การพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ

Deep learning model development for auto-delineation of organ at risks on CT simulation images in head and neck cancer

รัฐพงษ์ นันตาแพร่[1], ฐิติพงศ์ แก้วเหล็ก[1,2]

[1]สาขาฟิสิกส์การแพทย์ ภาควิชารังสีเทคนิค คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร

[2]ภาควิชารังสีเทคนิค คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร

นิพนธ์ผู้ประสานงาน

รัฐพงษ์ นันตาแพร่

สาขาฟิสิกส์การแพทย์ ภาควิชารังสีเทคนิค คณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร อ.เมือง จ.พิษณุโลก 65000

อีเมล: rutthapongn61@nu.ac.th

Rutthapong Nantaprae[1], Titipong Kaewlek[1,2]

[1]Medical physics program, Department of radiological technology, Faculty of allied health sciences, Naresuan University

[2]Department of radiological technology, Faculty of allied health sciences, Naresuan University

Corresponding author

Rutthapong Nantaprae

Medical physics program, Department of radiological technology, Faculty of allied health sciences, Naresuan University, Muang, Phitsanulok, Thailand 65000

E-mail: rutthapongn61@nu.ac.th

Submitted: Oct 31, 2020

Revised: Jan 5, 2020

Accepted: Feb 2, 2020

บทคัดย่อ

หลักการและเหตุผล: ในผู้ป่วยมะเร็งศีรษะและลำคอที่รักษาด้วยการฉายรังสี การกำหนดขอบเขตอวัยวะข้างเคียงปกติมีความสำคัญต่อการวางแผนการรักษา เนื่องจากอวัยวะข้างเคียงปกตินี้อาจสูญเสียการทำงาน หากได้รับปริมาณรังสีสูงเกินขีดจำกัด โดยปกติการกำหนดขอบเขตถูกกำหนดโดยแพทย์รังสีรักษาแต่มีข้อจำกัดเช่น ใช้เวลาการกำหนดขอบเขตนาน มีความแปรปรวนภายในและระหว่างแพทย์ผู้กำหนดขอบเขต รวมถึงประสบการณ์ของแพทย์

วัตถุประสงค์: เพื่อพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติบนภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษาผู้ป่วยมะเร็งศีรษะและลำคอ และประเมินโมเดลที่พัฒนาขึ้น

วัสดุและวิธีการ: ศึกษาโดยใช้ข้อมูลภาพเอกซเรย์คอมพิวเตอร์จำลองการรักษา และข้อมูลการกำหนดขอบเขตอวัยวะข้างเคียงปกติเป็นข้อมูลสอนโมเดล ใช้รูปแบบสถาปัตยกรรมโครงข่ายยูเน็ตและฟีเจอร์พีระมิดเน็ตเวิร์ค (เอฟพีเอ็น) ร่วมกับโมเดลแบคโบนแบบวีจีจี19 สอนโมเดลด้วยวิธีปรับจูน และวิธีการถ่ายทอดการเรียนรู้ ประเมินโมเดลด้วยค่าสัมประสิทธิ์ความเหมือน (ดีเอสซี) ค่าระยะห่าง 95 เปอร์เซ็นต์ไทล์ฮาวดรอฟ (95% เอชดี) ประเมินเวลาฝึก และเวลาทำนายมาสก์

ผลการศึกษา: โมเดล ทียูวีจีจี19 และ เอฟเอฟวีจีจี19 ให้ผลค่าดีเอสซีของแต่ละอวัยวะข้างเคียงปกติมากกว่า 0.80 ยกเว้นในต่อมน้ำลายข้างกกหูทั้งสองข้าง และไขสันหลังที่ให้ค่าดีเอสซีมากกว่า 0.72 ผลการประเมินระยะ 95% เอสดี ของแต่ละอวัยวะข้างเคียงปกติมีค่าน้อยกว่า 2 มิลลิเมตร ยกเว้นในขากรรไกรล่างและต่อมน้ำลายข้างกกหูทั้งสองข้างมีค่าน้อยกว่า 4 มิลลิเมตร เวลาเฉลี่ยในการทำนายมาสก์ เท่ากับ 1.286 และ 1.534 วินาทีต่อภาพ สำหรับโมเดลทียูวีจีจี 19 และเอฟเอฟวีจีจี 19 ตามลำดับ

ข้อสรุป: โมเดลที่พัฒนาขึ้นสามารถให้ผลการทำนายมาสก์ที่ดี ประเมินด้วยค่าดีเอสซี 95% เอชดี ได้ใกล้เคียงกับการศึกษาอื่นก่อนหน้า และสามารถทำนายมาสก์ได้อย่างรวดเร็ว

คำสำคัญ: สถาปัตยกรรมโครงข่ายยูเน็ต สถาปัตยกรรมโครงข่ายฟีเจอร์พีระมิดเน็ตเวิร์ค (เอฟพีเอ็น) การถ่ายทอดการเรียนรู้และการปรับจูน

Abstract

Backgrounds: The organs at risk (OARs) delineation process has an important role in treatment planning for head and neck radiation therapy. Overdose to these OARs may cause long term functional loss. In clinic practice, OARs are usually delineated by Radiation Oncologist (RO). Nevertheless, it takes a long time and depends on RO experience. Moreover, intra-inter observer variation is observed.

Objective: To develop a deep learning model for auto-delineation of OARs on head and neck cancer patients computed tomography images and evaluating the performance of the model.

Materials and Methods: Head and neck computed tomography simulation images and DICOM structure radiation therapy were used for training deep learning models. The models were developed by using U-Net architecture and Features Pyramid Networks (FPN) architecture. The architectures were modified by using the VGG19 backbone. Fine tune method and Transfer

learning method were used to train deep learning models. The models were evaluated by Dice Similarity Coefficient (DSC), 95 percentile Hausdorff Distance (95%HD), the time of model training, and the time of model auto-predictions.

Results: The averages DSC results of the TUVGG19 model and FFVGG19 were more than 0.80 for all the OARs, except DSC of both parotid glands and spinal cord were more than 0.72. The averages 95% HD were less than 2 millimeters for all the OARs, except 95% HD of mandible and both parotid glands were not more than 4 millimeters. The average times to predict mask were 1.286 and 1.534 seconds per image for the TUVGG19 model and FFVGG19 model, respectively.

Conclusion: Deep learning models for auto-delineation of OARs have been developed. They showed good prediction performance. The results were evaluated by DSC value and 95%HD likewise the value of previous studies and fast predictions mask.

Keywords: U-Net architecture, Features Pyramid Network (FPN) architecture, Transfer learning and fine tune

J Thai Assoc Radiat Oncol 2021; 27(1): R12-R28

บทนำ

ปัจจุบันมีแนวทางรักษามะเร็งหลายแนวทาง การรักษาด้วยการฉายรังสีเป็นหนึ่งในวิธีการรักษาที่ได้รับการยอมรับ สำหรับอวัยวะบริเวณที่มีความซับซ้อนทางกายวิภาคศาสตร์ เช่น บริเวณศีรษะและลำคอ เนื่องจากวิธีการรักษาอื่นเช่น การผ่าตัด ทำได้ยากและอาจมีผลกระทบหลังการรักษามากกว่า บริเวณศีรษะและลำคอมีตำแหน่งที่จะเกิดโรคมะเร็งได้หลายตำแหน่ง เช่น มะเร็งหลังโพรงจมูก หลอดอาหาร กล่องเสียง ต่อมไทรอยด์ ลิ้น และต่อมน้ำลายข้างกกหู ซึ่งโรคมะเร็งชนิดต่าง ๆ เหล่านี้พบได้มากในผู้ป่วยมะเร็งชาวไทย จากสถิติผู้ป่วยมะเร็งที่เข้ารับบริการที่สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์ ปี 2562[1] พบว่ามีจำนวนผู้ป่วยมะเร็งชนิดต่างๆ ที่กล่าวข้างต้นรวมมากถึง 453 ราย

