมโครงข่าย กำหนดให้ U แทนสถาปัตยกรรม U-Net และ F แทนสถาปัตยกรรม FPN อักษรอื่น ๆ ที่เหลืออยู่แทนชื่อโมเดล backbone ได้แก่ VGG19, ResNet50, SE-ResNet50 และ IncV3 = InceptionV3