Chulalongkorn University
Chula Digital Collections

Chulalongkorn University Theses and Dissertations (Chula ETD)

2023

การจำแนกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาวด้วยแขขจำลองทรานฟอร์มเมอร์สำหรับวิธีการแข่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุ

ภูมิพัฒน์ เจริญธนานุวัฒน์
คณะวิศวกรรมศาสตร์

Follow this and additional works at: https://digital.car.chula.ac.th/chulaetd

 Part of the Electrical and Electronics Commons

Recommended Citation

เจริญธนานุวัฒน์, ภูมิพัฒน์, "การจำแนกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาวด้วยแขขจำลองทรานฟอร์มเมอร์สำหรับวิธีการแข่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุ" (2023). *Chulalongkorn University Theses and Dissertations (Chula ETD)*. 10268.
https://digital.car.chula.ac.th/chulaetd/10268

This Thesis is brought to you for free and open access by Chula Digital Collections. It has been accepted for inclusion in Chulalongkorn University Theses and Dissertations (Chula ETD) by an authorized administrator of Chula Digital Collections. For more information, please contact ChulaDC@car.chula.ac.th.

การจำแนกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาวด้วยแบบจำลองทรานฟอร์มเมอร์สำหรับวิธีการแบ่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุ

นายภูมิพัฒน์ เจริญธนานุวัฒน์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2566

Classification of Red and White Blood Cells using Transformer-based Semantic Segmentation and Object Detection Joint Method

Mr. Phumiphat Charoentananuwat

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering
Department of Electrical Engineering
Faculty Of Engineering
Chulalongkorn University
Academic Year 2023

หัวข้อวิทยานิพนธ์	การจำแนกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาวด้วยแบบจำลองทรานฟอร์มเมอร์สำหรับวิธีการแบ่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุ
โดย	นายภูมิพัฒน์ เจริญธนานุวัฒน์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุรีย์ พุ่มรินทร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.นายแพทย์พลภัทร โรจน์นครินทร์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.สุรีย์ พุ่มรินทร์)

..... กรรมการ
(รองศาสตราจารย์ ดร.วันเฉลิม โปรา)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.สรรพฤทธิ์ มฤคทัต)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภูมิพัฒน์ เจริญธนานุวัฒน์ : การจำแนกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาวด้วยแบบจำลองทรานฟอร์มเมอร์สำหรับวิธีการแบ่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุ. (Classification of Red and White Blood Cells using Transformer-based Semantic Segmentation and Object Detection Joint Method) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุรีย์ พุ่มรินทร์

การตรวจเซลล์ลิมโฟบลาสติกเฉียบพลันในภาพฟิล์มเลือดเป็นวิธีการวินิจฉัยโรคมะเร็งเม็ดเลือดขาว การใช้การแบ่งภาพเชิงความหมายของเซลล์เม็ดเลือดขาวเฉียบพลันสามารถนำไปใช้ในการพัฒนาระบบการวิเคราะห์โดยใช้คอมพิวเตอร์ช่วย ในขอบเขตของการวิเคราะห์ฟิล์มส่วนปลาย วิธีการเรียนรู้เชิงลึก โดยเฉพาะอย่างยิ่งโครงข่ายประสาทเทียมแบบคอนโวลูชันมักถูกนำมาใช้ ปัจจุบัน โมเดลที่ใช้ทรานฟอร์มเมอร์สำหรับงานแบ่งภาพความหมายส่วนใหญ่ให้ผลลัพธ์ในเชิงความแม่นยำที่สูง ในการศึกษานี้ SegFormer ซึ่งเป็นแบบจำลองที่ใช้ทรานฟอร์มเมอร์สำหรับการแบ่งภาพความหมาย ถูกนำมาใช้เพื่อแบ่งส่วนและจำแนกเซลล์เม็ดเลือดขาวเฉียบพลันโดยใช้กลยุทธ์การฝึกอบรมที่แตกต่างกันสี่แบบ ผลลัพธ์ที่ดีที่สุดเกิดขึ้นได้โดยมีค่าเฉลี่ยของจุดตัด-โอเวอร์-ยูเนี่ยน (IoU) เท่ากับ 0.821 และความแม่นเฉลี่ย 0.917

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต .....

ปีการศึกษา 2566 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6372093721 : MAJOR ELECTRICAL ENGINEERING

KEYWORD: Blood smear image, Semantic segmenation, Transformer

Phumiphat Charoentananuwat : Classification of Red and White Blood Cells using Transformer-based Semantic Segmentation and Object Detection Joint Method. Advisor: Asst. Prof. SUREE PUMRIN, Ph.D.

The examination of peripheral blood smear images for acute lymphoblastic cells represents a diagnostic approach for leukemia. The utilization of semantic segmentation of acute lymphoblastic cells can be employed in the development of a computer-aided analysis system. In the realm of peripheral blood smear analysis, deep learning methods, particularly convolutional neural networks, are commonly utilized. Currently, transformer-based models have emerged as the state-of-the-art approach for semantic segmentation tasks. In this study, SegFormer, a transformer-based model for semantic segmentation, was utilized to segment and classify acute lymphoblastic cells using four distinct training strategies. The optimal outcome was achieved with a mean intersection-over-union (IoU) of 0.821 and a mean accuracy of 0.917.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study: Electrical Engineering Student's Signature .....

Academic Year: 2023 Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดีผู้ศึกษาขอขอบพระคุณบุคคล และกลุ่มบุคคลต่าง ๆ ที่ได้กรุณาให้คำปรึกษา แนะนำและช่วยเหลืออย่างดียิ่ง ดังรายนามดังต่อไปนี้ ผู้ช่วยศาสตราจารย์สุรีย์ พุ่มรินทร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำปรึกษาทั้งในด้านวิชาการ ขอบเขตและรูปแบบของวิทยานิพนธ์ ตลอดจนเป็นส่วนสำคัญที่ผลักดันให้งานสำเร็จได้ทันเวลา ศาสตราจารย์นายแพทย์ ดร.พลภัทร โรจน์นครินทร์ ประธานกรรมการสอบวิทยานิพนธ์ที่ได้ให้คำแนะนำในมุมมองของการนำไปใช้งานในทางการแพทย์ รองศาสตราจารย์ ดร.วันเฉลิม โปรา กรรมการสอบวิทยาพินธ์ และ ดร.สรรพฤทธิ์ มฤคทัต กรรมการสอบวิทยานิพนธ์ภายนอก สำหรับคำแนะนำในด้านวิชาการ เพื่อปรับปรุงให้งานวิจัยชิ้นนี้มีความสมบูรณ์ยิ่งขึ้น รวมทั้งแนวทางในการพัฒนาต่อยอด คณาจารย์ เจ้าหน้าที่และสมาชิกห้องปฏิบัติการ Embedded Systems and IC Design Research Laboratory (ESID) คณะวิศวกรรมศาสตร์จุฬาลงกรณ์มหาวิทยาลัย ที่ได้สนับสนุนและเป็นส่วนเติมเต็มสำคัญที่ทำให้งานนี้สำเร็จได้ โดยเฉพาะคุณณัทกร เกษมสำราญ สำหรับชุดข้อมูลที่นำมาใช้ในการพัฒนาแบบจำลองการเรียนรู้เชิงลึก และครอบครัวเจริญธนานุวัฒน์ ที่สนับสนุน และเป็นกำลังใจในทุกกิจกรรมของผู้ศึกษาตลอดมา

ภูมิพัฒน์ เจริญธนานุวัฒน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

หน้า

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในการวินิจฉัยของโรคแพทย์เป็นกระบวนการหนึ่งที่จำเป็นต้องทำเพื่อให้สามารถรักษาโรคได้ ซึ่งในการวินิจฉัยนั้น แพทย์จะต้องใช้ข้อมูลในหลายด้านเพื่อประกอบการวินิจฉัย ไม่ว่าจะเป็นประวัติอาการ ผลการตรวจร่างกาย และผลตรวจทางห้องปฏิบัติการของผู้ป่วย ในส่วนของผลตรวจทางห้องปฏิบัติการจะเป็นผลที่ได้จากการตรวจสารคัดหลั่ง (Secretions) ต่างๆของมนุษย์เช่นเลือด ปัสสาวะ อุจาระ น้ำลาย เสมหะ หรือเยื่อบุคอหอยเป็นต้น โดยการตรวจสารคัดหลั่งแต่ละชนิดก็จะมีวิธีการที่แตกต่างกันออกไปเช่นการตรวจเยื่อบุคอหอยเพื่อหาเชื้อไวรัสโคโรนา 2019 (COVID-19) ที่ทำโดยการเก็บตัวอย่างของเยื่อบุคอหอยไว้ในหลอดชนิด VTM/UTM เพื่อรอให้เชื้อเพิ่มจำนวน แล้วจึงวัดปริมาณสายดีเอ็นเอที่สร้างขึ้นใหม่โดยการใช้สารฟลูออเรสเซนซ์ หรือจะเป็นการตรวจฟิล์มเลือด (Blood smear) ที่ทำโดยการนำตัวอย่าง (Specimen) มาเติมลงบนแผ่นสไลด์ (Slide glass) แล้วนำไปย้อมสีก่อนที่จะนำมาตรวจด้วยกล้องจุลทรรศน์ แล้วจึงนับจำนวนเซลล์ต่างๆเช่นเซลล์เม็ดเลือดแดง (Red Blood Cells: RBCs) เซลล์เม็ดเลือดขาว (White Blood Cells: WBCs) และเกล็ดเลือด (Platelets) ด้วยมนุษย์ โดยจะเก็บข้อมูลจากเม็ดเลือดแดงโดยการเปรียบเทียบขนาดกับเม็ดเลือดขาวชนิดลิมโฟไซต์ (Lymphocytes) ขนาดเล็ก รูปแบบการกระจายตัว บริเวณที่มีสีขาวต่อพื้นที่เซลล์ รูปร่างและขนาด เก็บข้อมูลเม็ดเลือดขาวโดยระบุชนิดของเม็ดเลือดขาว นับจำนวนเม็ดเลือดขาวทั้งหมดและนับแยกชนิด อัตราส่วนนิวเคลียสต่อพื้นที่เซลล์เป็นต้น ส่วนเกล็ดเลือดจะทำการเก็บข้อมูลโดยการนับจำนวน ขนาด และการกระจายตัว [1] เพื่อที่จะนำค่าต่างๆที่ได้จากการสังเกตไปใช้ในการวิเคราะห์และวินิจฉัยโรคต่างๆต่อไป เช่นนำไปคำนวณเพื่อหาค่าความสมบูรณ์ของเซลล์เม็ดเลือด (Complete Blood Count: CBC) โดยวิธีการที่กล่าวมาข้างต้นนี้เป็นวิธีการที่ต้องทำเป็นประจำ (Routine) ที่จำเป็นต้องใช้นักเทคนิคการแพทย์ที่มีประสบการณ์เพื่อที่จะลดโอกาสผิดพลาดให้เหลือน้อยที่สุด ทำให้สถานพยาบาลขนาดเล็กที่มีทรัพยากรบุคคลและเครื่องมือไม่เพียงพอไม่สามารถให้บริการผู้ป่วยได้อย่างเต็มที่ และยังอาจต้องเสียค่าใช้จ่ายมากขึ้นจากการที่ต้องส่งตัวอย่างไปตรวจที่ห้องปฏิบัติการภายนอก

ตั้งแต่อดีตจนถึงปัจจุบัน ได้มีการพยายามใช้เทคนิคการประมวลภาพ (Image Processing) และ คอมพิวเตอร์วิทัศน์ (Computer Vision) มาประยุกต์ใช้ในการวิเคราะห์ภาพทางการแพทย์มาโดยตลอด ซึ่งงานด้านการวิเคราะห์ฟิล์มเลือดก็เป็นหนึ่งในงานที่มีการพัฒนามาอย่างต่อเนื่อง ตั้งแต่การพัฒนาเทคนิคการประมวลผลภาพไปจนถึงการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก (Deep

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

learning) ในการทำการตรวจจับวัตถุ (Object Detection) และการให้ความหมายภาพแต่ละพิกเซล (Semantic Segmentation) เพื่อที่จะแยกแยะชนิดของเซลล์ต่างๆในเลือด รวมไปถึงการนับจำนวนเซลล์เพื่อที่จะนำไปวิเคราะห์ความสมบูรณ์ของเซลล์เม็ดเลือด โดยงานวิจัยต่างๆที่ผ่านมา เข้ามามีส่วนช่วยในการสังเคราะห์ข้อมูลจำนวนมากของผู้ป่วย และทำให้กระบวนการรักษาทำได้อย่างรวดเร็วมากยิ่งขึ้น

ในงานวิจัยนี้ ผู้วิจัยมีความต้องการที่จะนำเทคนิคการเรียนรู้เชิงลึกชนิดหนึ่งที่มีการใช้งานในด้านการประมวลผลภาพที่เรียกว่าวิชั่นทรานฟอร์มเมอร์ (Vision transformer) มาประยุกต์ใช้ในการให้ความหมายภาพแต่ละพิกเซล และการประสานร่วมกันระหว่างการให้ความหมายภาพแต่ละพิกเซลกับการตรวจจับวัตถุ เพื่อที่จะนำมาใช้ในการแยกแยะเซลล์เม็ดเลือดแดง และเซลล์เม็ดเลือดขาว พร้อมทั้งนับจำนวนเซลล์ต่างๆที่ปรากฏในภาพ เพื่อที่นำมาใช้ในการคำนวณต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

- เพื่อพัฒนาแบบจำลองการเรียนรู้เชิงลึกแบบ Vision Transformer สำหรับคัดแยกและนับจำนวนเซลล์ในฟิล์มเลือด โดยใช้การให้ความหมายภาพแต่ละพิกเซล (Semantic segmentation) โดยใช้เครื่องมือ Pytorch ซึ่งเป็นเครื่องมือแบบ Open source
- เพื่อพัฒนาแบบจำลองการเรียนรู้เชิงลึกแบบ Vision Transformer สำหรับคัดแยกและนับจำนวนเซลล์ในฟิล์มเลือด โดยประสานการตรวจจับวัตถุ (Object detection) และการให้ความหมายภาพแต่ละพิกเซล (Semantic Segmentation) โดยใช้เครื่องมือ Pytorch ซึ่งเป็นเครื่องมือแบบ Open source
- เพื่อพัฒนาการทำการขยายชุดข้อมูล (Data augmentation) สำหรับการสอนแบบจำลองเพื่อการให้ความหมายภาพแต่ละพิกเซล และการตรวจจับวัตถุ เพื่อที่จะลดปัญหาจากชุดข้อมูลที่ไม่สมดุล (Imbalance dataset)

1.3 ขอบเขตของงานวิจัย

- พัฒนา Algorithm สำหรับการคัดแยกและนับจำนวนเซลล์ด้วยใช้ภาษา Python โดยใช้ OpenCV และ Numpy สำหรับการเตรียมข้อมูล และใช้ Pytorch หรับการสร้างแบบจำลองการเรียนรู้เชิงลึกแบบ Vision Transformer ซึ่งต้องมีความแม่นยำไม่น้อยกว่าร้อยละ 80

Chulalongkorn University

- สร้างชุดข้อมูลภาพถ่ายฟิล์มเลือด (Blood smear dataset) จากชุดข้อมูลส่วนตัว (Private dataset) จำนวน 200 ภาพ ที่ได้รับการติดป้าย (Labeling) ประกอบด้วยเม็ดเลือดแดง (Red Blood Cell: RBC) และเม็ดเลือดขาว (White Blood Cell: WBC)