กระบวนการรักษาผู้ป่วยมะเร็งด้วยการฉายรังสีประกอบด้วยหลายขั้นตอน[2] ได้แก่ การจัดท่าจำลองการรักษา การกำหนดขอบเขตเป้าหมายและอวัยวะข้างเคียงปกติ การวางแผนการรักษาและการกำหนดปริมาณรังสี การประกันคุณภาพก่อนการฉายรังสี การฉายรังสี และการติดตามการรักษา ซึ่งแต่ละกระบวนการดังกล่าวมานั้นมีความเกี่ยวข้องกัน หากมีความคลาดเคลื่อนของกระบวนใดกระบวนการหนึ่งจะส่งผลต่อกระบวนการอื่นของการรักษา เช่น การฉายรังสีเพื่อทำลายเซลล์มะเร็งที่บริเวณศีรษะและลำคอหากมีการกำหนดขอบเขตเป้าหมายและอวัยวะข้างเคียงปกติที่ใหญ่หรือเล็กกว่าขนาดจริง จะส่งผลกระทบต่อการวางแผนการรักษาและการกำหนดปริมาณรังสี ซึ่งอาจส่งผลให้ก้อนมะเร็งอาจได้รับปริมาณรังสีน้อยกว่าที่กำหนด ทำให้ไม่สามารถทำลายเซลล์มะเร็งได้ หรืออวัยวะข้างเคียงปกติอาจจะได้รับ

ปริมาณรังสีเกินขีดจำกัดได้ ซึ่งส่งผลให้อวัยวะข้างเคียงปกติสูญเสียการทำงานปกติได้ มีผลกระทบต่อคุณภาพชีวิตของผู้ป่วยหลังได้รับการรักษามะเร็งด้วยวิธีการฉายรังสี เช่น การเกิดภาวะน้ำลายแห้ง (xerostomia)[3] การตายของกระดูกขากรรไกรหลังการฉายรังสี (post extraction osteoradionecrosis)[4] การเกิดความผิดปกติต่อดวงตาและระบบประสาทตา[5] การเกิดอาการผิดปกติจากการสูญเสียการทำงานของก้านสองและไขสันหลัง[6] ดังนั้นงานวิจัยนี้จึงมุ่งให้ความสนใจกับขั้นตอนของการกำหนดขอบเขตของอวัยวะข้างเคียงปกติ

การกำหนดขอบเขตอวัยวะข้างเคียงปกติในปัจจุบันถูกกำหนดโดยแพทย์รังสีรักษา แต่มีข้อจำกัดคือ ขั้นตอนการกำหนดขอบเขตใช้ระยะเวลานาน มีความแปรปรวนระหว่างและภายในแพทย์ผู้กำหนดขอบเขต และมีปัจจัยด้านประสบการของแพทย์ร่วมด้วย[7] มีหลายการศึกษาก่อนหน้าได้พัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตของอวัยวะข้างเคียงปกติบริเวณศีรษะและลำคอ โดยมีเป้าหมายเพื่อลดข้อจำกัดในการกำหนดขอบเขตที่กล่าวข้างต้น จากการทบทวนวรรณกรรม การศึกษาของ Wu X. และคณะ[8] ศึกษาการกำหนดขอบเขตอวัยวะข้างเคียงปกติอัตโนมัติด้วยวิธีการ Automatic Anatomy Recognition สร้างโมเดลการกำหนดขอบเขตหลายวัตถุบนภาพเอกซเรย์คอมพิวเตอร์สำหรับการวางแผนการรักษาบริเวณศีรษะและลำคออัตโนมัติ การศึกษาของ Ibragimov B. และ Xing L.[9] ศึกษาการแบ่งส่วนอวัยวะข้างเคียงปกติในภาพเอกซเรย์คอมพิวเตอร์บริเวณศีรษะและลำคอ โดยใช้โครงข่ายประสาทเทียมคอนโวลูชันในการตรวจจับตำแหน่งลักษณะภาพจากรูปแบบความเข้มรอบ Voxel ของอวัยวะข้างเคียงปกติที่สนใจบนภาพเอกซเรย์คอมพิวเตอร์ (Computed Tomography: CT) การศึกษาของ Zhu W. และคณะ[10] พัฒนาโมเดล Anatomy Net สำหรับการแบ่งส่วนปริมาตรทางกายวิภาคศาสตร์บริเวณศีรษะและลำคออัตโนมัติ กำหนดขอบเขตอัตโนมัติของปริมาตรอวัยวะข้างเคียงปกติได้ 9 อวัยวะ โดยใช้โครงสร้างสถาปัตยกรรมโครงข่าย 3D U-net เป็นโครงข่ายหลักของโมเดลและเพิ่มการปรับปรุงส่วนการเข้ารหัสและการถอดรหัสของโมเดล และใช้ฟังก์ชันการสูญเสียแบบใหม่ จากที่กล่าวมานี้แสดงให้เห็นถึงความสามารถของโมเดลการเรียนรู้เชิงลึกที่นำมาช่วยลดข้อจำกัดในขั้นตอนกำหนดขอบเขตอวัยวะข้างเคียงปกติได้ และมีรูปแบบสถาปัตยกรรมโครงข่ายอีกหลายรูปแบบที่น่าสนใจรวมถึงรูปแบบวิธีการสอนโมเดลการเรียนรู้เชิงลึก การศึกษานี้จึงมีวัตถุประสงค์เพื่อพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติบนภาพเอซเรย์คอมพิวเตอร์จำลองการรักษาที่ใช้รูปแบบสถาปัตยกรรมโครงข่าย U-net[11] และ Features Pyramid Networks (FPN)[12] และประเมินโมเดลที่พัฒนาขึ้น

วัสดุและวิธีการศึกษา

วัสดุอุปกรณ์

เครื่องมือสำหรับการศึกษานี้ใช้งาน Kaggle Virtual Machines ที่มีลักษณะการทำงานเป็น Jupyter notebooks สำหรับเขียนโปรแกรม ใช้งาน segmentation model library[13] บนพื้นฐานโครงสร้าง Keras (tensor flow) สำหรับการสร้างรูปสถาปัตยกรรมโครงข่าย และใช้งาน Dicom-contour library[14] สำหรับการแปลงภาพ DICOM และข้อมูลการกำหนดขอบเขตบนภาพ DICOM ให้อยู่ในรูปแบบอาร์เรย์

ข้อมูลภาพเอกซเรย์คอมพิวเตอร์วางแผนการรักษาประกอบด้วยภาพ CT และข้อมูลการกำหนดขอบเขตอวัยวะข้างเคียงปกติจากฐานข้อมูลสาธารณะออนไลน์ The Cancer Image Achieve ในชุดข้อมูล Head-Neck-PET-CT[15] จาก 4 สถาบันที่แตกต่างกันใน Québec จำนวน 101 ราย และจากชุดข้อมูล TCGA-HNSC[16] จำนวน 1 ราย

วิธีการ

การศึกษานี้ดำเนินการพัฒนาโมเดลการเรียนเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะ

ข้างเคียงปกติบนภาพเอกซเรย์คอมพิวเตอร์วางแผนการรักษาด้วยสถาปัตยกรรมโครงข่าย U-Net และ FPN กำหนดศึกษาใน 7 อวัยวะข้างเคียงปกติ ได้แก่ brain stem, left eye globe, right eye globe, mandible, left parotid gland, right parotid gland และ spinal cord การดำเนินงานประกอบด้วยขั้นตอนการเตรียมข้อมูล การพัฒนาโมเดล การทดสอบและประเมินผล มีรายละเอียด ดังนี้

1. เตรียมข้อมูล image และ mask

1.1 การเตรียมข้อมูลภาพ CT

เนื่องจากภาพ CT เป็นภาพ DICOM ที่ข้อมูลของแต่ละ pixel อยู่รูปแบบ Hounsfield Unit (HU) ซึ่งเป็นค่าข้อมูลที่มีช่วงค่อนข้างกว้าง (-1000, 1000 HU) แต่อวัยวะข้างเคียงปกติที่สนใจศึกษามีความแตกต่างของเนื้อเยื่อกับบริเวณโดยรอบน้อย จึงกำหนดข้อมูลภาพ CT นำเข้าในรูปแบบของ window ที่เหมาะสม และเพื่อให้สามารถแยกขอบเขตอวัยวะข้างเคียงปกติที่ชัดเจนขึ้น กำหนดให้ภาพ input มีขนาด 512, 512, 3 โดยแต่ละ channel คือภาพ 1 window ได้แก่ brain window, soft tissue window และ bone window