1.4 วิธีการดำเนินการวิจัย

- ศึกษาการทำงานของแบบจำลองการเรียนรู้เชิงลึกแบบ Vision Transformer
- ทำการอธิบายข้อมูล (Annotate data) สำหรับการตรวจจับวัตถุและการให้ความหมายภาพแต่ละพิกเซล
- ทำการตรวจสอบข้อมูลเบื้องต้น (Exploratory Data Analysis: EDA) เพื่อวิเคราะห์การกระจายตัวของข้อมูล
- ศึกษาการทำการขยายชุดข้อมูลสำหรับการตรวจจับวัตถุและการให้ความหมายภาพแต่ละพิกเซล เพื่อที่จะลดปัญหาจากการกระจายตัวของข้อมูลที่ไม่สมดุล
- พัฒนาขั้นตอนทวิธีสำหรับการขยายชุดข้อมูล
- สร้างแบบจำลองการเรียนรู้เชิงลึก และฝึกแบบจำลองโดยใช้ชุดข้อมูลจริง และชุดข้อมูลที่ได้จากการขยายชุดข้อมูล
- ทดสอบความแม่นยำของแบบจำลอง และเปรียบเทียบผลลัพธ์ของแบบจำลองทั้ง 2 แบบ
- สรุปผลการทดลอง และทำบทสรุปการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- ชุดข้อมูลให้สำหรับงานวิจัยในอนาคต
- ขั้นตอนวิธีการขยายชุดข้อมูลสำหรับการตรวจจับวัตถุและการให้ความหมายภาพแต่ละพิกเซล และลดปัญหาจากการกระจายตัวของข้อมูลที่ไม่สมดุล
- แบบจำลองการเรียนรู้เชิงลึกสำหรับการคัดแยก และนับจำนวนเซลล์ในฟิล์มเลือดที่มีความแม่นยำไม่น้อยกว่าร้อยละ 80 โดยวิธีการวัด IoU

บทที่ 2 ความรู้พื้นฐาน ทฤษฎีและวรรณกรรมงานวิจัยที่เกี่ยวข้อง

2.1 ทบทวนวรรณกรรมงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ประยุกต์ใช้เทคนิคการให้ความหมายภาพแต่ละพิกเซล และการประสานเทคนิคการตรวจจับวัตถุและการให้ความหมายภาพในแต่ละพิกเซลในสร้างแบบจำลองการเรียนรู้เชิงลึกเพื่อระบุตำแหน่งพร้อมทั้งจำแนกชนิดของเซลล์เม็ดเลือดที่ปรากฏในภาพฟิล์มเลือด โดยในขั้นตอนการเตรียมข้อมูลก่อนที่จะสอนแบบจำลองการเรียนรู้เชิงลึก ได้มีการประยุกต์ใช้เทคนิคการขยายชุดข้อมูลเพื่อให้ชุดข้อมูลมีขนาดใหญ่มากขึ้น ซึ่งมีแนวโน้มที่จะทำให้แบบจำลองการเรียนรู้เชิงลึกมีประสิทธิภาพที่ดียิ่งขึ้น

ในงานวิจัยของ A. Tareef. [2] ได้แบ่งการทำงานเป็นสองส่วนคือการแยกนิวเคลียส และการแยกไซโตพลาสซึม ในส่วนของการแยกนิวเคลียสได้ทำการเพิ่มความเข้มสีของนิวเคลียสโดยใช้การะบวนการแกรม (Gram-Schmidt orthogonalization) จากนั้นทำการแยกนิวเคลียสด้วยเทคนิคการสลับสัณฐานหลายระดับแบบคู่ด้วยตนเอง (Self-dual Multi-scale Morphological Toggle: SMMT) ที่เป็นหนึ่งในเทคนิคการทำสัณฐานภาพ (Image Morphological) ในส่วนของการแยกไซโตพลาสซึมได้ใช้วิธีการควบคุมลายน้ำด้วยเครื่องหมาย (Marker-controlled watershed) จากนั้นทำการแยกไซโตพลาสซึมด้วยช่วงที่มีความเข้มข้นสูงสุดของฮิสโตแกรม (Histogram) จากนั้นนำผลลัพธ์ขอทั้งสองส่วนมารวมกัน ซึ่งงานวิจัยนี้มีข้อจำกัดที่มุ่งเน้นไปที่การแยกเซลล์เม็ดเลือดขาวเพียงอย่างเดียว G. K. Chadha. [3] ได้ทำการแยกเซลล์เม็ดเลือดแดงโดยการใช้เกณฑ์ (Thresholding) ซึ่งจะทำการลดระดับสีเทา (Gray scale level) ให้เหลือสองระดับหรือก็คือสีดำและสีขาว แล้วทำการสกัดลักษณะเด่นโดยการใช้การตรวจจับขอบ (Edge Detection) ด้วยวิธีการตรวจจับขอบของโซเบล (Sobel edge detection) และใช้การกัดกร่อนรูปภาพ (Image erosion) เพื่อทำให้ภาพที่ได้มีความเรียบมากขึ้น ในส่วนต่อมาได้มีการใช้การเปลี่ยนแปลงของฮาว์ก (Hough transformation) ที่เป็นหนึ่งในเทคนิคการตรวจจับขอบมาใช้ในการนับจำนวนเซลล์เม็ดเลือดแดงในภาพ ซึ่งงานวิจัยนี้มีจุดมีเน้นไปที่การแยกและนับเซลล์เม็ดเลือดแดงเพียงอย่างเดียว A.M.P.G. Vale [4] ได้นำเสนอวิธีการแยกเซลล์ต่างๆในภาพฟิล์มเลือดโดยการใช้ตรรกะคลุมเครือ (Fuzzy logic) โดยนำช่องสีเขียว (Green channel of the RGB) มาหาช่วงที่มืดที่สุด (Dark peak) กลางที่สุด (Medium peak) และสว่างที่สุด (Light peak) จากฮิสโตแกรม ซึ่งจะนำไปใช้ในการกำหนดช่วงของสิ่งที่สนใจในภาพและกำหนดจุดกลาง (Centroid) เพื่อที่จะนำตรรกะคลุมเครือมาใช้ในการหาพื้นที่จริงที่จะใช้ในการกำหนดขอบเขตของการแยกภาพต่อไป จากชุดข้อมูลทั้งหมด 530 ภาพ เมื่อผ่านกระบวนการดังกล่าวนักวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

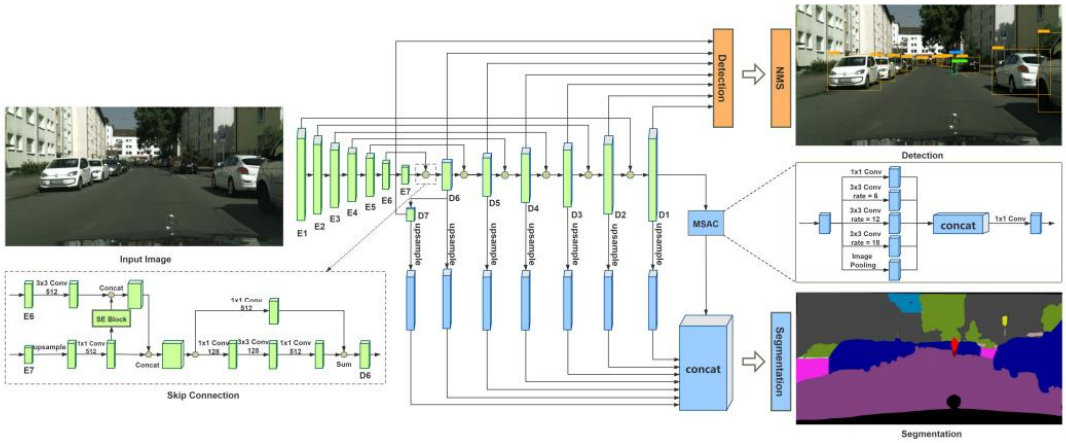
กล่าวว่าได้ความแม่นยำเฉลี่ยที่ร้อยละ 97.31 M. I. Razzak [5] ได้นำเสนอวิธีในการแยกภาพเซลล์ด้วยการใช้แบบจำลองการเรียนรู้เชิงลึกสำหรับการแบ่งส่วนการรับรู้รูปร่าง (Contour aware segmentation) อย่าง Deep Contour Aware Network (DCAN) [6] ที่สามารถสร้างลักษณะเด่นหลายระดับ (Multi-scale feature representation) เพื่อจัดการปัญหาเกี่ยวกับความหลายหลากและการซ้อนทับกันของเซลล์ ทำให้สามารถแบ่งแยกเซลล์ที่ซ้อนทับกันได้ แล้วจึงทำการนับจำนวนเซลล์เม็ดเลือดขาวและเซลล์เม็ดเลือดแดงด้วยการวิเคราะห์สัณฐานของภาพ โดยได้ทำการสอนแบบจำลองด้วยชุดข้อมูล ALL-IDB [7] ซึ่งเป็นชุดข้อมูลภาพเซลล์เม็ดเลือดแบบเปิด (Open dataset) โดยได้ความแม่นยำเฉลี่ยอยู่ที่ร้อยละ 98.36 T. Tran [8] ได้นำแบบจำลองการเรียนรู้เชิงลึกสำหรับการแยกส่วนภาพอย่าง SegNet [9] ซึ่งเป็นแบบจำลองชนิดตัวเข้ารหัสอัตโนมัติแบบสังวัตนาการ (Convolutional autoencoder) ที่มีแกนหลัก (Backbone) มาจาก VGG-16 [10] โดยนำมาประยุกต์ใช้ในการแยกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาว โดยได้ใช้ชุดข้อมูล ALL-IDB เช่นเดียวกับ [5] โดยได้ความแม่นยำเฉลี่ยร้อยละ 91 และคะแนนในหน่วย IoU (Intersection over Union) ที่ 0.81 S. Fujita [11], N. Dhieb [12], D. Loh [13] และ M. P. Paing [14] ได้ใช้แบบจำลองการเรียนรู้เชิงลึกสำหรับการแยกภาพอย่าง Mask R-CNN [15] สำหรับ [12] ได้ทำการสอนและทดสอบแบบจำลองโดยใช้ชุดข้อมูลจาก Isfahan medical and signal processing (MISP) ในส่วนของการทำการสกัดคุณลักษณะเด่น นักวิจัยได้ใช้วิธีพีรามิดของลักษณะเด่น (Feature Pyramid Network) [16] โดยใช้แบบจำลองการเรียนรู้เชิงลึกอย่าง ResNet [17] โดยแบบจำลองที่ได้มีประสิทธิภาพในการตรวจจับเซลล์เม็ดเลือดแดงร้อยละ 92 และเซลล์เม็ดเลือดขาวร้อยละ 96 ส่วน [11] สอนแบบจำลองและทดสอบแบบจำลองโดยใช้ชุดข้อมูล DSB2018 [18] ซึ่งเป็นชุดข้อมูลเกี่ยวกับเซลล์เนื้อเยื่อต่างทั้งที่ปกติและผิดปกติ ทำให้ชุดนี้เป็นชุดข้อมูลที่ไม่สมดุล นักวิจัยจึงได้ใช้ความสูญเสียโฟคัล (Focal loss) ในการตรวจสอบความถูกต้อง (Validation) เพื่อลดปัญหาที่เกิดจากชุดข้อมูลที่ไม่สมดุล เมื่อนำไปทดสอบกับชุดข้อมูลทดสอบ ได้คะแนน F1 ที่ 0.888 ส่วนของ [13] ใช้ชุดข้อมูลที่เก็บจากฟิล์มเลือดที่ซื้อมาจากธนาคารเลือดระหว่างรัฐ (Interstate blood bank) นักวิจัยได้แบ่งชุดข้อมูลออกเป็นสามส่วนคือ 173 ภาพ สำหรับการสอนแบบจำลอง 64 ภาพ สำหรับการตรวจสอบความถูกต้อง และ 60 ภาพสำหรับการทดสอบ รวมเป็นจำนวน 297 ภาพ ได้ความแม่นยำที่ร้อยละ 94.57 ส่วนของ [14] นักวิจัยได้ทำการพัฒนาแบบจำลองเพื่อตรวจจับโรคมะเร็งไขกระดูกมัยอิโลมา (Multiple myeloma) โดยใช้ชุดข้อมูล TCIA [19] และ SegPC [20] ซึ่งถูกแบ่งไปเป็นชุดข้อมูลสำหรับการสอนแบบจำลอง 298 ภาพ สำหรับการตรวจสอบความถูกต้องของแบบจำลอง 200 ภาพ และสำหรับทดสอบแบบจำลอง 277 ภาพ รวมทั้งหมด 775 ภาพ โดยได้ทำการแยกเซลล์ที่ถูก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

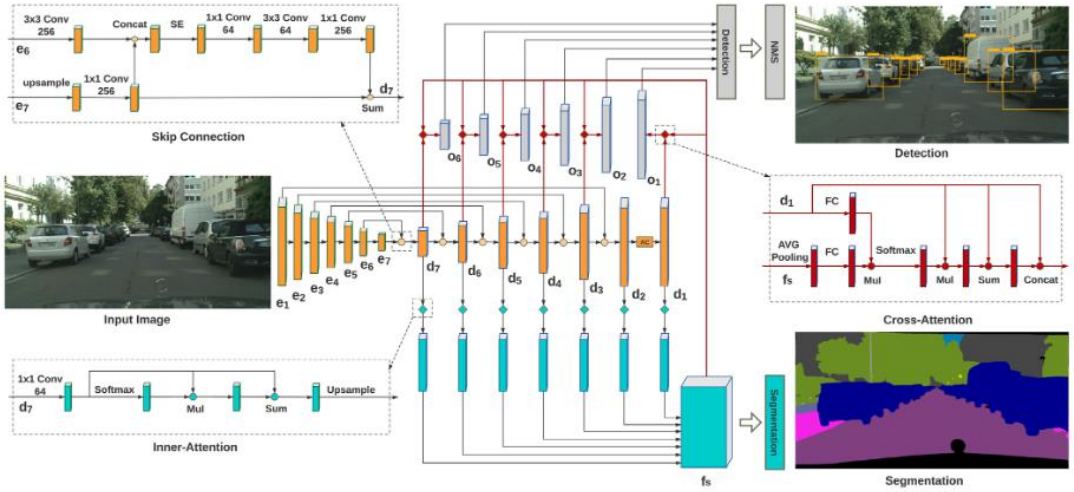
ย้อมสีโดยการใช้การยืดภาพเพื่อความชัดเจน (Contrast stretching) ของเฉียดสี (Hue) ของภาพ จากนั้นได้ทำการขยายชุดข้อมูลโดยใช้วิธีการขยายชุดข้อมูลเชิงลึกอย่างชาญฉลาด (Deep wise data augmentation) ซึ่งเป็นวิธีการเรียนรู้เชิงลึกอย่างหนึ่งที่จะทำการคัดลอกพื้นที่ที่ต้องการแยก (Segmented region) ไปวางเพิ่มในส่วนอื่นๆของภาพ ทำให้ได้ภาพที่มีพื้นที่หรือวัตถุเป้าหมายเพิ่มมากขึ้น โดยมีพื้นฐานมาจากการผสมภาพเชิงลึก (Deep Image Blending) [21] ผู้วิจัยได้เปรียบเทียบประสิทธิภาพของแบบจำลองที่สอนโดยใช้ชุดข้อมูลหลักและชุดข้อมูลที่ได้จากการทำการขยายข้อมูล กับแบบจำลองที่สอนโดยใช้เพียงชุดข้อมูลหลัก ได้ประสิทธิภาพในหน่วย IoU อยู่ที่ 0.8721 และ 0.7689 ตามลำดับ

Z. Nan [22] [23] ได้นำเสนอวิธีการที่ทำการแบ่งภาพเชิงความหมายร่วมกับการตรวจจับวัตถุบนภาพการจราจร โดยในปี 2020 ได้นำเสนอ [22]ที่ใช้โครงข่ายประสาทเทียมแบบการเข้ารหัสอัตโนมัติ (Autoencoder) ในการสกัดคุณลักษณะเด่นที่มีการใช้การข้ามการเชื่อมต่อ (Skip Connection) จากชั้นต่างๆของส่วนเข้ารหัสไปยังชั้นต่างๆในส่วนถอดรหัสในลำดับเดียวกัน และนำผลลัพธ์ของจากการถอดรหัสในแต่ละชั้นไปเข้าแบบจำลองของสำหรับถอดรหัสของการตรวจจับวัถตุและการแบ่งภาพเชิงความหมายดังรูปที่ 1 โดยได้ทำการสอนและทดสอบแบบจำลองบน [24] ชุดข้อมูล VOC2012 โดยได้แยกการทำงานเป็น 3 ชนิดคือการตรวจจับวัตถุ การแบ่งภาพเชิงความหมายและการตรวจจับวัตถุร่วมกับการแบ่งภาพเชิงความหมาย เมื่อทำการตรวจจับวัตถุได้ผลลัพธ์ในหน่วย mAP ที่ร้อยละ 36.3 และเมื่อทำการแบ่งภาพได้ผลลัพธ์ในหน่วย mIoU ที่ร้อยละ 55.0 และเมื่อทำงานร่วมกันได้ผลลัพธ์ในหน่วย mAP และ mIoU ที่ร้อยละ 40.0 และ 55.5 ตามลำดับ ต่อมาในปี 2021 ได้นำเสนอวิธี [23] ที่มีพื้นฐานคล้ายแบบจำลองก่อนหน้า แต่เพิ่มความสนใจภายใน (Inner-attention) เข้าไปในส่วนของการถอดรหัสของการสกัดคุณลักษณะเด่นแต่ละชั้น เพิ่มขยายขนาดของข้อมูลก่อนทำการแบ่งภาพเชิงความหมาย และเพิ่มความสนใจภายนอก (Cross-attention) เพื่อทำการรวมข้อมูลจากส่วนของการถอดรหัสของการสกัดคุณลักษณะเด่นแต่ละชั้นกับส่วนการถอดรหัสสำหรับการแบ่งภาพเชิงความหมายและนำผลลัพธ์ที่ได้ไปใช้ในการตรวจจับวัตถุดังรูปที่ 2 เมื่อทดสอบแบบจำลองโดยใช้ชุดข้อมูลเดียวกันกับ [22] โดยใช้วิธีการทำงานร่วมกันระหว่างการตรวจจับวัตถุและการแบ่งภาพเชิงความหมายให้ผลลัพธ์ที่ดียิ่งขึ้นเป็นร้อยละ 41.2 และ 57.4 ในหน่วย mAP และ mIoU ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



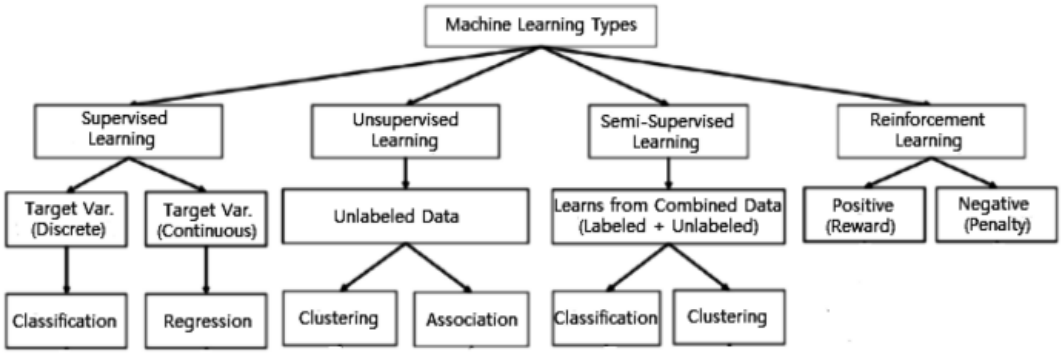
รูปที่ 1 block diagram ของแบบจำลองการตรวจจับวัตถุร่วมกับการแบ่งภาพเชิงความหมายของ [22]



รูปที่ 2 block diagram ของแบบจำลองการตรวจจับวัตถุร่วมกับการแบ่งภาพเชิงความหมาย[23]

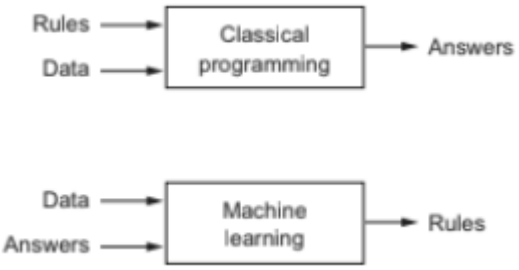
2.2 ความรู้พื้นฐาน และทฤษฎีที่เกี่ยวข้อง

2.2.1 การเรียนรู้ของเครื่อง



รูปที่ 3 ประเภทของการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่องเป็นขั้นตอนวิธีที่ทำให้คอมพิวเตอร์สามารถสร้างความรู้โดยเรียนรู้จากข้อมูล แล้วจึงนำความรู้นั้นไปใช้ในการวิเคราะห์ คาดการณ์ หรือตัดสินใจได้ [25] โดยการเรียนรู้ของเครื่องสามารถแบ่งออกได้เป็น 4 ประเภทได้แก่



รูปที่ 4 ความแตกต่างระหว่างขั้นตอนวิธีการเรียนรู้ของเครื่อง และการเขียนโปรแกรมแบบทั่วไป

1. การเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นรูปแบบการเรียนรู้ของเครื่องที่เรียนรู้จากชุดข้อมูลสำหรับฝึก (Training set) หรือที่เรียกว่าวิธีการขับเคลื่อนด้วยภาระหน้าที่ (Task-driver approach) โดยจะนำคู่ของข้อมูลที่เป็นคุณลักษณะ (Feature) และเป้าหมาย (Target) จากชุดข้อมูลสำหรับฝึกมาปรับพารามิเตอร์ของฟังก์ชัน เพื่อให้ฟังก์ชันนี้ทำการดำเนินการกับคุณลักษณะของข้อมูลแล้วให้ผลลัพธ์ใกล้เคียงกับเป้าหมายในชุดข้อมูลสำหรับตรวจสอบความถูกต้อง (Validation set) มากที่สุด จากรูปที่ 3 จะเห็นว่าการเรียนรู้แบบ Supervised Learning แบ่งออกได้อีก 2 ประเภทคือการจำแนกแยกแยะ (Classification) และการทดถอย (Regression) โดยการ

จำแนกแยกแยะเป็นการแยกชนิดของข้อมูลที่ไม่มีความต่อเนื่อง (Discrete data) ส่วนการทดถอยเป็นการทำให้ฟังก์ชันเหมาะกับข้อมูลที่มีความต่อเนื่อง (Continuous data)

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นขั้นตอนวิธีที่ใช้ในการวิเคราะห์ชุดข้อมูลที่ไม่มีการติดป้าย (Unlabeled dataset) หรือที่เรียกว่าวิธีการขับเคลื่อนด้วยข้อมูล (Data-driven approach)เพื่อที่จะสกัดคุณลักษณะ ระบุแนวโน้มและโครงสร้างของข้อมูล และแบ่งกลุ่มข้อมูล โดยทั่วไปแล้วจะถูกนำไปใช้งานในด้านการแบ่งกลุ่มข้อมูล (Clustering) การคาดการณ์ความหนาแน่น (Density estimation) การเรียนรู้คุณลักษณะ (Feature learning) การลดมิติข้อมูล (Dimensionality reduction) การหาความสัมพันธ์ของข้อมูล (Finding association rules) และการตรวจจับสิ่งผิดปกติ (Anomaly detection) เป็นต้น ซึ่งในงานด้านการประมวลผลภาพ ได้มีการนำวิธีการแบ่งกลุ่มอย่าง K-means ซึ่งเป็นหนึ่งในขั้นตอนวิธีการเรียนรู้แบบไม่มีผู้สอนไปใช้ในการทำการแบ่งภาพด้วยเช่นกัน

3. การเรียนรู้กึ่งมีผู้สอน (Semi-supervised learning) เป็นขั้นตอนวิธีที่ผสมผสานระหว่างการเรียนรู้แบบมีผู้สอนและการเรียนรู้แบบไม่มีผู้สอน ซึ่งดำเนินการกับทั้งข้อมูลที่มีป้ายกำกับและข้อมูลที่ไม่มีป้ายกำกับ ซึ่งมีเป้าหมายเพื่อให้ได้แบบจำลองที่สามารถให้ความแม่นยำได้ดีกว่าแบบจำลองที่ใช้ข้อมูลที่ติดป้ายเพียงอย่างเดียว มักถูกนำไปใช้ในการสร้างแบบจำลองในการแปลภาษา การตรวจจับการฉ้อโกง และการติดป้ายในกับข้อมูล

4. การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) เป็นขั้นตอนวิธีที่ทำให้เครื่องสามารถเรียนรู้เพื่อที่จะปรับปรุงแบบจำลองได้อย่างอัตโนมัติ โดยอาศัยข้อมูลจากผลลัพธ์ของงานที่ต้องการทำและสภาพแวดล้อม หรือที่เรียกว่าวิธีการขับเคลื่อนด้วยสภาพแวดล้อม (Environment-driven approach) มักถูกนำมาใช้งานในด้านหุ่นยนต์ การขับเคลื่อนอัตโนมัติ สายการผลิตและการกระจายสินค้า

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

[2] งานของการเรียนรู้ของเครื่องจะขึ้นอยู่กับปัญหา และข้อมูลที่มีอยู่ เช่นการจัดประเภทข้อมูลจะทำการกำหนดหมวดหมู่ของข้อมูลนั้นๆ ส่วนการจัดกลุ่มจะจัดตามความคล้ายของข้อมูล ซึ่งการจะเลือกใช้ขั้นตอนวิธีต่างๆของการเรียนรู้ของเครื่อง จะต้องขึ้นอยู่กับรูปแบบของข้อมูลนั้นๆ [1] โดยขั้นตอนวิธีของการเรียนรู้ของเครื่องสามารถแบ่งได้เป็นการวิเคราะห์การจำแนก (Classification analysis) การวิเคราะห์การทดถอย (Regression Analysis) การวิเคราะห์การจัดกลุ่ม (Cluster analysis) การลดมิติข้อมูลและการเรียนรู้คุณลักษณะ (Dimensionality reduction and feature learning) การเรียนรู้กฎหาสัมพันธ์ (Association rule learning) การเรียนรู้แบบเสริมกำลัง และ

โครงข่ายประสาทเทียมและการเรียนรู้เชิงลึก (Artificial neural network and deep learning) ซึ่งงานสำคัญที่ใช้ในงานวิจัยนี้ และจะกล่าวถึงในส่วนถัดไป

2.2.2 การเรียนรู้เชิงลึก

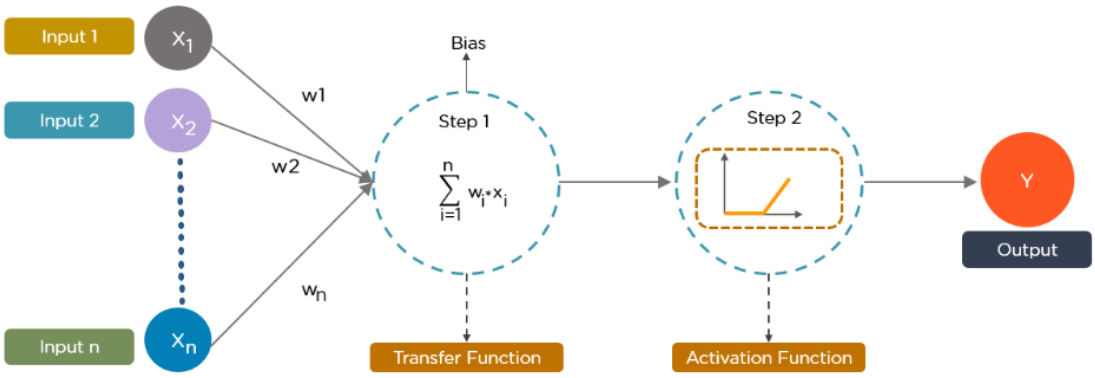
การเรียนรู้เชิงลึกเป็นส่วนหนึ่งของการเรียนรู้ของเครื่อง ที่ประกอบไปด้วยโครงข่ายประสาทเทียมมากกว่า 3 ชั้นขึ้นไป โดยโครงข่ายประสาทเทียมเหล่านี้จะจำลองหน้าที่เหมือนสมองของมนุษย์ ทำให้ขั้นตอนวิธีนี้มีความสามารถในการเรียนรู้คล้ายกับมนุษย์

หน่วยย่อยของโครงข่ายประสาทเทียมจะถูกเรียกว่าเพอร์เซปตรอน (Perceptron) ดังที่แสดงในรูปที่ 5 โดยเมื่อใช้ฟังก์ชันกระตุ้นเป็น ReLU (Rectified Linear Unit) ความสัมพันธ์จะเป็นดังสมการต่อไปนี้

$$F\big(T(x,w)\big) = \begin{cases} T(x,w), & T(x,w) > 0 \\ 0, & otherwise \end{cases} \quad (1)$$

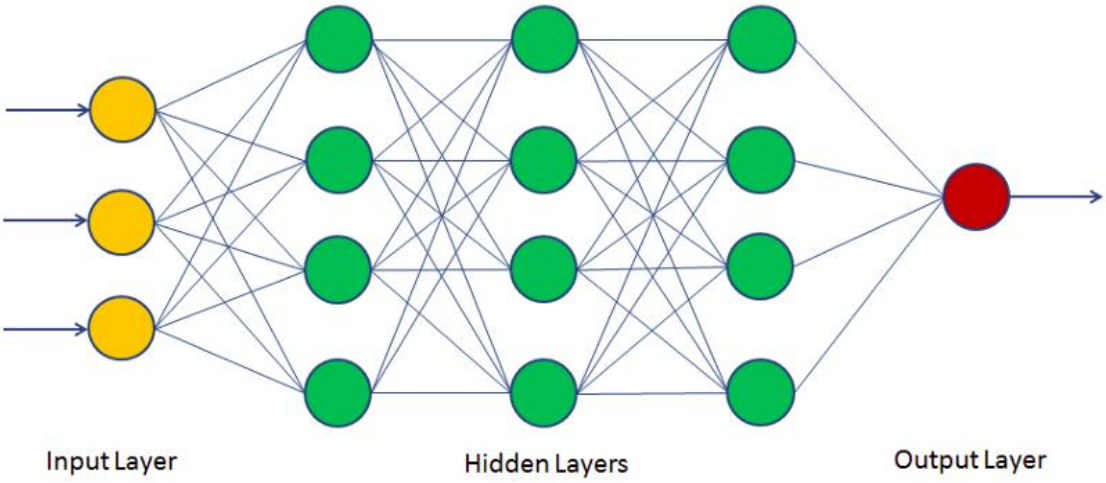
$$T(x,w) = \sum_{i=0}^{n} x_i w_i + b \quad (2)$$