1.2 การเตรียมข้อมูล mask

นำข้อมูลจากชุดข้อมูล DICOM RT Structure มาแยกข้อมูลขอบเขตของอวัยวะข้างเคียงปกติที่สนใจศึกษาให้อยู่ในรูป mask array โดยใช้ DICOM-contour library จากนั้นแปลงเป็นภาพลักษณะ binary image สำหรับแต่ละอวัยวะข้างเคียงปกติ แล้วทำการรวมภาพลักษณะ binary image เป็น 1 ชุดภาพ มีขนาดเป็น 512, 512, 7 กำหนดให้แต่ละ channel เป็นภาพ binary image ของแต่ละอวัยวะข้างเคียงปกติที่สนใจ

1.3 การแบ่งชุดข้อมูล

ข้อมูลที่เตรียมจากข้อ 1.1 และ 1.2 จะถูกแบ่งเป็น 3 ชุดข้อมูล ได้แก่ 1. ชุดข้อมูลสำหรับการสอนโมเดล จำนวน 75 ราย (เป็นภาพ CT และ ภาพ mask จำนวนอย่างละ 7,683 ภาพ) 2. ชุดข้อมูลสำหรับการตรวจสอบโมเดล 12 ราย (เป็นภาพ CT และ ภาพ mask จำนวนอย่างละ 1,136 ภาพ) และ 3. ชุดข้อมูลสำหรับการทดสอบโมเดล 15 ราย (เป็นภาพ CT และ ภาพ mask อย่างละ 1,504 ภาพ) ข้อมูลภาพทั้งหมดแสดงการกระจายความถี่ mask ของข้อมูลแต่ละอวัยวะข้างเคียงปกติแสดงดัง**ภาพที่ 1**

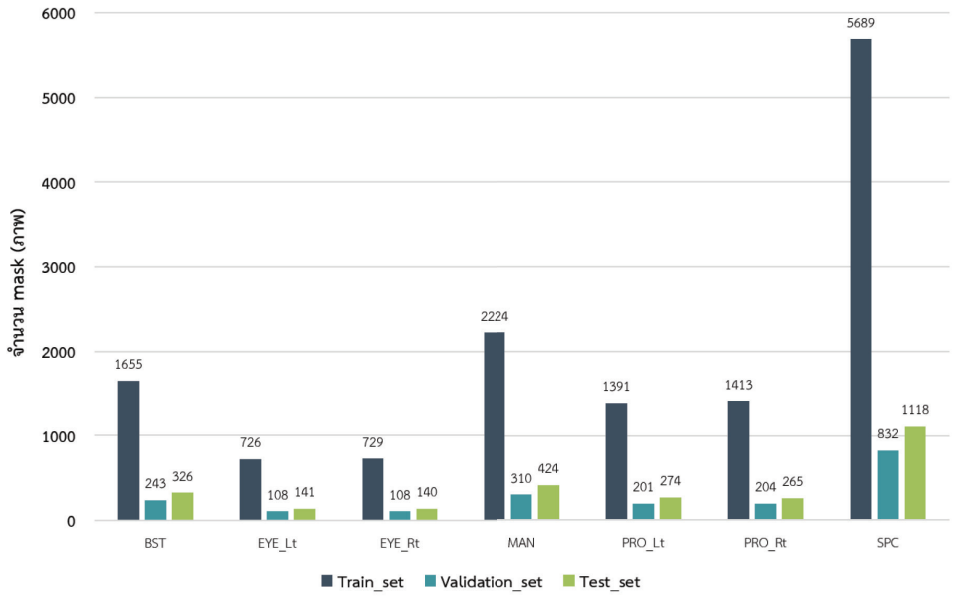
2.การพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะข้างเคียงปกติ

2.1 สถาปัตยกรรมโครงข่าย ทำการสร้างโมเดลการเรียนรู้เชิงลึกที่มีรูปแบบสถาปัตยกรรมโครงข่ายแบบ U-Net และ FPN โดยใช้ segmentation model library เลือกใช้โมเดล backbone แบบ VGG19 สำหรับทั้งสองรูปแบบสถาปัตยกรรมโครงข่าย

2.1.1 สถาปัตยกรรมโครงข่าย U-Net ประกอบด้วย 2 ส่วน ได้แก่ ส่วน down sampling และส่วน up sampling แสดงดัง**ภาพที่ 2**

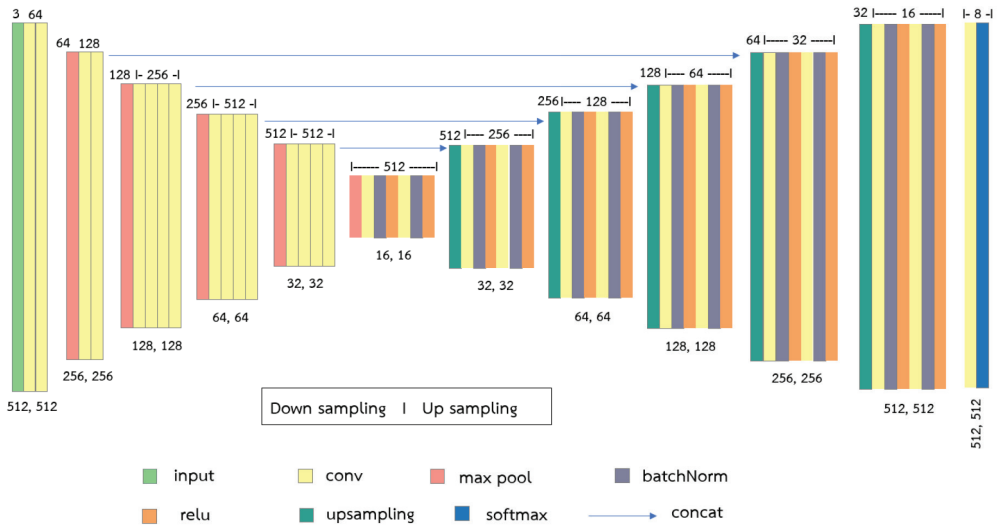
ส่วนของ down sampling ใช้โมเดล backbone แบบ VGG19 ประกอบด้วย 5 block ใน block ที่ 1 และ 2 ประกอบด้วยชั้นคอนโวลูชัน 2 ชั้นเชื่อมต่อกับชั้น pooling 1 ชั้น ใน block ที่ 3 ถึง 5 ประกอบด้วยชั้นคอนโวลูชัน 4 ชั้น ตามด้วยชั้น pooling 1 ชั้น

ส่วนของ up sampling ประกอบด้วย 5 block ที่สัมพันธ์กับส่วน down sampling โดยแต่ละ block ประกอบด้วย ชั้นคอนโวลูชัน ชั้น batch normalization และชั้น activation (Relu) จำนวน block ละ 2 ชุด ตามด้วยชั้น up sampling ยกเว้นใน block 5 ที่ไม่เชื่อมต่อด้วยชั้น up sampling แต่เชื่อมต่อด้วยชั้นคอนโวลูชันและชั้น activation (soft max) การเชื่อมกันในส่วนของ down sampling และ up sampling เชื่อมต่อด้วย block ตรงกลางระหว่างสองส่วน เรียกว่า center block มีการเรียงของชั้นดำเนินการเหมือนกับส่วนของ up sampling ใน block ที่ 1 ถึง 4 และใช้วิธี concatenation ในการเชื่อต่อ feature map ระหว่าง block ที่ 2 ถึง 5 ในส่วน down sampling กับ block ที่ 1 ถึง 4 ในส่วน up sampling



ภาพที่ 1 การกระจายความถี่ mask อวัยวะข้างเคียงปกติของชุดข้อมูลสอน ชุดข้อมูลตรวจสอบ และชุดข้อมูลทดสอบ

คำย่อ: BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord



ภาพที่ 2 รูปแบบสถาปัตยกรรมโครงข่าย U-Net

คำย่อ: conv = convolution layer, max pool = max pooling layer, batchNorm = batch normalization layer, relu = rectified linear activation function layer, upsampling = upsampling layer, softmax = soft max activation function layer, concat = concatenation