โดย x คืออินพุต w คือค่าถ่วงน้ำหนักของอินพุต (Weight) และ b คือค่าความเอนเอียง (Bias) ของฟังก์ชันส่งผ่าน (Transfer function)



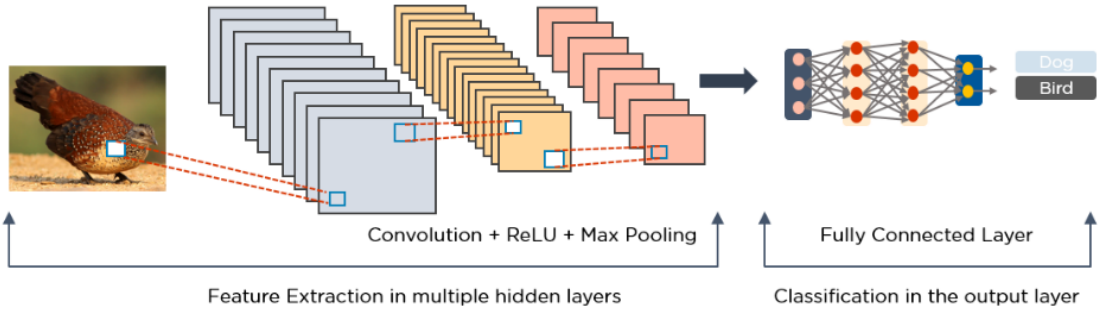
รูปที่ 5 โครงสร้างของเพอร์เซปตรอน

เมื่อนำเพอร์เซปตรอนหลายตัวมาต่อกันจะได้เพอร์เซปตรอนหลายชั้น (Multi-layer perceptron) หรือที่เรียกว่าโครงข่ายประสาทเทียม



รูปที่ 6 โครงข่ายประสาทเทียมที่เกิดจากการนำเพอร์เซปตรอนมาต่อกัน

ต่อมาได้มีการพัฒนาแบบจำลองการเรียนรู้เชิงลึกขึ้นมาหลากหลายรูปแบบได้แก่โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network: CNN), โครงข่ายประสาทเทียมแบบวนกลับ (Recurrent Neural Networks: RNN) เป็นต้น ในงานด้านการรู้จำรูปภาพ (Image recognition) มักจะนิยมนำโครงข่ายประสาทเทียมแบบสังวัตนาการมาประยุกต์ใช้



รูปที่ 7 รู้จำภาพโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ

จากรูปที่ 7 จะเห็นว่าแบบจำลองชนิดนี้จะแบ่งออกเป็นส่วนหลักๆ 2 ส่วนคือส่วนที่ใช้ในการสกัดคุณลักษณะเด่น และส่วนที่ใช้ในการจำแนกข้อมูล ในการสกัดคุณลักษณะเด่นจะประกอบด้วย 2 ส่วนหลักๆ คือชั้นคอนโวลูชันและชั้นการรวมข้อมูล (Pooling)

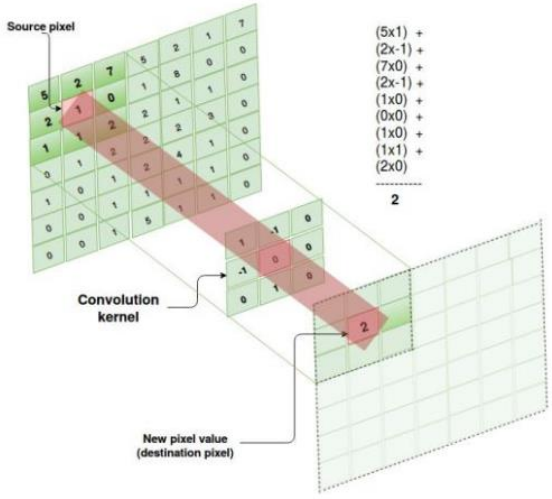
ในชั้นคอนโวลูชัน (Convolution layer) จะมีการทำคอนโวลูชัน ที่เป็นการคำนวณเชิงสเกลาร์ (Dot product) ระหว่างพื้นที่ส่วนย่อย (Local region) ของรูป ซึ่งโดยทั่วไปจะมีการเติมศูนย์ (Zero padding) กับตัวกรอกเชิงพื้นที่ (Spatial filter) หรือที่เรียกว่าเคอร์นอล (Kernel) ที่เป็นเมท

ริกซ์จัตุรัส (Square Matrix) เพื่อหารูปแบบของคุณสมบัติ (Feature map) ดังสมการที่ 3 และ 4 แสดงดังรูปที่ 8

$$G = h * F \quad (3)$$

$$G_{i,j} = \sum_{u=-k}^{k} \sum_{v=-k}^{k} h_{u,v} F_{i-u,j-v} \quad (4)$$

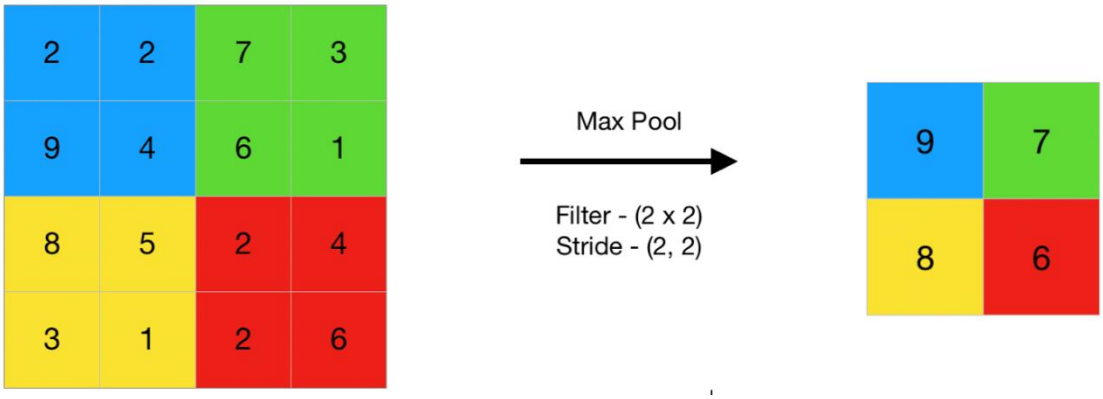
โดยที่ h คือรูปภาพ และ F คือเคอร์นอลขนาด k × k



รูปที่ 8 คอนโวลูชันแบบ 2 มิติ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

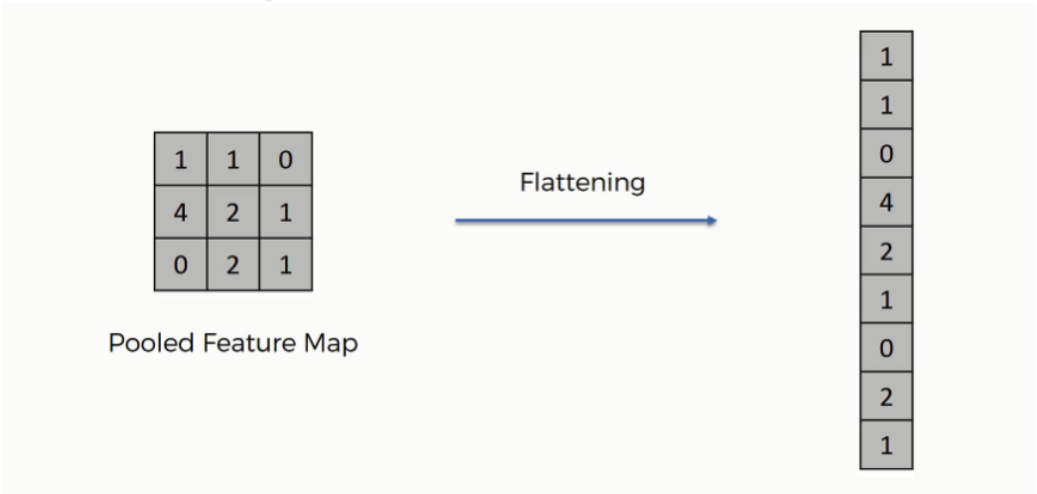
ชั้นรวมข้อมูล (Pooling layer) เป็นชั้นที่อยู่ระหว่างชั้นคอนโวลูชันกับชั้นคอนโวลูชัน หรือชั้นระหว่างชั้นลดมิติของข้อมูล (Flatten layer) มีหน้าที่ในการลดขนาด (Down sampling) ของรูปแบบคุณสมบัติของภาพ ซึ่งสามารถใช้ได้ทั้งฟังก์ชันหาค่าสูงสุด ค่าเฉลี่ย และค่าต่ำสุด โดยทั่วไปจะมีการใช้ฟังก์ชันหาค่าสูงสุด (Max pooling) เพื่อเลือกค่าที่สูงที่สุดของพื้นที่ย่อยของรูปแบบคุณสมบัติ ดังรูปที่ 9



รูปที่ 9 ผลลัพธ์ของฟังก์ชันหาค่าสูงสุด

โดยปกติในการสร้างแบบจำลองเพื่อสกัดคุณลักษณะเด่นจะมีการนำคู่ของชั้นคอนโวลูชันกับชั้นรวมข้อมูลมาต่อกัน โดยจะขึ้นอยู่กับความซับซ้อนของข้อมูล แต่ถ้าหากจำนวนคู่ของชั้นทั้งสองมีมากเกินไป อาจจะส่งผลให้เกิดปัญหาแบบจำลองเข้ากับข้อมูลสอนมากเกินไป (Overfitting) และหากน้อยเกินไปก็อาจทำให้เกิดปัญหาที่ไม่สามารถสกัดคุณลักษณะเด่นได้ดีพอ ซึ่งจะส่งผลต่อประสิทธิภาพของการจำแนกข้อมูล

เมื่อรูปภาพผ่านแบบจำลองสำหรับการสกัดคุณลักษณะเด่นแล้ว จะได้ผลลัพธ์ออกมาเป็นรูปแบบของคุณลักษณะเด่น ซึ่งจะต้องนำไปผ่านฟังก์ชันลดมิติของข้อมูล หรือชั้นลดมิติของข้อมูล เพื่อให้ได้เวคเตอร์ของคุณลักษณะ (Feature vector) ดังรูปที่ 10 เพื่อที่จะนำไปใช้เป็นข้อมูลขาเข้าของชั้นเชื่อมโยงแบบสมบูรณ์ (Fully connected layer) ที่ใช้ในการจำแนกข้อมูล



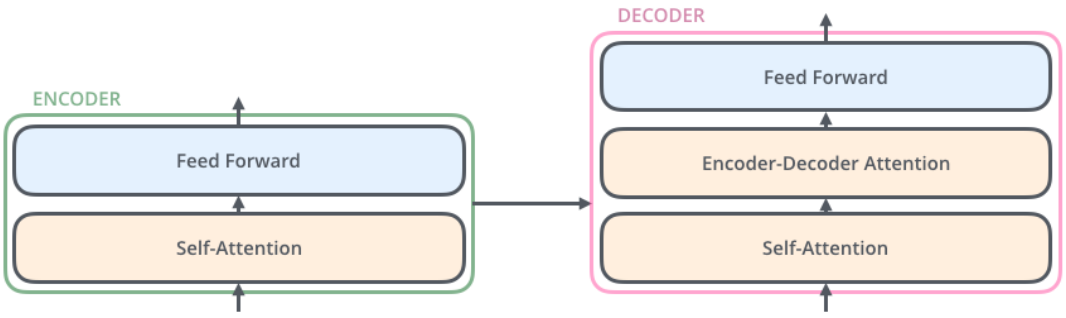
รูปที่ 10 ฟังก์ชันลดมิติข้อมูล

2.2.3 ทรานฟอร์มเมอร์

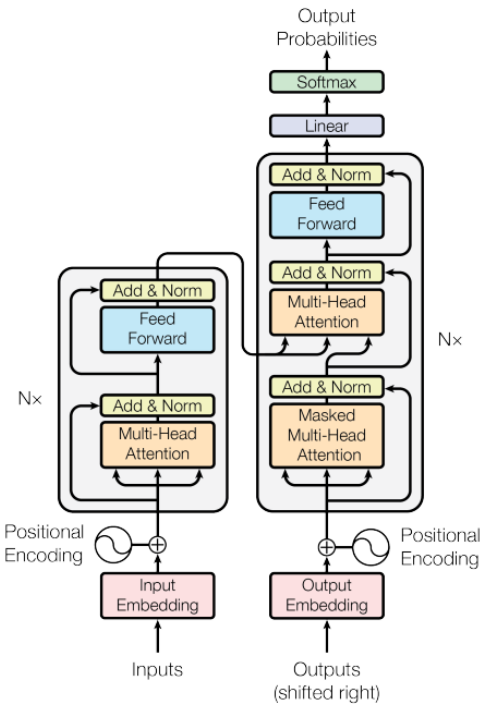
ทรานฟอร์มเมอร์เป็นแบบจำลองการเรียนรู้เชิงลึกประเภทหนึ่ง ที่ถูกพัฒนามาเพื่อใช้งานในด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ไม่ว่าจะเป็นงานทางด้านการแปลภาษาของเครื่อง (Machine translation) การตอบคำถาม (Question answering) เป็นต้น โดย A. Vaswani [26] ได้พัฒนาแบบจำลองที่มีพื้นฐานจากกลไกความสนใจ (Attention mechanism) และยังเป็นแบบจำลองที่ถูกสอนมาแล้ว (Pre-trained) โดยใช้ข้อมูลขนาด ผู้พัฒนาจึงสามารถนำไปใช้ต่อหรือนำไปปรับปรุงได้ทันที ทำให้ประหยัดทรัพยากรณ์ที่ใช้สำหรับสอนแบบจำลองไปได้

แนวคิดหลักของทรานฟอร์มเมอร์แบบดั้งเดิมคือการใช้กลไกความสนใจตนเอง (self-attention) ทำให้สามารถเห็นความสัมพันธ์ระหว่างข้อมูลได้

แบบจำลองทรานฟอร์มเมอร์ประกอบไปด้วย 2 ส่วนหลักคือส่วนเข้ารหัส (Encoder) และส่วนถอดรหัส (Decoder) โดยในแต่ละส่วนจะประกอบด้วยกลไกความสนใจตนเอง และโครงข่ายประสาทเทียมแบบป้อนข้อมูลไปข้างหน้า (Feed-forward neural network) แต่ในส่วนของตัวถอดรหัส ที่รับข้อมูลมาจากทั้งส่วนของอินพุตและจากตัวเข้ารหัส จะมีชั้นที่เรียกว่ากลไกความสนใจตัวเข้ารหัสและถอดรหัส (Encoder-decoder attention) ดังรูปที่ 11 และรูปที่ 12

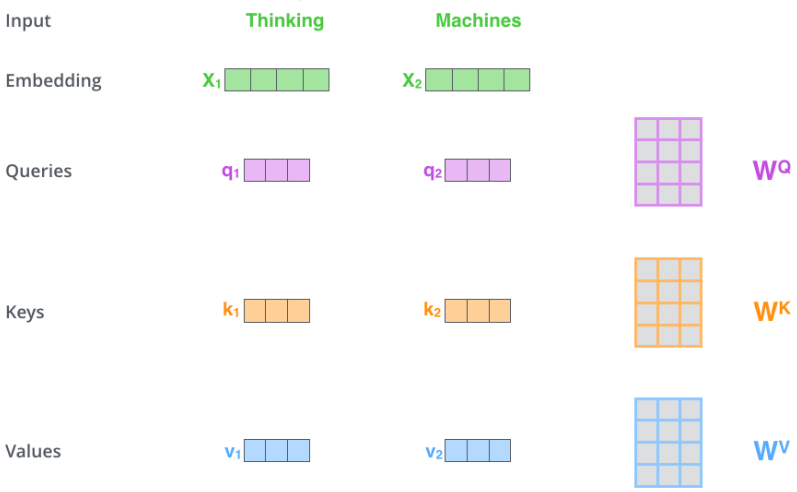


รูปที่ 11 โครงสร้างอย่างง่ายของทรานฟอร์มเมอร์แบบพื้นฐาน



รูปที่ 12 โครงสร้างของแบบจำลองทรานฟอร์มเมอร์

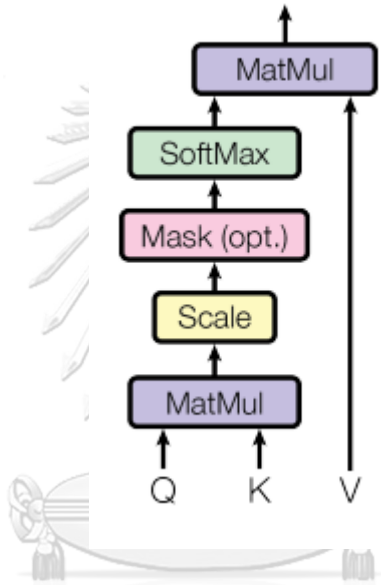
ในการคำนวณค่าความสนใจ (Attention) แบบจำลองจะนำอินพุตที่ผ่านการแปลงเป็นเวคเตอร์ (Embedding vector) มาใช้ในการสร้างเวคเตอร์ใหม่อีก 3 ชนิดได้แก่ q (Queries) k (Keys) และ v (Values) โดยการทำคูณกับค่าถ่วงน้ำหนักของเวคเตอร์แต่ละชนิด ดังรูปที่ 13



รูปที่ 13 การสร้างเวคเตอร์ q k และ v

ทรานฟอร์มเมอร์ใช้ความสนใจตนเองของผลคูณเชิงสเกลาร์ที่ปรับขนาด (Scaled dot-product self-attention) โดยการนำ q มาหาผลคูณเชิงสเกลาร์กับทุก k แล้วหารด้วยรากที่ 2 ของ d_k จากนั้นใช้ฟังก์ชัน SoftMax เพื่อที่จะหาความน่าจะเป็นของความสนใจ และปรับขนาดด้วยเวคเตอร์ v ดังสมการที่ 5 และรูปที่ 14

$$Attetion(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{d_k}}\right)v \quad (5)$$



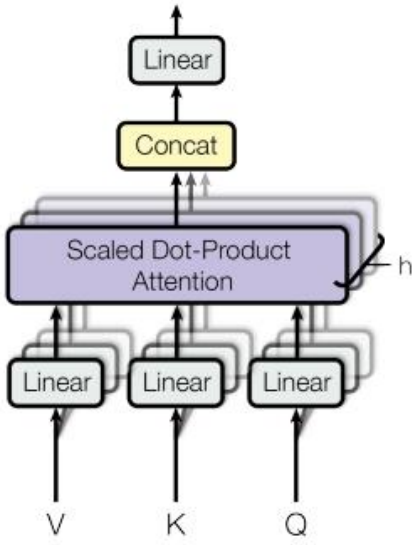
รูปที่ 14 block diagram ของ Scaled dot-product self-attention

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ในส่วนของความสนใจโดยรวม (Multi-head attention) เป็นการนำความสนใจเพียงอย่างเดียว (single self-attention) ที่เพิ่มตัวถ่วงน้ำหนัก W สำหรับเวคเตอร์ q k และ v เข้าไปแล้วนำความน่าสนใจต่างมารวมกันเพื่อที่จะนำคูณกับค่าถ่วงน้ำหนักโดยรวมอีกครั้ง ทำให้แบบจำลองสามารถหาความสัมพันธ์ระหว่างข้อมูลที่ต่างกันในตำแหน่งที่ต่างจากเดิม ดังสมการที่ 6 และ 7 และดังรูปที่ 15

$$MultiHead(Q,K,V) = Concat(head_1, head_2, \dots, head_h)W^0 \quad (6)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (7)$$



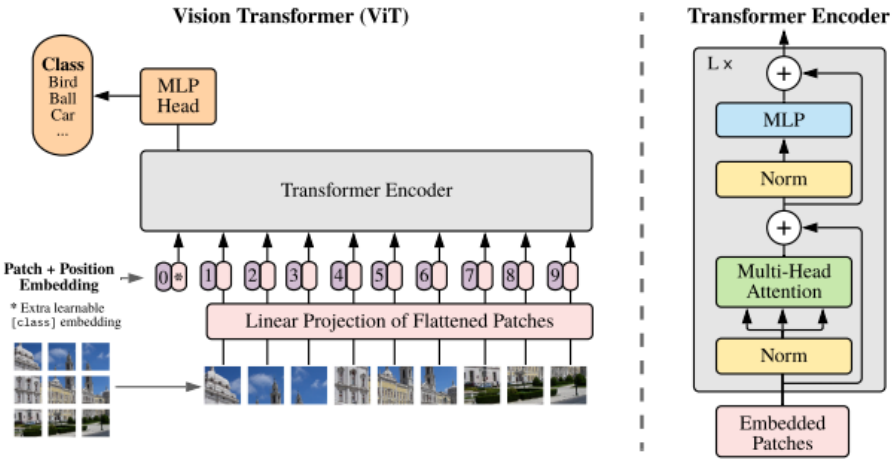
รูปที่ 15 block diagram ของ Multi-Head attention

A. Dosovitskiy [27] ได้นำเสนอวิธีการทำการจดจำรูปภาพโดยใช้ทรานฟอร์มเมอร์ โดยการแบ่งภาพออกเป็นส่วนย่อยๆ (Patching) เพื่อใช้ในการสร้างเวคเตอร์ของส่วนของภาพ (Patch embedding) ที่จะนำไปใช้เป็นอินพุตสำหรับส่วนเข้ารหัสของทรานฟอร์มเมอร์ ผลลัพธ์ที่ได้จากส่วนเข้ารหัสจะไม่ได้นำไปเป็นอินพุตของส่วนถอดรหัสเหมือนกับทรานฟอร์มเมอร์ทั่วไป แต่จะถูกนำไปใช้เป็นอินพุตสำหรับโครงข่ายประสาทแบบเพอร์เซปตรอนหลายชั้น (Multi-layer perceptron) เพื่อจำแนกประเภทของข้อมูล ดังรูปที่ 16 เมื่อเทียบกับแบบจำลองโครงข่ายประสาทเทียมแบบสังวัตนาการแบบจำลองทรานฟอร์มเมอร์ในส่วนของชั้นการสร้างเวคเตอร์ของส่วนของภาพและชั้นการเข้ารหัสข้อมูล เปรียบเสมือนเป็นชั้นการสกัดคุณลักษณะเด่นของข้อมูล ก่อนที่จะจำแนกประเภทของข้อมูลด้วยโครงข่ายประสาทเทียม

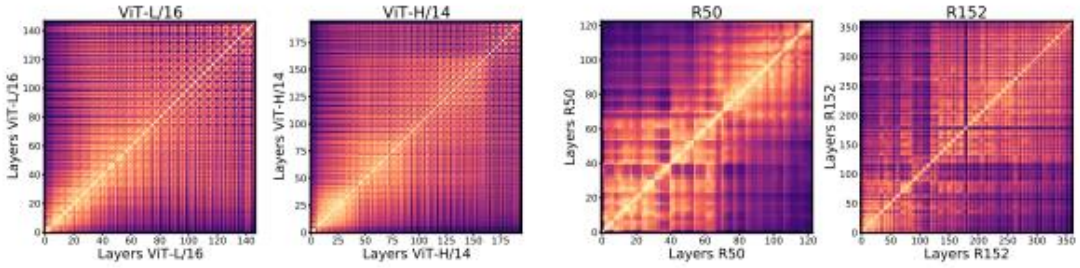
M. Raghu [28] เปรียบเทียบการสกัดคุณลักษณะเด่นของข้อมูลโดยใช้ Vision Transformer (ViT) [27] และโครงข่ายประสาทเทียมแบบสังวัตนาการแบบ ResNet [17] โดยใช้ Central Kernel Alignment (CKA) จากการเปรียบเทียบ รูปแบบการเรียนรู้ของแบบจำลองทั้งสองนั้นแตกต่างกัน โดยลักษณะเด่นของข้อมูลที่สกัดจาก ViT จะมีการกระจายตัวที่สม่ำเสมอกันในทุกชั้นของแบบจำลอง ต่างจากลักษณะเด่นของข้อมูลที่ถูกสกัดจาก ResNet จะมีความแตกต่างกันอย่างชัดเจนดังรูปที่ 17 ซึ่งคุณสมบัติดังกล่าวส่งผลให้แบบจำลองแบบ ViT สามารถเรียนรู้ข้อมูลภาพที่มรการกระจายตัวของค่าพิกเซลล์เท่าๆกันในแต่ละส่วนของภาพได้ดีกว่า เช่นภาพฟิล์มเลือดเป็นต้น

K. Patel [29] เสนอวิธีการแบ่งภาพของ ViT เพื่อการเพิ่มประสิทธิภาพของการสกัด

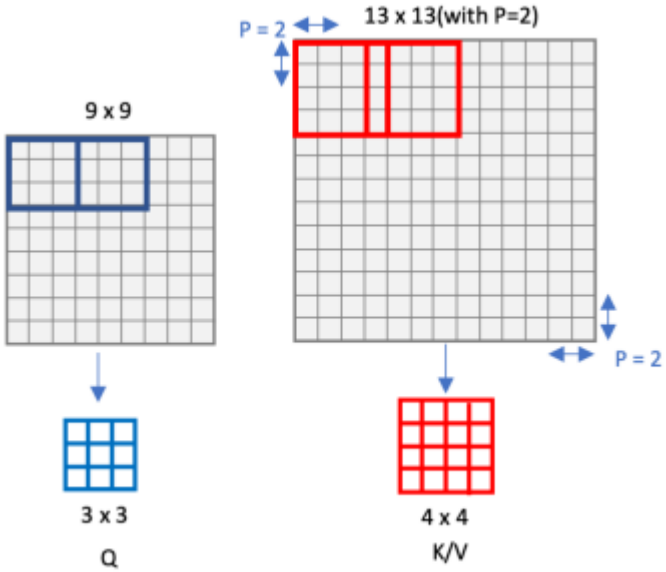
คุณลักษณะเด่นของ ViT โดยแบ่งภาพให้มีส่วนที่ซ้อนทับกัน ซึ่งจะทำให้ลักษณะเด่นที่สกัดมามีความต่อเนื่องกันมากยิ่งขึ้น จากการทดลองการแบ่งภาพให้มีส่วนที่ซ้อนทับกัน 17% ของพื้นที่ภาพย่อยจะให้ผลลัพธ์ที่ดีที่สุดทั้งในเชิงของความแม่นและความซับซ้อนในการคำนวณ โดยได้เปรียบเทียบกับสัดส่วนการซ้อนทับที่ 33%, 50%, และ 66%



รูปที่ 16 block diagram ของแบบจำลอง vision transformer



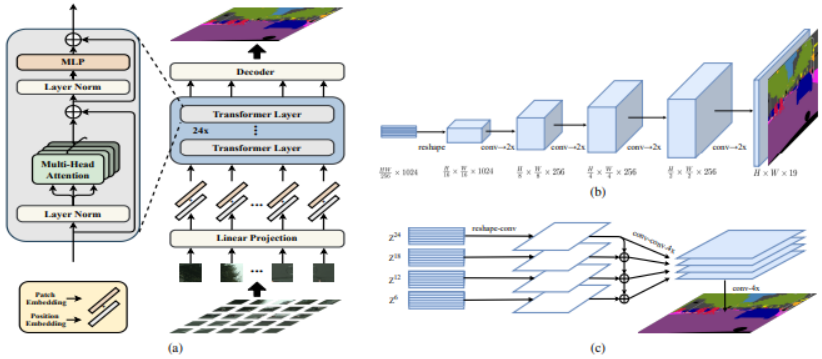
รูปที่ 17 การกระจายตัวของคุณลักษณะเด่นของข้อมูลที่ถูกสกัดจากแบบจำลอง ViT และ ResNet



รูปที่ 18 เปรียบเทียบการแบ่งภาพของ ViT แบบธรรมดาและแบบซ้อนทับ

ภายหลัง นักวิจัยได้มีการนำทรานฟอร์มเมอร์สำหรับการจดจำรูปภาพไปประยุกต์ใช้กับงานในด้าน [30] การสังเคราะห์รูปภาพ (Image generation), [31] การสังเคราะห์รูปภาพจากข้อความ (Text to image synthesis), [32] การทำความเข้าใจวิดีโอ (Video understanding) รวมทั้ง [33] การตรวจจับวัตถุหรือ [34] [35] การแบ่งภาพเชิงความหมายเป็นต้น

S. Zheng [35] ได้นำเสนอวิธีการที่ใช้ทรานฟอร์มเมอร์ในการแบ่งภาพเชิงความหมายโดยการเพิ่มขั้นตอนการสร้างเวคเตอร์ของตำแหน่ง (Position embedding) รวมกับเวคเตอร์ของส่วนย่อยของภาพเพื่อเป็นอินพุตสำหรับส่วนการเข้ารหัสของทรานฟอร์มเมอร์ และเพิ่มส่วนถอดรหัสที่เป็นโครงข่ายประสาทเทียมแบบสังวัตนาการเข้ามาแทนที่เพอร์เซปตรอนหลายชั้น เพื่อใช้ในการเพิ่มขนาดของภาพผลลัพธ์ (Up-sampling) เพื่อให้ได้ขอบเขต (Region) ของส่วนที่ต้องการแบ่งในขนาดเดียวกับภาพที่ต้องการแบ่ง ดังรูปที่ 17



รูปที่ 19 block diagram ของ SEgmentation TRansformer (SETR)

2.2.4 การประเมินผลแบบจำลอง

การประเมินผลแบบจำลอง (Model evaluation) เป็นส่วนสำคัญของกระบวนการพัฒนาแบบจำลอง ที่จะสามารถช่วยเราในการเลือกแบบจำลองที่เหมาะสมกับชุดข้อมูลของเรา และแบบจำลองที่เลือกจะทำงานได้ดีเพียงใน ในทางวิทยาการข้อมูล (Data science) จะใช้การประเมินผลแบบจำลอง 2 วิธีคือวิธีการ Hold-Out และวิธีการ Cross-Validation

ในวิธี Hold-Out ชุดข้อมูลจะถูกแบ่งออกเป็น 3 ส่วนได้แก่

1. ชุดข้อมูลสำหรับสอนแบบจำลอง เป็นชุดข้อมูลที่ใช้ในการสร้างแบบจำลอง

2. ชุดข้อมูลสำหรับวัดผลแบบจำลอง เป็นชุดข้อมูลที่ใช้ในการวัดผลแบบจำลองในขณะที่สอน เพื่อปรับปรุงพารามิเตอร์ต่างๆให้แบบจำลองทำงานได้ดีขึ้น

3. ชุดข้อมูลสำหรับทดสอบแบบจำลอง เป็นชุดข้อมูลที่ใช้ในการทดสอบแบบจำลองหลังการสอน เพื่อประเมินประสิทธิภาพของแบบจำลอง

ในการสอนแบบจำลองด้วยข้อมูลจำนวนที่มีไม่มากพอ จะมีโอกาสทำให้การประเมินผลแบบจำลองเกิดความเอนเอียง (Biased model) ซึ่งการทำ k-fold cross-validation จะสามารถนำมาใช้แก้ปัญหานี้ได้โดยมีขั้นตอนดังนี้