2.1.2 สถาปัตยกรรมโครงข่าย FPN ประกอบด้วยการเชื่อมต่อของชั้นดำเนินการ 3 แนวได้แก่ bottom-up (BU), top-down (TD) และแนว lateral I, II (L–I, L–II) การเชื่อมต่อของแนว BU ใช้โมเดล backbone แบบ VGG19 เช่นเดียวกับส่วน down sampling ของสถาปัตยกรรมโครงข่าย U-Net ประกอบด้วย 5 block แสดงดัง**ภาพที่ 3**

การเชื่อมต่อในแนว L–I ใช้ผลลัพธ์จากชั้นคอนโวลูชันสุดท้ายของ block 2 BU ถึง block 5 BU และผลลัพธ์จากชั้น max pooling ของ block 5 BU ถูกนำมาคอนโวลูชันด้วย kernel ขนาด (1, 1) ได้ผลลัพธ์แสดงเป็น block 2 L–I ถึง block 5 L–I สำหรับการคอนโวลูชันของ block 2 BU ถึง block 5 BU ตามลำดับ และผลลัพธ์ที่ได้จากการคอนโคลูชันชั้น max pooling ใน block 5 BU จะถูกใช้เป็นข้อมูลเริ่มต้นสำหรับการดำเนินการในแนวการเชื่อมต่อ TD

การเชื่อมต่อในแนว TD ใช้ผลลัพธ์จากการเชื่อมต่อในแนว L–I ของชั้น max pooling ใน block 5 BU เป็นข้อมูลเริ่มต้นเพื่อทำการ up sampling ผลลัพธ์จากการ up sampling ในชั้นนี้จะถูกรวมเข้ากับผลลัพธ์ block 5 L–I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 5 TD จากนั้นผลลัพธ์นี้จะถูก up sampling และรวมกับผลของ block 4 L–I ด้วยวิธีการ add ผลที่ได้แสดงเป็นผลลัพธ์ของ block 4 TD จากนั้นใช้รูปแบบการเชื่อมต่อลักษณะเดียวกันกับ block 3 L–I กับ ผล up sampling ของ block 4 TD แสดงเป็นผลลัพธ์ของ block 3 TD และ block 2 L–I กับ ผล up sampling ของ block 3 TD แสดงเป็นผลลัพธ์ของ block 2 TD ตามลำดับ

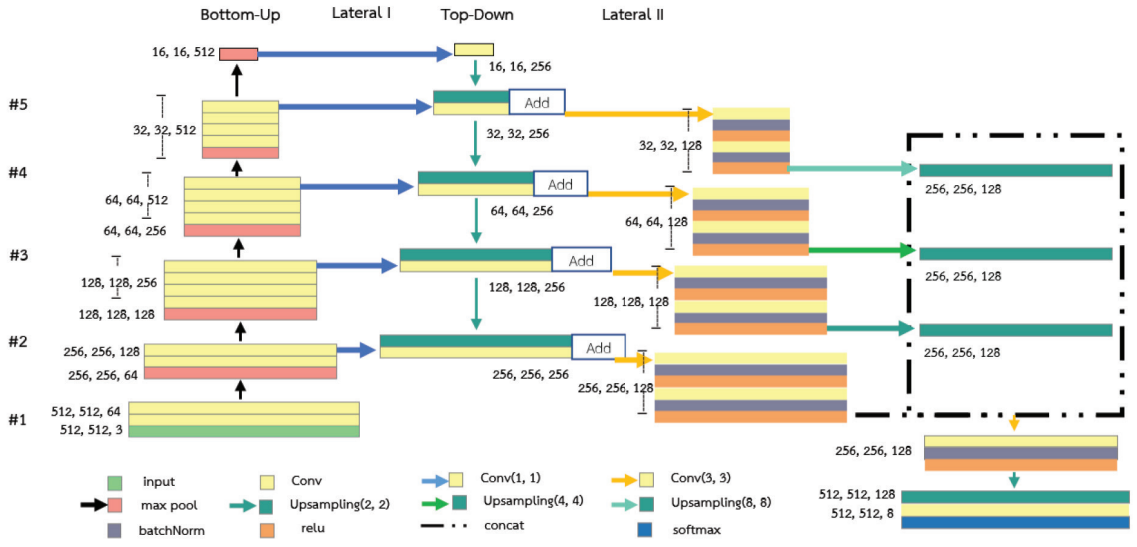
การเชื่อมต่อในแนว L–II ใช้ผลลัพธ์ของ block 2 TD ถึง block 5 TD ถูกดำเนินการด้วยชั้นคอนโวลูชัน ชั้น batch normalization และชั้น activation (Relu) สำหรับแต่ละ block TD ละ 2 ชุด ผลลัพธ์สุดท้ายที่ได้แสดงเป็น block 2 L–II ถึง block 5 L–II โดยผลลัพธ์ block 3 L–II ถึง block 5 L–II จะถูก up sampling ด้วย kernel ขนาด (2, 2), (4, 4), (8, 8) ตามลำดับ ผลลัพธ์จากการดำเนินการ up sampling ถูกเชื่อมผสานเข้ากับผลลัพธ์ block 2 L–II ด้วยวิธีการ concatenation ผลลัพธ์ที่ได้จากการ concatenation จะถูกนำมาดำเนินการด้วยชั้นคอนโวลูชัน ชั้น batch normalization ชั้น activation (Relu) แล้วทำการ up sampling ตามด้วยการคอนโวลูชัน และใช้ชั้น activation (soft max) เพื่อทำนายผลลัพธ์ การเชื่อมต่อของสถาปัตยกรรมฟีเจอร์พีระมิดแสดงดัง**ภาพที่ 3**

2.2 การสอนโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะข้างเคียงปกติบริเวณศีรษะและลำคอ แบ่งออกเป็น 2 วิธี ในแต่ละสถาปัตยกรรมโครงข่าย วิธีที่ 1 สอนโมเดลด้วยวิธี fine tune (ค่าพารามิเตอร์ของโมเดลที่สัมพันธ์กับโมเดล backbone แบบ VGG19 ซึ่งมีค่า weight ที่ถูกสอนด้วยชุดข้อมูล ImageNet 2012 ILSVRC[17] ถูกนำมาใช้เป็นค่าพารามิเตอร์เริ่มต้นของโมเดล) และวิธีที่ 2 สอนโมเดลโดยใช้วิธีการถ่ายทอดการเรียนรู้[18] ด้วยค่าถ่วงน้ำหนัก ImageNet (ค่าพารามิเตอร์ของโมเดลที่สัมพันธ์กับโมเดล backbone แบบ VGG19 ถูกตรึง (freeze) เป็นพารามิเตอร์ของโมเดล) โมเดลสำหรับการศึกษานี้จึงมีด้วยกัน 4 โมเดล ค่าพารามิเตอร์ต่างๆ แสดงดัง**ตารางที่ 1** ขั้นตอนการสอนทำการกำหนดอัตราการเรียนรู้ที่ 0.0001 กำหนด batch size = 4 สำหรับขั้นตอนการสอน (train) และ batch size = 1 ในขั้นตอนการตรวจสอบ (validation) กำหนด epoch = 20 เลือกใช้ Adam optimizer เลือกใช้ Dice Loss + Categorical Focal Loss เป็นฟังก์ชันการสูญเสีย และกำหนดให้หยุดการฝึกเมื่อค่า validation loss ไม่มีการเปลี่ยนแปลงต่อเนื่องกัน 3 ครั้ง

3. การทดสอบและการวัดผล

3.1 การประเมินความถูกต้องการทำงานของโมเดล

ดำเนินการประเมินผลความเหมือนของภาพ mask array จากการกำหนดขอบเขตอวัยวะข้างเคียงปกติจากโมเดลที่พัฒนาขึ้นกับ mask array