1. แบ่งชุดข้อมูลออกเป็น k กลุ่ม

2. ทำการสอนโมเดลใหม่และประเมินผล k ครั้ง โดยใช้ชุดข้อมูลสำหรับสอน k-1 กลุ่ม ส่วนอีก 1 กลุ่มที่เหลือจะนำไปใช้ในการประเมินผล

หากใช้ k เท่ากับจำนวนข้อมูลทั้งหมด จะเรียกวิธีนี้ว่า leave-one-out

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปที่ 20 การทำ k-fold cross-validation

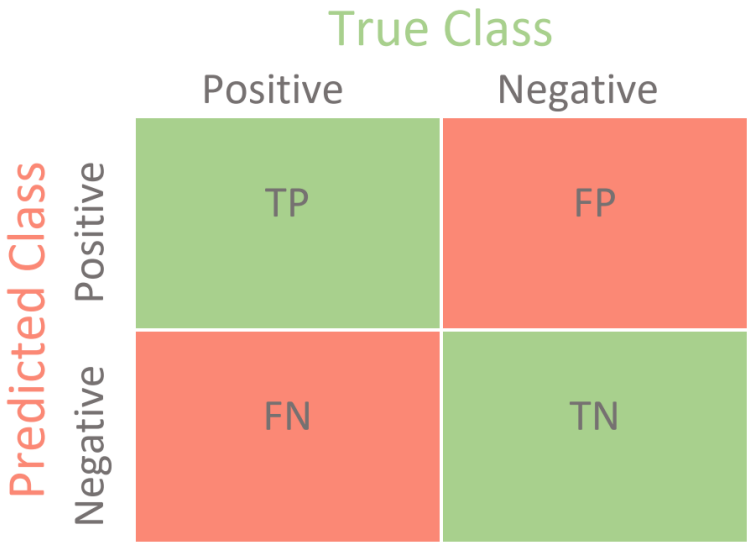
สำหรับการประเมินผลแบบจำลองสำหรับการจำแนกข้อมูล วิธีการคำนวณความแม่นยำและความคลาดเคลื่อน จะใช้ข้อมูลจริง (Actual) และค่าพยากรณ์ (Predicted) ซึ่งจะถูกบันทึกไว้ใน Confusion matrix ซึ่งข้อมูลจะถูกแบ่งออกเป็น 4 ส่วนดังรูปที่ 19 ได้แก่ True positive (TP), False positive (FP), False negative (FN) และ True negative (TN) โดย

1. True positive คือจำนวนข้อมูลที่แบบจำลองพยากรณ์ได้อย่างถูกต้อง โดยที่ข้อมูลนั้นมีค่าความจริงเป็นจริง และแบบจำลองพยากรณ์ว่าเป็นจริง เช่นในแบบจำลองสำหรับการทำนายว่าใช่ภาพสุนัขหรือไม่ ข้อมูลที่ใส่มาเป็นภาพสุนัข และแบบจำลองพยากรณ์ว่าเป็นภาพสุนัข

2. False positive คือจำนวนข้อมูลที่มีค่าความจริงเป็นจริง แต่แบบจำลองพยากรณ์ผิดว่าเป็นเท็จ เช่นข้อมูลเป็นภาพสุนัข แต่แบบจำลองพยากรณ์ว่าไม่ใช่สุนัข

3. False negative คือจำนวนข้อมูลที่แบบจำลองพยากรณ์ได้อย่างถูกต้อง โดยที่ข้อมูลนั้นมีค่าความจริงเป็นเท็จ และแบบจำลองพยากรณ์ว่าเป็นเท็จ เช่นข้อมูลเป็นไม่ใช่ภาพสุนัข และแบบจำลองพยากรณ์ว่าไม่ใช่สุนัข

4. True negative คือจำนวนข้อมูลที่ค่าความจริงเป็นเท็จ แต่แบบจำลองพยากรณ์ว่าเป็นจริง เช่นข้อมูลไม่ใช่ภาพสุนัข แต่แบบจำลองพยากรณ์ว่าเป็นภาพสุนัข



รูปที่ 21 confusion matrix

ค่าความแม่นยำ (Accuracy) สามารถคำนวณได้ดังสมการที่ 8

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

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

โดยปกติแล้ว ในการประเมินผลแบบจำลองสำหรับการแบ่งภาพเชิงความหมายและการตรวจจับวัตถุ มักจะใช้การ Intersection over Union (IoU) มาเป็นหน่วยในการวัดผล ซึ่งเป็นการนำขอบเขตจริงมาทำการดำเนินการเชิงบิต (Bitwise operation) ดังสมการที่ 9

$$IoU = \frac{target \cap prediction}{target \cup prediction} \quad (9)$$

ซึ่งสามารถคำนวณได้โดยการใช้ไลบารี่ Numpy ในภาษา Python ได้โดยง่ายดังรูปที่ 20

```
intersection = np.logical_and(target, prediction)
union = np.logical_or(target, prediction)
iou_score = np.sum(intersection) / np.sum(union)
```

รูปที่ 22 ตัวอย่างการเขียนโปรแกรมด้วยภาษา Python เพื่อคำนวณค่า IoU

โดยผลลัพธ์ที่ได้จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยยิ่งค่า IoU มีค่ามากก็หมายความว่าแบบจำลองมีประสิทธิในการทำงานของชุดทดสอบมากด้วยเช่นกัน

ค่าความเที่ยงตรง (Precision) เป็นวิธีการบ่งบอกแนวโน้มการรวมกลุ่มกันของค่าพยากรณ์ที่ถูกต้องของคลาสบวก ว่ามีค่าเป็นเท่าไรของกลุ่มข้อมูลคลาสบวกทั้งหมด หากค่า Precision มีค่าเข้าใกล้ 1 แสดงว่าโมเดลโครงข่ายประสาทนั้นมีความเที่ยงตรงในการพยากรณ์ค่าของคลาสบวกได้ดี

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

ค่าการเรียกคืน (Recall) เป็นวิธีการบ่งบอกความสามารถในการพยากรณ์ได้ถูกต้องของคลาสบวก หากค่า Recall มีค่าเข้าใกล้ 1 แสดงว่าโมเดลโครงข่ายประสาทนั้นมีจำนวนของการพยากรณ์ถูกต้องสูง นั้นคือหากถูกระบุว่าเป็นพบขโมย ความเป็นจริงต้องเจอขโมย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

การบ่งบอกความสามารถด้วยค่า Average Precision (AP) จะเป็นการบอกด้วยค่า Average Precision (AP) ด้วยการหาพื้นที่ใต้โค้ง (Under curve area) ระหว่าง ค่าความเที่ยงตรง (Precision) และ ค่าการเรียกคืน (Recall) มีค่าอยู่ระหว่าง 0 ถึง 1 หากค่า AP มีค่าเข้าใกล้ 1 แสดงว่าโมเดลโครงข่ายประสาทนั้นสามารถในการพยากรณ์ได้อย่างถูกต้องแม่นยำ

$$AP = \int_0^1 p(r)dr = \sum_{i=1}^{n} P(i)\Delta Re(i) \quad (12)$$

เมื่อ AP คือพื้นที่ใต้โค้งของค่าความเที่ยงกับค่าการเรียกคืน $p(r)dr$ คือฟังก์ชันความสัมพันธ์ระหว่างค่าความเที่ยงกับค่าการเรียกคืน n คือจำนวนจุดตัดข้อมูลในช่วงกราฟ $P(i)$ คือช่วงค่าความเที่ยงตรงในพื้นที่ใต้โค้ง และ $\Delta Re(i)$ คือการเปลี่ยนแปลงของค่าการเรียกคืน

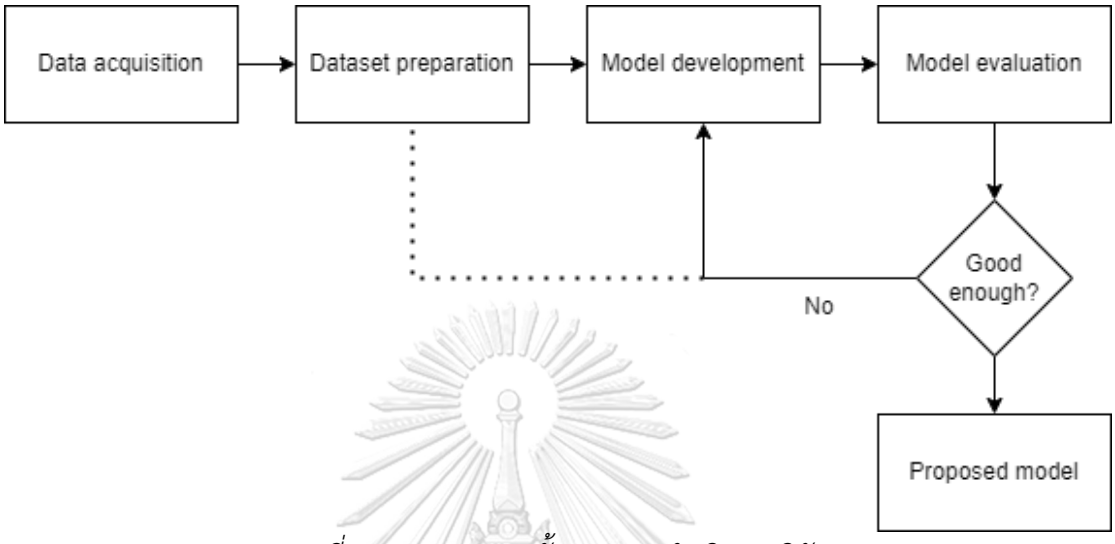
การบ่งบอกความสามารถด้วยค่า Mean Average Precision (mAP) เป็นการประเมินความสามารถของแบบจำลองโดยการนำค่า average precision ของแต่ละประเภทมาหาค่าเฉลี่ย หาก mAP มีค่าเข้าใกล้ 1 แสดงว่าโมเดลโครงข่าย ประสาทนั้นมีความสามารถในการพยากรณ์ได้อย่างถูกต้องแม่นยำในทุกประเภท

$$mAP = \frac{1}{n_{classes}} \sum_{i=1}^{n_{classes}} AP_i \quad (13)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3 ขั้นตอนวิธีที่นำเสนอ

3.1 ขั้นตอนวิธีโดยรวม



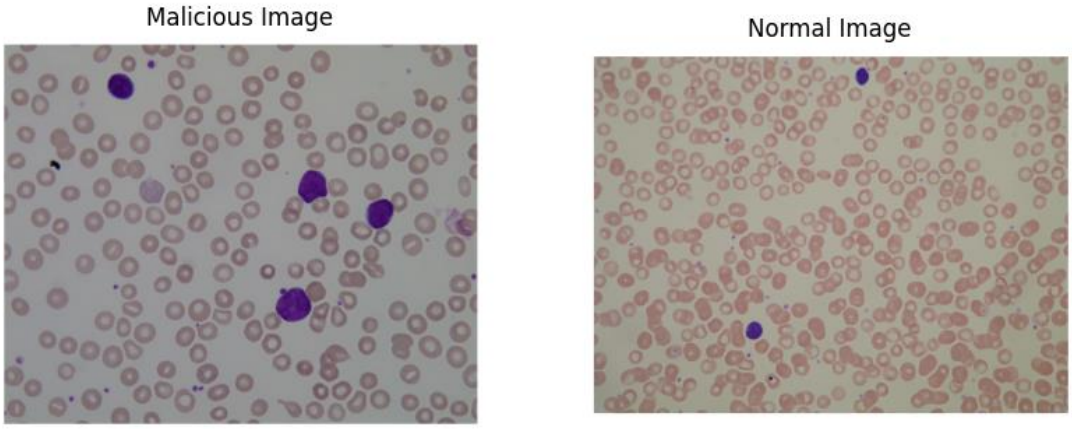
รูปที่ 23 ภาพรวมของขั้นตอนการดำเนินงานวิจัย

จากรูปที่ 23 ที่แสดงถึงขั้นตอนต่างๆในการดำเนินงานวิจัยซึ่งประกอบไปด้วยการเก็บรวมรวบข้อมูล (Data acquisition) การเตรียมชุดข้อมูล (Dataset preparation) การพัฒนา แบบจำลอง (Model development) และการประเมินผลของแบบจำลอง (Model evaluation)

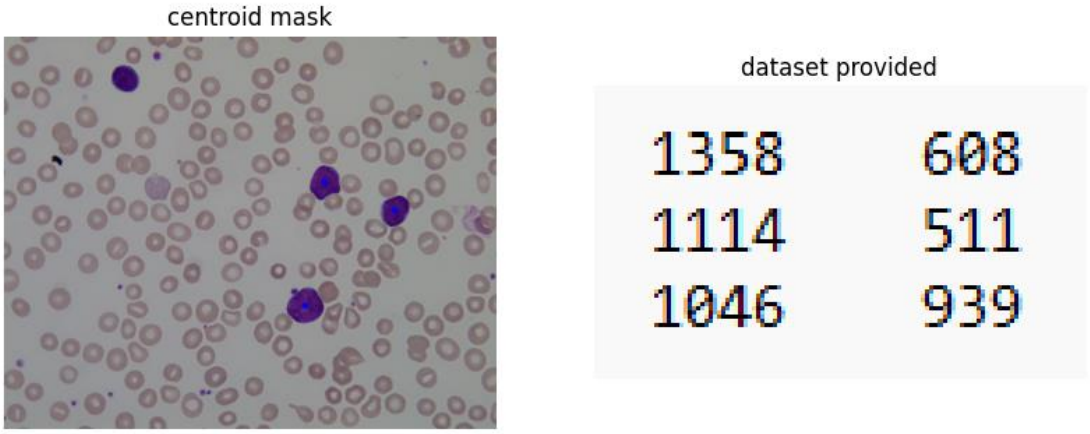
3.2 การเก็บรวมรวบข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในการพัฒนาแบบจำลองเพื่อการแบ่งภาพเซลล์ลิมโฟบลาสท์ ผู้วิจัยได้ใช้ชุดข้อมูลเปิด ALL-IDB ซึ่งประกอบด้วยภาพฟิล์มเลือดจำนวน 108 ภาพ โดยภาพทั้งหมดนี้ถูกถ่ายด้วยกล้องจุลทรรศน์ที่กำลังขยายตั้งแต่ 300 – 500 เท่า ตัวไฟล์ภาพถูกจัดเก็บในรูปแบบ JPEG ที่ความละเอียด 2592 x 1944 พิกเซล ผู้ที่รวมรวบข้อมูลของชุดข้อมูลนี้ได้แบ่งประเภทของข้อมูลออกเป็น 2 ชนิดได้แก่ 1. ภาพฟิล์มเลือดที่ถูกเก็บจากผู้ป่วยมะเร็งเม็ดเลือดขาว และ 2.ภาพฟิล์มเลือดที่ถูกเก็บจากคนสุขภาพดี โดยในภาพฟิล์มเลือดที่เก็บจากผู้ป่วยมะเร็งเม็ดเลือดขาวจะมีเซลล์ลิมโฟบลาสท์อยู่ภายในภาพ โดยที่ตำแหน่งของเซลล์ลิมโฟบลาสท์ได้ถูกระบุไว้ด้วยตำแหน่งของจุดกึ่งกลางของแต่ละเซลล์