ภาพที่ 3 รูปแบบสถาปัตยกรรมโครงข่าย FPN

คำย่อ: conv = convolution layer, conv (1,1) = convolution kernel size 1×1 layer, conv (3,3) = convolution kernel size 3×3 layer, max pool = max pooling layer, batchNorm = batch normalization layer, relu = rectified linear activation function layer, upsamplimg (2,2) = upsampling kernel size 2×2 layer, upsampling (4,4) = upsampling kernel size 4×4 layer, upsamplimg (8,8) = upsampling kernel size 8×8 layer, softmax = soft max activation function layer, concat = concatenation, Add = addition method

ตารางที่ 1 พารามิเตอร์ของโมเดลการเรียนรู้เชิงลึก

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

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19

จากการกำหนดขอบเขตอวัยวะข้างเคียงปกติด้วยการวาดมือในการวางแผนการรักษา ด้วยภาพลักษณะ2 มิติ (2D CT) ของบริเวณศีรษะและลำคอ โดยใช้ชุดข้อมูลทดสอบตามที่กำหนดไว้ในข้อ 1.3

การเปรียบเทียบความเหมือนเชิงรูปร่างถูกคำนวนหาค่า Dice Similarly Coefficient (DSC)[19] แสดงการคำนวณดัง**สมการ 1** และการเปรียบเทียบเพื่อหาความสัมพันธ์ของ mask array ทั้งสองในรูปแบบระยะห่างของขอบ mask ถูกคำนวนด้วยค่า 95 Percentile Hausdorff Distance (95%HD)[19] แสดงการคำนวณดังสมการ 2

$$DSC = \frac{2|A \cap B|}{|A|+|B|} \quad (1)$$

เมื่อ |A| และ |B| เป็นพื้นที่ของ mask ที่ได้จากการทำนายและ mask อ้างอิง และ $|A \cap B|$ เป็นพื้นที่ ที่ซ้อนทับกันของ mask จากการทำนายและ mask อ้างอิง โดยค่า DSC เท่ากับ 1 ถ้า mask ของ 2 ชุด ซ้อนทับกันอย่างสมบูรณ์ และเท่ากับ 0 ถ้าถ้า mask ของ 2 ชุด ไม่ซ้อนทับกัน

$$\vec{d}_{H,\ r}(\mathrm{x},\mathrm{y}) = \mathrm{K}_r\left\{\min_{y \in Y} d(x,y)\right\}, \forall x \in X \quad (2)$$

$$d_{H,\ r}(\mathrm{x},\mathrm{y}) = \frac{\vec{d}_{H,\ r}(\mathrm{x},\mathrm{y}) + \vec{d}_{H,\ r}(\mathrm{y},\mathrm{x})}{2} \quad (3)$$

เมื่อ K_r คือ ค่าเปอร์เซ็นต์ไทล์ที่ใช้ในการคำนวน (95 Percentile) และ $d(x,y)$ คือ ระยะทางระหว่าง จุด x บนเส้นแสดงรูปร่างขอบเขต mask จากการทำนายและจุด y บนเส้นแสดงรูปร่างขอบเขต mask อ้างอิง

3.2 การประเมินเวลา

ดำเนินการประเมินระยะเวลาการฝึกจากการจับเวลาการเริ่มต้นถึงเสร็จสิ้นการฝึกในแต่ละ epoch และประเมินระยะเวลาการทำนายภาพ โดยการนำข้อมูลภาพ CT เข้าสู่โมเดลแล้วให้โมเดลทำนายภาพผลลัพธ์ mask array ของแต่ละโมเดล ทำการจับเวลาด้วย time library แล้วหาผลต่างของระยะเวลาที่ได้จากการบันทึกเวลาเริ่มการทำนายของโมเดล และเวลาการทำนายแล้วเสร็จของโมเดล

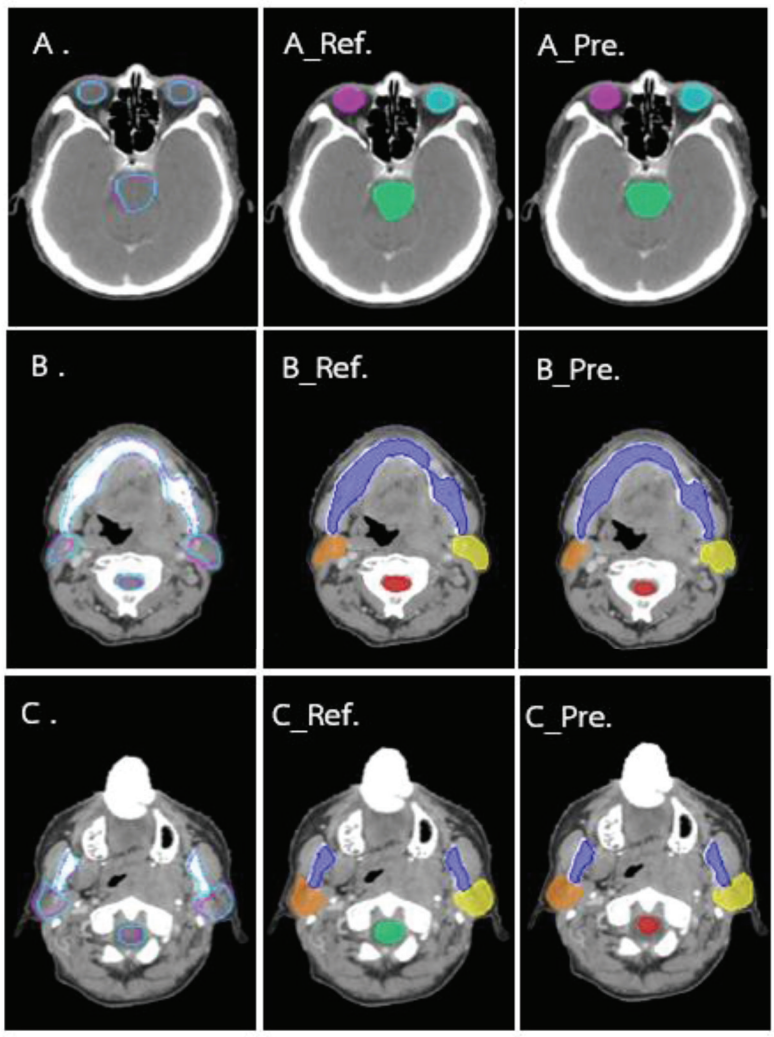
ผลการศึกษา

ผลการทำนาย mask จากโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอวัยวะข้างเคียงปกติบริเวณศีรษะและลำคอที่พัฒนาขึ้นแสดงตัวอย่างผลของการทำนายดัง**ภาพที่ 4** พบว่า เมื่อเทียบกับ mask อ้างอิงในภาพที่ 4A_Ref, 4B_Ref กับภาพ 4A_Pre, 4B_Pre โมเดลสามารถทำนาย mask ได้ดี แต่มีการทำนายผิดพลาดในบางภาพ CT ดังแสดงในภาพเปรียบเทียบระหว่างภาพ 4C_Ref กับภาพที่ 4C_Pre ผลพบการทำนายผิดของโมเดลว่าเป็น spinal cord แต่ในความเป็นจริงแล้วเป็นภาพ mask อ้างอิงของ brain stem นอกจากนี้แล้ว ผลการทำนาย mask ของทั้ง 4 โมเดลบนภาพจำนวน 15 ชุด พบว่ามี 1 ชุดข้อมูลที่ไม่สามารถให้ผลการทำนาย mask ได้ ดังนั้น การศึกษาในส่วนของค่า DSC และ 95%HD ของการศึกษานี้จะคำนวนเฉพาะข้อมูลใน 14 ชุดข้อมูลที่สามารถทำนาย mask ได้เท่านั้น

ผลการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ จากการคำนวณค่า DSC จากภาพการทำนาย 7 อวัยวะข้างเคียงปกติ เปรียบเทียบกับภาพอ้างอิง ได้ผลดังแสดงใน**ภาพที่ 5**

พบว่าโมเดล TUVGG19 ให้ค่าเฉลี่ย DSC สูงที่สุดในอวัยวะข้างเคียงปกติ brain stem, left eye globe, right eye globe และ mandible มีค่าเฉลี่ยของ DSC เท่ากับ 0.815 ± 0.052, 0.869 ± 0.040, 0.864 ± 0.056 และ 0.887 ± 0.030 ตามลำดับ