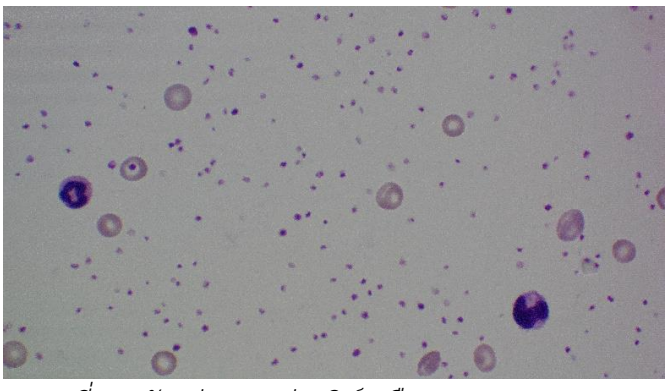
รูปที่ 24 ตัวอย่างภาพจากชุดข้อมูล ALL-IDB ในแต่ละประเภท



รูปที่ 25 การระบุตำแหน่งเซลล์ลิมโฟไซท์โดยผู้เชี่ยวชาญที่ทำการรวมรวบข้อมูล ALL-IDB โดยรูปซ้ายคือตำแหน่งในแนวแกน xy และรูปขวาแสดงการวาดจุดสีน้ำเงินตามคู่ลำดับ

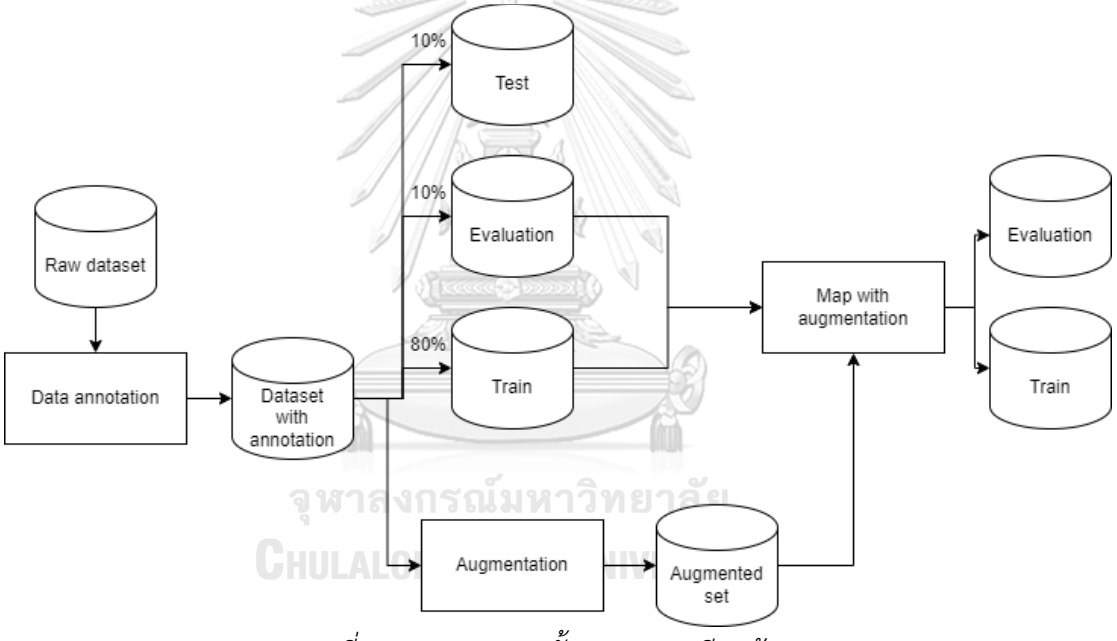
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในการพัฒนาแบบจำลองสำหรับการจำแนกและแบ่งภาพของเซลล์เม็ดเลือดแดง และเซลล์เม็ดเลือดขาวผู้วิจัยได้ใช้ชุดข้อมูลภาพถ่ายแผ่นนสไลด์ฟิล์มเลือดแบบ Buffy coat ย้อมสีแบบ Giemsa โดยบริษัท ward's science ถ่ายด้วยกล้องจุลทรรศน์ Olympus CX33 กำลังขยายของเลนส์ใกล้วัตุที่ 40 เท่า จำนวน 200 ภาพ ซึ่งถูกถ่ายโดยนายณัทกร เกษมสำราญ



รูปที่ 26 ตัวอย่างภาพถ่ายฟิล์มเลือดแบบ Buffy coat

3.3 การเตรียมชุดข้อมูล



รูปที่ 27 ภาพรวมของขั้นตอนการเตรียมข้อมูล

จากรูปที่ 27 ภาพถ่ายทั้งหมดจะถูกนำไปใส่คำอธิบายข้อมูล (Data annotate) จากนั้นผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น 3 ชุดได้แก่ชุดข้อมูลสำหรับการสอนแบบจำลอง ชุดข้อมูลสำหรับการวัดผลแบบจำลองขณะสอน และชุดข้อมูลสำหรับการมดสอบแบบจำลอง ในอัตราส่วนร้อยละ 80 10 และ 10 ตามลำดับ และทำการขยายชุดข้อมูลทั้งหมด ชุดข้อมูลสำหรับสอนแบบจำลองและชุดข้อมูลสำหรับทดสอบแบบจำลองขณะสอนจะถูกนำมารวมกับชุดข้อมูลที่ถูกสร้างขึ้นใหม่ โดยจะเพิ่มเฉพาะภาพที่ถูกสร้างจากภาพต้นฉบับที่อยู่ในชุดข้อมูลสำหรับสอนแบบจำลองและชุดข้อมูลสำหรับทดสอบแบบจำลองขณะสอนเท่านั้น

3.3.1 การใส่คำอธิบายข้อมูล

ในการพัฒนาแบบจำลองการเรียนรู้เชิงลึก โดยปกติเราจำเป็นที่จะต้องใช้ผลเฉลย (ground truth) ในการสอนแบบจำลอง ซึ่งในการพัฒนาแบบจำลองสำหรับการแบ่งภาพเชิงความหมาย โดยปกติแล้วจะใช้ผลเฉลยที่เป็นภาพที่ถูกแบ่งแล้ว (segmentation map) ผู้วิจัยได้ทำการระบุข้อมูลโดยการระบุขอบเขตของเซลล์ในโปรแกรม Computer Vision Annotation Tools (CVAT) [36] และส่งออกผลการระบุข้อมูลออกมาเป็นไฟล์สกุล JSON ตามโครงสร้างของ COCO (Common Object in Context) [37]

```
categories = [{
    "id": int,
    "name": str,
    "supercategory": str,
}]
image = {
    "id": int,
    "width": int,
    "height": int,
    "file_name": str,
}
annotation = {
    "id": int,
    "image_id": int,
    "category_id": int,
    "segmentation": RLE or [polygon],
    "area": float,
    "bbox": [x,y,width,height],
    "iscrowd": 0 or 1,
}
```

รูปที่ 28 รูปแบบโครงสร้างข้อมูลของ COCO

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

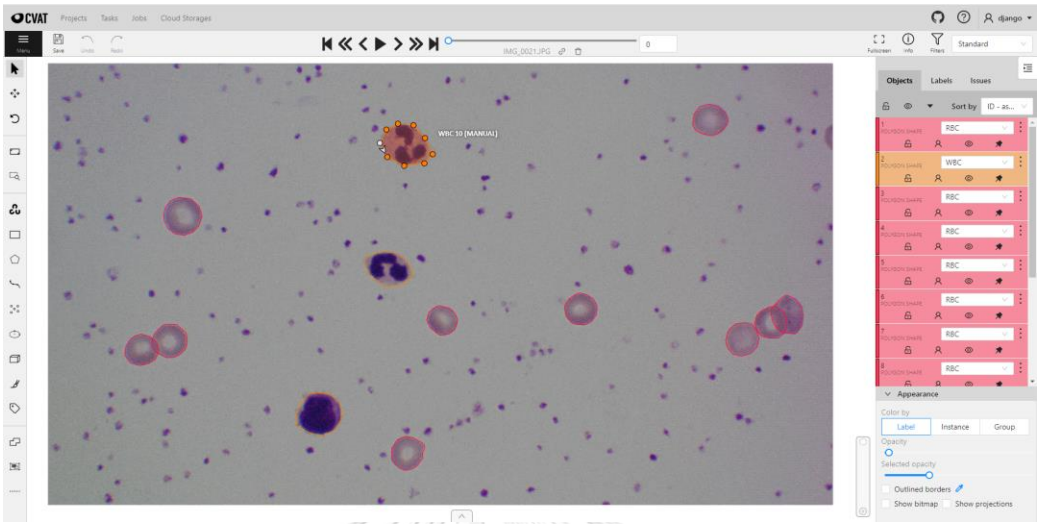
จากรูปที่ 28 คำอธิบายภาพรูปแบบ COCO สำหรับงานด้านการแบ่งภาพเชิงความหมายจะแบ่งออกเป็น 3 ส่วนหลักๆคือ

1. categories ที่จะเก็บชื่อของชนิดของข้อมูลไว้ใน name และเก็บรหัสไว้ใน id

2. Image จะเก็บรหัสไว้ใน id และเก็บชื่อไฟล์ไว้ใน file_name อีกทั้งยังเก็บวความกว้างและความสูงของภาพไว้ใน width และ height ตามลำดับ

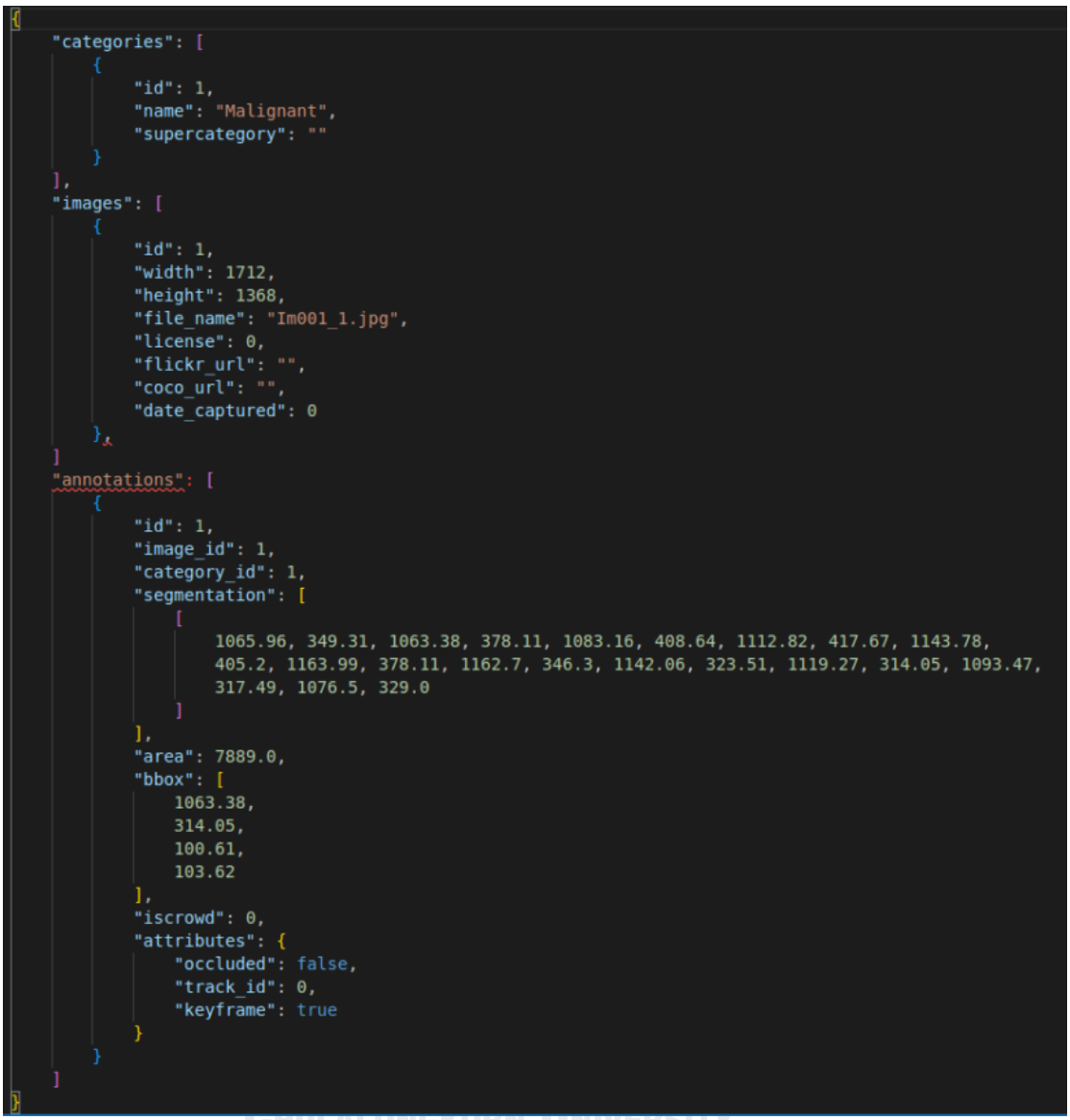
3. Annotation เป็นส่วนที่เก็บคำอธิบายข้อมูลไว้ ซึ่งจะเก็บรหัสของคำอธิบายไว้ที่ id เก็บรหัสของรูปที่อธิบายไว้ที่ image_id เก็บรหัสของชนิดของข้อมูลไว้ที่ category_id สำหรับการแบ่งภาพ จะเก็บขอบเขตไว้ใน segmentation ซึ่งสามารถเก็บได้ 2 แบบคือ Run Length Encoding (RLE) หรือการเก็บแบบ polygon เก็บขนาดของพื้นที่ของข้อมูลไส้ที่ area และเป็นตำแหน่งของจุดบนซ้ายพร้อมทั้งความกว้าง และความสูงของกล่องขอบเขตวัตถุ (bounding box) ไว้ใน bbox

ในรูปที่ 29 และ 30 เป็นตัวอย่างของการอธิบายรูปภาพในเครื่องมือ CVAT โดยการบันทึกในรูปแบบ COCO 1.0



รูปที่ 29 ตัวอย่างการวาด polygon รอบขอบเขตที่สนใจบนเครื่องมือ CVAT

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



CHULALONGKORN UNIVERSITY

รูปที่ 30 ตัวอย่างอย่างส่วนหนึ่งของไฟล์คำอธิบายรูปภาพสำหรับการแบ่งภาพเชิงความหมาย ในรูปแบบ COCO

ในการสร้างภาพผลเฉลยสำหรับการแบ่งภาพเชิงความหมาย ผู้วิจัยได้ใช้ชุดเครื่องมือ pandas, cv2, และ pycocotools ซึ่งเป็นชุดเครื่องมือในภาษา Python สำหรับการสร้างคำสั่ง create_class_csv และ create_segmentation_mask

create_class_csv มีไว้สำหรับการสร้างไฟล์ที่ทำหน้าที่เก็บข้อมูลประเภทของวัตถุในภาพ โดยจะเก็บชื่อ, ลำดับ, และค่าในภาพผลเฉลยของแต่ละประเภท