โมเดล FFVGG19 มีค่าเฉลี่ยของ DSC สูงที่สุดในอวัยวะข้างเคียงปกติ left parotid gland และ spinal cord มีค่าเฉลี่ยของ DSC เท่ากับ 0.746 ± 0.082 และ 0.792 ± 0.050 ตามลำดับ



	DSC
BST	0.9101
EYE_Lt.	0.9087
EYE_Rt.	0.9424

	DSC
MAN	0.9418
PRO_Lt.	0.8379
PRO_Rt.	0.7473
SPC	0.7710

	DSC
MAN	0.9330
PRO_Lt.	0.8872
PRO_Rt.	0.8389
SPC	0.0000

ภาพที่ 4 ตัวอย่างผลการทำนาย mask และค่า DSC จากโมเดล TUVGG19

คำย่อ: ภาพ 4A - 4C แสดงเส้นขอบเขตของ mask อ้างอิง (สีฟ้า) และ mask ที่ได้จากโมเดล (สีชมพู), ภาพ 4A_Ref - 4C_Ref แสดงพื้นที่ mask อ้างอิง, ภาพ4A_Pre - 4C_Pre แสดงพื้นที่ mask จากการทำนายของโมเดล , BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord

โมเดล FUVGG19 มีค่าเฉลี่ย DSC สูงที่สุดในอวัยวะข้างเคียงปกติ right parotid gland มีค่าเฉลี่ยของ DSC เท่ากับ 0.760 ± 0.095

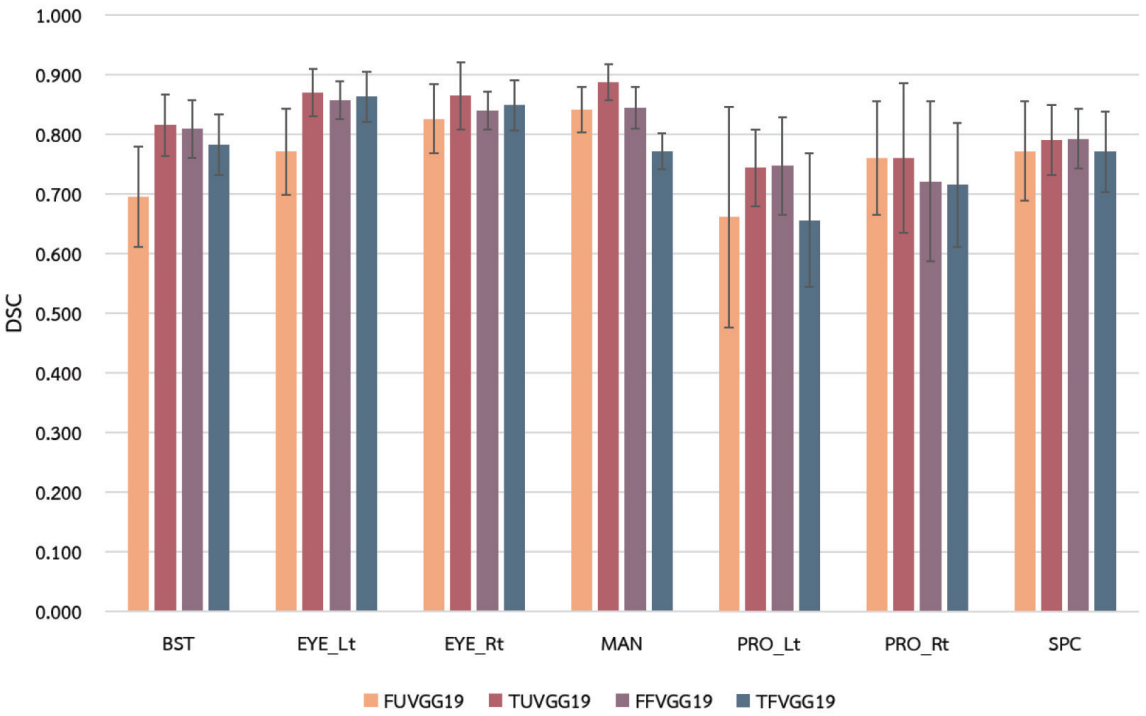
ผลการประเมินระยะห่างของขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ ผลการคำนวณค่า 95%HD แสดงดัง**ภาพที่ 6**

ผลของโมเดล FUVGG19 มีค่าเฉลี่ย 95%HD ต่ำที่สุดในอวัยวะข้างเคียงปกติ mandible, right parotid gland และ spinal cord มีค่าเฉลี่ย 95%HD เท่ากับ 3.288 ± 1.241, 3.008 ± 0.960 และ 1.209 ± 0.281 มิลลิเมตร ตามลำดับ

โมเดล TUVGG19 มีค่าเฉลี่ย 95%HD ต่ำที่สุดในอวัยวะข้างเคียงปกติ brain stem และ right eye globe มีค่าเฉลี่ย 95%HD เท่ากับ 1.764 ± 0.486 และ 1.079 ± 0.242 มิลลิเมตร ตามลำดับ

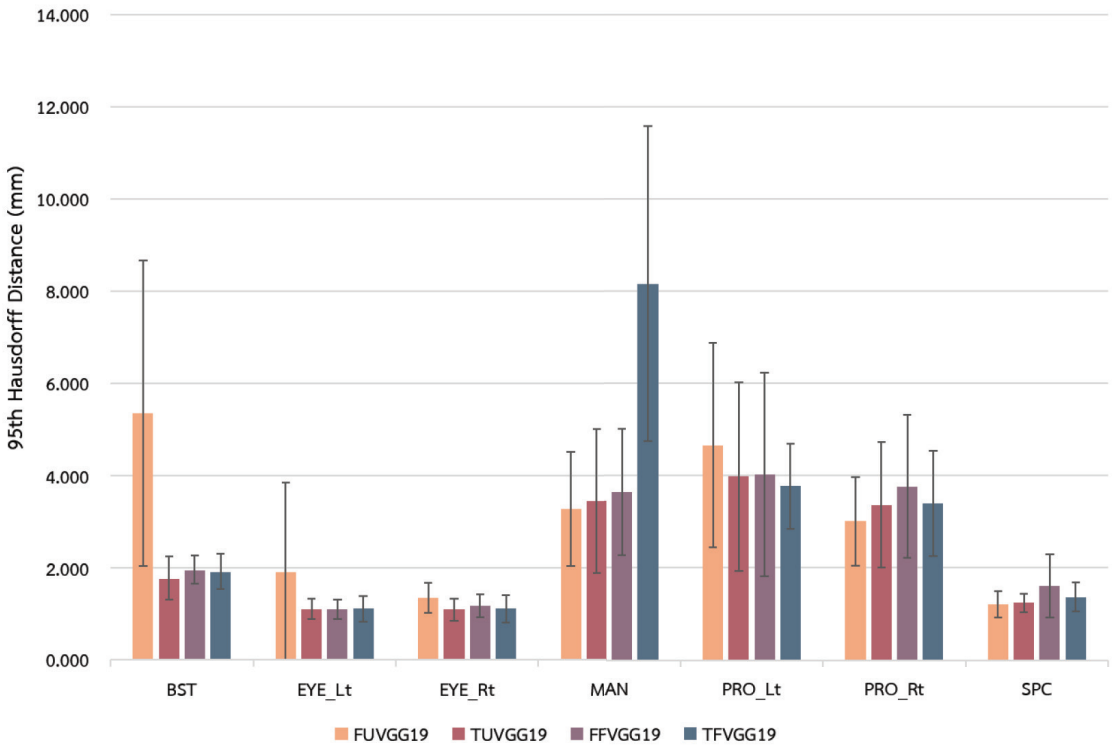
โมเดล FFVGG19 มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะข้างเคียงปกติ left eye globe มีค่าเฉลี่ย 95%HD เท่ากับ 1.075 ± 0.212 มิลลิเมตร

โมเดล TFVGG19 มีค่าเฉลี่ย 95%HD ต่ำสุดในอวัยวะข้างเคียงปกติ left parotid gland ค่าเฉลี่ย 95%HD เท่ากับ 3.770 ± 0.919 มิลลิเมตร



ภาพที่ 5 ค่า DSC เฉลี่ยของแต่ละอวัยวะข้างเคียงปกติที่ถูกทำนายด้วยโมเดลการเรียนรู้เชิงลึก 4 โมเดล

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord



ภาพที่ 6 ค่าเฉลี่ย 95%HD ของแต่ละอวัยวะข้างเคียงปกติที่ถูกทำนายด้วยโมเดลการเรียนรู้เชิงลึก 4 โมเดล

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord

ผลการประเมินเวลาการทำงานของโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติ ข้อมูลเวลาการสอนโมเดลและข้อมูลเวลาการทำนาย mask ของแต่ละโมเดล แสดงดัง**ตารางที่ 2** พบว่าโมเดลที่สอนโดยใช้วิธีการ transfer learning ใช้เวลาน้อยกว่าการสอนโมเดลด้วยวิธีการ fine tune ทั้งใน 2 รูปแบบสถาปัตยกรรมโครงข่าย (U-Net และ FPN) เวลาลดลง 258.4 และ 296 วินาทีต่อ epoch ตามลำดับ

เวลาการทำนาย mask ของโมเดลที่ถูกสอนด้วยวิธี transfer learning และวิธีการ fine tune มีความใกล้เคียงกันทั้งสองวิธีในแต่ละสถาปัตยกรรมโครงข่าย โดยโมเดลที่ใช้สถาปัตยกรรมโครงข่าย U-Net ใช้เวลาในการทำนาย mask น้อยกว่าโมเดลที่ใช้ FPN

บทวิจารณ์

การศึกษานี้ได้ทำการพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติบนภาพเอกซเรย์คอมพิวเตอร์วางแผนการรักษา โดยใช้รูปแบบสถาปัตยกรรมโครงข่าย 2 รูปแบบ ได้แก่ U-Net และ FPN และใช้โมเดล backbone แบบ VGG 19 ร่วมกับ

ตารางที่ 2 เวลาการสอนและเวลาการทำนาย mask ของโมเดล

Models	Training time (second per epoch)	Prediction time (second per image)
FUVGG19	737.2	1.307
TUVGG19	478.8	1.286
FFVGG19	1056	1.534
TFVGG19	760	1.430

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19

การสอน 2 วิธี ได้แก่ วิธี fine tune และ วิธี transfer learning ด้วยค่าถ่วงน้ำหนัก ImageNet

พบว่ามีบางภาพ ให้ผลการทำนาย mask ผิดพลาด จาก mask อ้างอิงที่ถูกกำหนดว่าเป็น brain stem เป็นอวัยวะข้างเคียงปกติ spinal cord ในกรณีนี้จากการศึกษาพบว่ามักเกิดในบริเวณรอยต่อขอบบนของ spinal cord กับขอบล่างของ brain stem มีสาเหตุมาจากชุดข้อมูลสอนซึ่งถูกเก็บมาจากหลายแหล่งข้อมูล ซึ่งการกำหนดขอบเขตส่วนขอบบนและขอบล่างของ spinal cord และ brain stem ไม่ชัดเจน แม้ว่าการปฏิบัติทางคลินิกมีการแนะนำตามคู่มือการกำหนดขอบเขตอวัยวะข้างเคียงปกติอย่างชัดเจนที่ให้กำหนดขอบบนของ spinal cord ที่ตำแหน่ง Tip of dens ของ C2 ซึ่งพบว่ามีความคลาดเคลื่อนของขอบเขตนี้เช่นเดียวกับการศึกษาของ van der Veen J. และคณะ[7] หากมีการเพิ่มชุดข้อมูลสำหรับการสอนในส่วนนี้จะช่วยลดความแปรปรวนของการกำหนดขอบเขตดังกล่าวนี้ได้

จากผลการทดลองพบว่าทั้ง 4 โมเดล มี 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 ของชุดข้อมูลในการสอนโมเดล ด้วยเหตุผลดังกล่าวนี้แสดงให้เห็นว่าขนาดของ FOV เป็นปัจจัยหนึ่งที่ส่งผลต่อการทำนายของโมเดลที่พัฒนาขึ้น โดยแนะนำให้ใช้ทำนายภาพ CT ที่มีขนาด FOV ตั้งแต่ 500 ถึง 600 มิลลิเมตร จึงจะทำให้โมเดลสามารถทำนายผล mask ได้

จากผลการศึกษาการประเมินความเหมือนของ mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ ทั้ง 14 ชุดข้อมูล จากการประเมินค่า DSC พบว่าโมเดลที่สามารถทำนาย mask แล้วให้ค่า DSC สูง คือ โมเดล TUVGG19 และ FFVGG19 ที่ให้ค่า DSC มากกว่า 0.80 ยกเว้นใน parotid ทั้ง 2 ข้างและ spinal cord ที่ให้ค่า DSC มากกว่า 0.72 และมีแนวโน้มผลการทำนายที่ให้ค่า DSC ใกล้เคียงกับผล DSC จากการศึกษาก่อนหน้านี้ ดังแสดงใน**ตารางที่ 3**

จากผลการคำนวณค่า DSC พบว่าโมเดล TUVGG19 และ FFVGG19 สามารถทำนายได้ดีในอวัยวะข้างเคียงปกติที่มีความแตกต่างของเนื้อเยื่อสูงได้ดีกว่าการทำนาย

ตารางที่ 3 ค่าเฉลี่ย DSC จากโมเดล TUVGG19, FFVGG19 และค่า DSC จากการศึกษาก่อนหน้า (หน่วย: ร้อยละ)

Organs	Wu X.[8]	Ibragimov B.[9]	Zhu W.[10]	Chen A.[20]	TUVGG19	FFVGG19
BST	NA	NA	86.65±2.00	80.30	81.50 ± 5.20	80.90 ± 4.80
EYE_Lt	NA	88.40±2.70	NA	NA	86.90 ± 4.00	85.70± 3.10
EYE_Rt	NA	87.70±3.70	NA	NA	86.40 ± 5.60	83.90 ± 3.20
MAN	89.00	89.50±3.60	92.51±2.00	91.70	88.70 ± 3.00	84.40 ± 3.50
PRO_Lt	77.00	76.60±6.10	88.07±2.00	81.10	74.40 ± 6.40	74.70 ± 8.20
PRO_Rt	76.00	77.90±5.40	87.35±4.00	81.40	76.00± 12.50	72.00 ± 3.40
SPC	NA	87.00±3.20	NA	NA	79.00 ± 5.90	79.20 ± 5.00

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord, NA = not available

ในอวัยวะที่มีความแตกต่างของเนื้อเนื้อต่ำสอดคล้องกับผลค่า DSC จากการศึกษาของ Wu X. และคณะ[8] และ Ibragimov B. และ Xing L.[9] แม้ว่าในการศึกษานี้จะใช้วิธีการเตรียมข้อมูลภาพที่กำหนดค่า windows ในชุดข้อมูลภาพ CT นำเข้า โดยพบว่าการเลือกใช้ soft tissue windows ยังไม่สามารถช่วยให้โมเดลแยกขอบเขตของ parotid grands กับเนื้อเยื่อโดยรอบได้ดีเท่าที่ควร การกำหนดค่าข้อมูลภาพ CT นำเข้าด้วย bone windows และ brain windows โมเดลที่พัฒนาขึ้นสามารถแยกขอบเขตอวัยวะข้างเคียงปกติในบริเวณที่สัมพันธ์กับทั้ง 2 windows นี้ให้ผลค่อนข้างดี นอกจากนี้พบว่าโมเดล TUVGG19 ที่ใช้สถาปัตยกรรมโครงข่าย U-Net สอนโมเดลด้วยวิธีการ transfer learning ให้ผลค่า DSC สูงกว่าการสอนด้วยวิธีการ fine tune ในขณะที่โมเดล FFVGG19 ที่ใช้สถาปัตยกรรมโครงข่าย FPN ให้ผลค่า DSC สูงในวิธีการสอนที่ตรงข้ามกับโมเดล TUVGG19 เมื่อพิจารณาความสัมพันธ์ของค่า DSC รูปแบบสถาปัตยกรรมโครงข่ายและวิธีการสอน พบว่าการเลือกใช้รูปแบบสถาปัตยกรรมโครงข่ายและวิธีการสอนที่ทำให้โมเดลมีค่าพารามิเตอร์ที่สามารถสอนได้ในจำนวนที่เหมาะสมจะส่งผลให้โมเดลมีการเรียนรู้ได้ดีและสามารถทำนาย mask ได้ถูกต้อง เช่น โมเดล TUVGG19 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่า โมเดล FUVGG19 ที่มีจำนวนพารามิเตอร์ที่สอนได้มากเกินไป หรือในโมเดล FFVGG19 มีจำนวนพารามิเตอร์ที่สามารถสอนได้เหมาะสมกว่าโมเดล TFVGG19 ที่มีจำนวนพารามิเตอร์ที่สอนได้น้อยเกินไป

จากผลการประเมินระยะห่างขอบเขต mask ที่ทำนายจากชุดข้อมูลภาพ CT กลุ่มทดสอบ ผลการวัดค่า 95%HD พบว่า mask ของอวัยวะข้างเคียงปกติจากโมเดลที่มีผลค่า DSC สูง จะส่งผลให้มีค่า 95%HD ที่ต่ำกว่า โมเดลที่ให้มีค่า DSC ต่ำ โดยค่า 95%HD จากโมเดล TUVGG19 และโมเดล FFVGG19 มีค่าใกล้เคียงกับการศึกษาก่อนหน้าแสดงดัง**ตารางที่ 4**

จากผลการประเมินเวลาการทำงานของโมเดล พบว่าโมเดลที่ใช้วิธีการสอนโดยวิธีการ transfer learning ใช้

ตารางที่ 4 ค่าเฉลี่ย 95%HD จากโมเดล FUVGG19 และโมเดล FVGG19 และค่าเฉลี่ย 95%HD จากการศึกษาก่อนหน้า (หน่วย: มิลลิเมตร)

Organs	Wu X.[8]	Ibragimov B.[9]	Zhu W.[10]	TUVGG19	FFVGG19
BST	NA	5.15	6.42 ± 2.38	1.76 ± 0.49	1.95 ± 0.32
EYE_Lt	NA	NA	NA	1.08 ± 0.22	1.08 ± 0.21
EYE_Rt	NA	NA	NA	1.07 ± 0.24	1.16 ± 0.24
MAN	1.60	2.49	6.28 ± 2.21	3.45 ± 1.56	3.65 ± 1.36
PRO_Lt	3.25	6.97	9.31 ± 3.32	3.98 ± 2.05	4.02 ± 2.21
PRO_Rt	3.23	6.43	10.08 ± 5.09	3.37 ± 1.36	3.77 ± 1.56
SPC	NA	NA	NA	1.24 ± 0.20	1.61 ± 0.68

คำย่อ: FUVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, TUVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม U-net โมเดล backbone แบบ VGG19, FFVGG19 = โมเดลที่สอนด้วยวิธี fine tune รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, TFVGG19 = โมเดลที่สอนด้วยวิธี transfer learning รูปแบบสถาปัตยกรรม FPN โมเดล backbone แบบ VGG19, BST = brain stem, EYE_Lt = left eye globe, EYE_Rt = right eye globe, MAN = mandible, PRO_Lt = left parotid gland, PRO_Rt = right parotid gland, SPC = spinal cord, NA = not available

เวลาน้อยกว่า เมื่อเปรียบเทียบกับการสอนด้วยวิธี fine tine ทั้ง 2 รูปแบบสถาปัตยกรรมโครงข่าย ซึ่งสัมพันธ์กับจำนวนพารามิเตอร์ กล่าวคือวิธีการ transfer learning ทำให้จำนวนพารามิเตอร์ที่สามารถสอนได้ในแต่ละสถาปัตยกรรมโครงข่ายลดลง จึงทำให้ใช้เวลาในการสอนลดลง ในขณะที่เวลาการทำนาย mask ของโมเดลมีความแตกต่างตามรูปแบบสถาปัตยกรรมโครงข่าย เวลาการทำนาย mask ของโมเดลในรูปแบบสถาปัตยกรรมโครงข่ายเดียวกันที่ใช้วิธีการสอนที่ต่างกัน มีความต่างเพียงเล็กน้อย เนื่องจากการศึกษานี้ทำการศึกษาผ่าน virtual machines บนระบบออนไลน์ทำให้มีความคลาดเคลื่อนของเวลาจากการประมวลผลผ่านเครือข่ายอินเตอร์เน็ต

สรุป

การพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับกำหนดขอบเขตอัตโนมัติของอวัยวะข้างเคียงปกติบริเวณศีรษะและลำคอ จากการใช้รูปแบบสถาปัตยกรรมโครงข่าย U-Net โมเดล backbone แบบ VGG19 สอนโมเดลด้วยวิธีการ transfer learning และ FPN โมเดล backbone แบบ VGG19 สอนโมเดลด้วยวิธี fine tune ให้ผลการกำหนดขอบเขตที่มีค่า DSC มากกว่า 0.80 ในอวัยวะข้างเคียงปกติ brain stem, left and right eye globe, mandible และมีค่า DSC มากกว่า 0.72 ใน left and right parotid glands และ spinal cord ผลการประเมินค่า 95%HD น้อยกว่า 2 มิลลิเมตรในอวัยวะข้างเคียงปกติ brain stem, left and right eye globes, spinal cord และค่า 95%HD น้อยกว่า 4 มิลลิเมตร ในอวัยวะข้างเคียงปกติ mandible, left and right parotid gland เวลาในการทำนาย 1.286 และ 1.534 วินาทีต่อภาพ ตามลำดับ

กิตติกรรมประกาศ

ขอบคุณคณะสหเวชศาสตร์ มหาวิทยาลัยนเรศวร สำหรับสถานที่ทำงานวิจัย

เอกสารอ้างอิง

1. สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์. สถิติผู้ป่วยที่มารับบริการที่สาขารังสีรักษาและมะเร็งวิทยา 20 อันดับโรค (ราย) ปี 2562. 2562 [cited 27 ตุลาคม 2563]. Available from: https://www.chulacancer.net/service-statistics-inner.php?id=699.
2. 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. Radiat Oncol 2014;9:264.
3. Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. Lancet Oncol 2010;11:85-91.
4. 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. Int J Radiat Oncol Biol Phys 2012;82:1858-65.
5. Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. Radiation dose–volume effects of optic nerves and chiasm. Int J Radiat Oncol Biol Phys 2010; 76:S28-35.
6. Kirkpatrick JP, van der Kogel AJ, Schultheiss TE. Radiation dose-volume effects in the spinal cord. Int J Radiat Oncol Biol Phys 2010;76:S42-9.
7. 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. Radiother Oncol 2019;138:68-74.
8. 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. Proc SPIE Int Soc Opt Eng 2018;10576.
9. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys 2017;44:547-57.
10. 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. Med Phys 2018;46.
11. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. 2015;9351:234-41.
12. 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.
13. Yakubovskiy P. Segmentation Models 2018 [cited 2020 Oct 27]. Available from: https://segmentation-models.readthedocs.io/en/latest/.
14. Turgutlu K. dicom-contour 2018 [cited 2020 Oct 27]. Available from: https://github.com/KeremTurgutlu/dicom-contour.

15. Head-Neck-PET-CT [Internet]. Available from: https://wiki.cancerimagingarchive.net/display/Public/Head-Neck-PET-CT.
16. TCGA-HNSC [Internet]. [cited Oct 27]. Available from: https://wiki.cancerimagingarchive.net/display/Public/TCGA-HNSC.
17. IMAGENET Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [cited 2020 Oct 29]. Available from: http://image-net.org/challenges/LSVRC/2012/index#data.
18. Transfer learning & fine-tuning 2020 [updated 2020 May 12; cited 2020 Oct 29]. Available from: https://keras.io/guides/transfer_learning/.
19. 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. Med Phys 2018;45:2063-75.
20. Chen A, Dawant BM. Multi-atlas approach for the automatic segmentation of multiple structures in head and neck CT images. The MIDAS 2016.