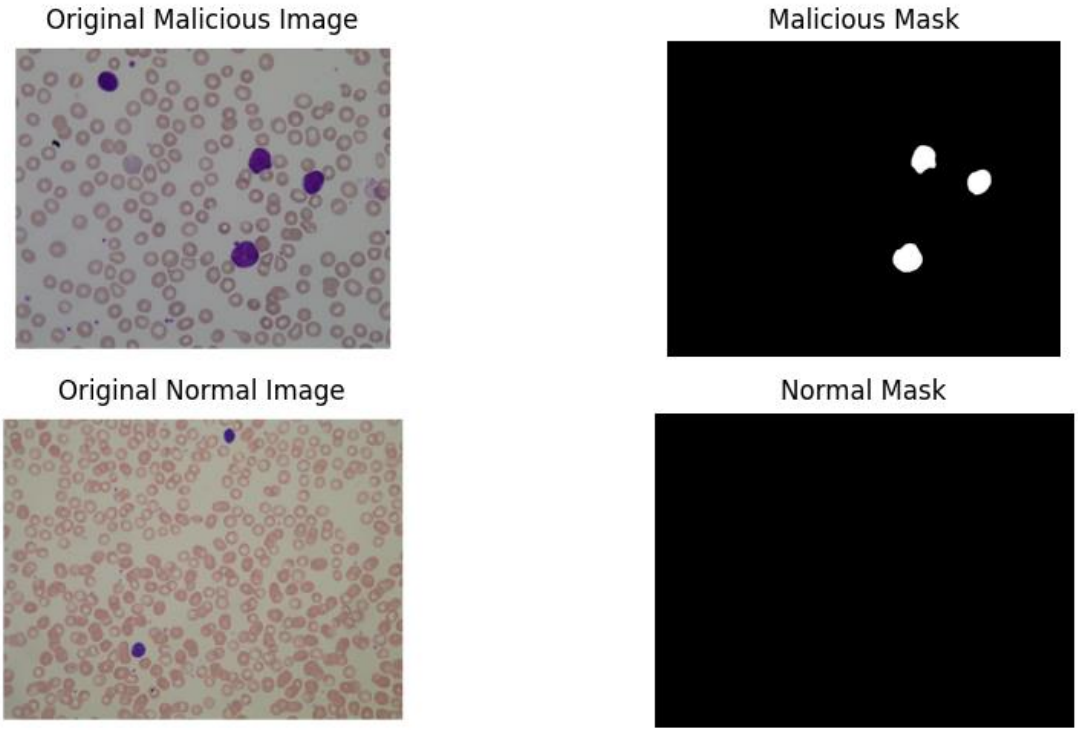
ส่วนคำสั่ง create_segmentation_mask มีไว้สำหรับการสร้างภาพผลเฉลยจากภาพจริงและไฟล์คำอธิบายชุดข้อมูล

```python
def create_class_csv(annotation_file: str):
    f = open(annotation_file)
    dataset_coco = json.load(f)

    annotations_dataframe = pd.DataFrame({"Class ID": [0],"Pixel Value": [0], "Class": ["background"]})

    class_name = [item["name"] for item in dataset_coco["categories"]]
    class_id = [item["id"] for item in dataset_coco["categories"]]
    class_value = [int(255/id) for id in class_id]

    annotations_dataframe = pd.concat([annotations_dataframe,
                                       pd.DataFrame({"Class ID": class_id,"Pixel Value": class_value, "Class": class_name})],
                                       ignore_index=True)

    dirname = os.path.dirname(annotation_file)
    annotations_dataframe.to_csv(os.path.join(dirname, "classes.csv"))
```

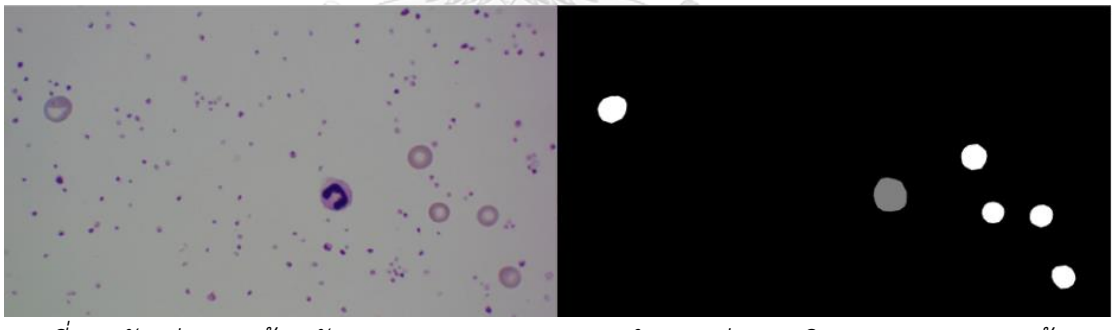
รูปที่ 31 คำสั่ง create_class_csv

```python
def create_segmentation_mask(image_dir: str, annotation_file: str):
    coco = COCO(annotation_file)

    save_path = os.path.join(os.path.dirname(image_dir), "masks")
    if os.path.exists(save_path) == False:
        os.makedirs(save_path, exist_ok=True)

    img_list = coco.imgs
    cat_ids = coco.getCatIds() # Load annotation classes
    pixel_value = [int(255/cat) for cat in cat_ids]

    for idx in range(1, len(img_list)+1):
        # Load image infomations
        coco_img = img_list[idx]
        # Load image annotation infomation
        img_anns = coco.getAnnIds(imgIds=coco_img["id"], catIds=cat_ids, iscrowd=None)
        anns = coco.loadAnns(img_anns)

        # Create segmentation mask
        mask = np.zeros((coco_img["height"], coco_img["width"]))

        for i in range(len(anns)):
            mask = np.maximum(mask, coco.annToMask(anns[i]))

        for i in range(len(cat_ids)):
            mask[mask==cat_ids[i]] = pixel_value[i]

        # Save segmentation mask
        mask = Image.fromarray(mask)
        mask = mask.convert("L")
        mask.save(os.path.join(save_path, f"{coco_img['file_name'].split('.')[0]}.png"), "PNG")
```

รูปที่ 32 คำสั่ง create_segmentation_mask



รูปที่ 33 ตัวอย่างภาพต้นฉบับและภาพผลเฉลยของการทำการแบ่งภาพเชิงความหมายของชุดข้อมูล ALL-IDB



รูปที่ 34 ตัวอย่างภาพต้นฉบับและภาพผลเฉลยของการทำการแบ่งภาพเชิงความหมายของชุดข้อมูล ภาพถ่าย Buffy coat

3.3.2 การขยายขนาดของชุดข้อมูล

การขยายขนาดของชุดข้อมูลเป็นขั้นตอนที่ทำการสร้างข้อมูลใหม่ ซึ่งในที่นี้คือรูปภาพ เพื่อให้ชุดข้อมูลมีการกระจายตัวและมีความหลากหลายมากยิ่งขึ้น อีกทั้งการสอนแบบจำลองด้วยข้อมูลที่มีจำนวนน้อยเกินไปอาจทำให้แบบจำลองเกิดภาวะเข้ากันได้กับชุดข้อมูลฝึกมากเกินไป (Overfitting) ซึ่งการเพิ่มขนาดของข้อมูลจะสามารถช่วยลดอัตราการเกิดปัญหานี้ลงได้ ในงานวิจัยนี้ ผู้วิจัยได้ใช้กล

ยุทธการขยายขนาดของชุดข้อมูลที่ใช้วิธีการประมวลผลรูปภาพทั้งในเชิงพื้นที่ (Spatial transform) และในเชิงความถี่ (Frequency domain transform) ดังตารางที่ 1 โดยใช้ชุดเครื่องมือ albumentations และ cv2 บนภาษา Python

ในการขยายขนาดของข้อมูล จะทำการรวมกลยุทธที่เตรียมไว้ในทุกความเป็นไปได้ เพื่อทำการสร้างภาพและผลเฉลยใหม่ที่แตกต่างกัน โดยจำนวนของชุดข้อมูลที่ถูกสร้างขึ้นใหม่จะเป็นไปตามสมการที่ 14

$$N_{aug} = N_{dataset} 2^{N_{strategy}} \quad (14)$$

N_{aug} คือจำนวนข้อมูลที่ถูกสร้างใหม่
$N_{dataset}$ คือจำนวนข้อมูลต้นฉบับ
$N_{strategy}$ คือจำนวนกลยุทธการขยายชุดข้อมูล

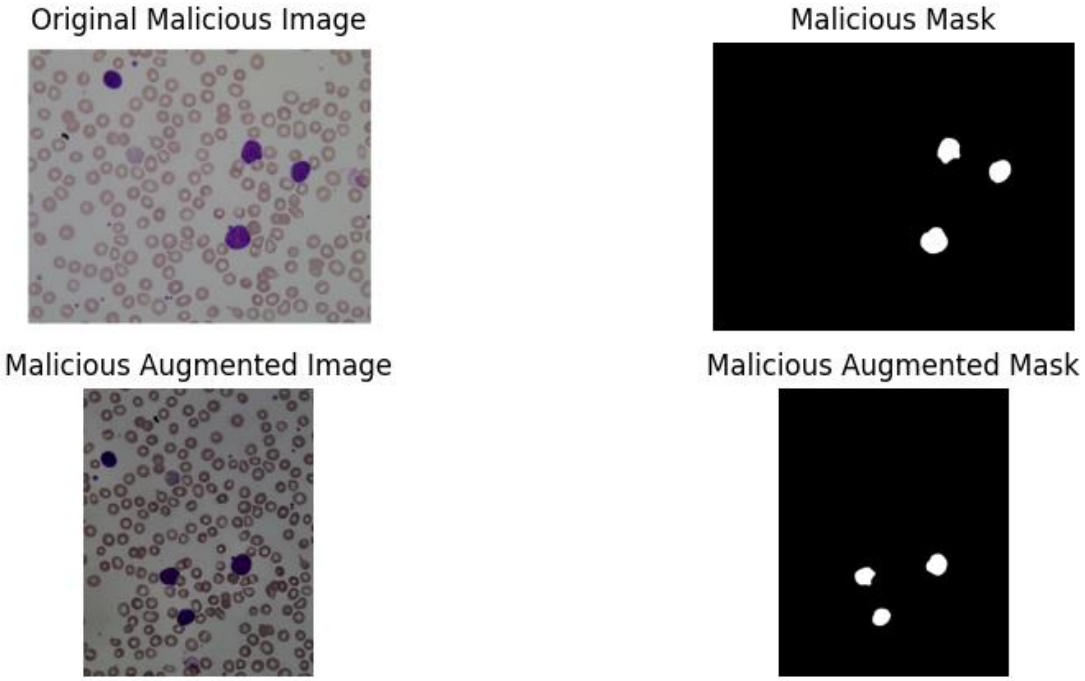
ตารางที่ 1 กลยุทธการขยายขนาดของชุดข้อมูล

กลยุทธการขยายขนาดของชุดข้อมูล	คำอธิบาย
การสุ่มผลิกรูป (Random axis flip)	สุ่มผลิกรูปในแนวนอนและแนวตั้ง
การสุ่มหมุนรูป (Random rotate)	หมุนรูปโดยการสุ่มมุมที่จะทำการหมุน
การสุ่มเลือกพื้นที่ของภาพ (Random crop)	สุ่มเลือกเฉพาะบางพื้นที่ในภาพตามขนาดที่ต้องการ
การเพิ่มความต่างของฮิสโตแกรมแบบจำกัด (CLAHE)	สุ่มเพิ่มความต่างของค่าในแต่ละพิกเซลต่างการกระจายตัวของข้อมูลในรูปภาพ
การเพิ่มตัวรบกวนแบบ Gaussian (Add gaussian noise)	เพิ่มตัวรบกวนเข้าไปในภาพ
การสุ่มค่าแกมม่า (Random gamma, p=0.8)	สุ่มปรับค่าแกมม่าเพิ่มปรับความสว่างโดยรวมของภาพ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```python
def get_img_mask_path(image_dir: str, mask_dir: str) -> dict:
    """
        args:
            image_dir (str): datast image directory
            mask_dir (str): dataset segmentation map directory
        return:
            {"images": [list of image path], "masks": [list of mask path]}
    """

    image_list = []
    mask_list = []

    for _, _, files in os.walk(image_dir):
        for name in files:
            image_path = os.path.join(image_dir,name)
            image_list.append(image_path)

            mask_path = os.path.join(mask_dir, f"{name.split('.')[0]}.png")
            mask_list.append(mask_path)

    return {"images": image_list, "masks": mask_list}

def create_aug(transform: A.transforms, data_dict: dict, save_dir: str, aug_method: str):
    save_images_dir = os.path.join(save_dir, "images")
    save_masks_dir = os.path.join(save_dir, "masks")

    if os.path.exists(save_images_dir) == False:
        os.makedirs(save_images_dir, exist_ok=True)
    if os.path.exists(save_masks_dir) == False:
        os.makedirs(save_masks_dir, exist_ok=True)

    for idx in range(len(data_dict["images"])):
        img = cv2.imread(data_dict["images"][idx])
        mask = cv2.imread(data_dict["masks"][idx])

        transformed = transform(image=img, mask=mask)

        save_img_name = f"{aug_method}_{data_dict['images'][idx].split('/')[-1]}"
        save_mask_name = f"{aug_method}_{data_dict['masks'][idx].split('/')[-1]}"

        cv2.imwrite(os.path.join(save_images_dir, save_img_name), transformed["image"])
        cv2.imwrite(os.path.join(save_masks_dir, save_mask_name), transformed["mask"])
```

รูปที่ 35 ชุดคำสั่งที่ใช้ในการขยายชุดข้อมูล



รูปที่ 36 ตัวอย่างของภาพต้นฉบับ และผลเฉลยที่ถูกสร้างขึ้นใหม่

3.4 การพัฒนาแบบจำลอง

สำหรับการพัฒนาแบบจำลองการเรียนรู้เชิงลึก ผู้วิจัยได้แบ่งออกเป็น 2 ขั้นตอนดังนี้ 1.การแบ่งชุดข้อมูลและการประมวลผลภาพเบื้องต้น และ 2.การสอนแบบจำลองการเรียนรู้เชิงลึก

3.4.1 การแบ่งชุดข้อมูลและการประมวลผลภาพเบื้องต้น

ในส่วนของการประมวลผลเบื้องต้น ผู้วิจัยได้ทำการทำให้ข้อมูลเป็นมาตรฐาน (Data normalization) โดยใช้วิธีการทำให้เป็นมาตรฐานโดยการใช้ค่าสูงสุดและต่ำสุด (Max-min normalization) เพื่อลดขอบเขตของข้อมูลในแต่ละพิกเซล จาก 0 – 255 ให้กลายเป็น 0 – 1 ซึ่งจะช่วยให้การเรียนรู้ของแบบจำลองดีขึ้นในขั้นตอนการแพร่กลับ (Backpropagation) และได้ทำการลดขนาดของรูปภาพ (Resize) เหลือ 512 x 512 พิกเซลล์ ดังตารางที่ 2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 2 ชุดคำสั่งสำหรับการประมวลผลเบื้องต้น

ชุดคำสั่ง	คำอธิบาย
การทำให้เป็นมาตรฐานโดยการใช้ค่าสูงสุด-ต่ำสุด	เพื่อลดขอบเขตของข้อมูล
การลดขนาดของรูปภาพเป็น 512 x 512 พิกเซลล์	เพื่อลดการใช้ทรัพยากรณ์ในการประมวล
การแปลงชนิดของข้อมูลให้เป็น torch.Tensor	แปลงชนิดของข้อมูลเพื่อประมวลผลด้วยชุดเครื่องมือ PyTorch

3.4.2 การสอนแบบจำลองการเรียนรู้เชิงลึก

ในปัจจุบัน มีแบบจำลองการเรียนรู้เชิงลึกสำหรับงานด้านการตรวจจับจับวัตถุและการแบ่งภาพเชิงความหมายในหลากหลายรูปแบบ ไม่ว่าจะเป็นการใช้โครงข่ายประสาทเทียมแบบสังวัตนาการเชิงพื้นที่ (Region-based Convolutional Neural Network: R-CNN) อย่าง Mask R-CNN [14] หรือจะเป็นแบบจำลองโครงข่ายประสาทแบบกราฟ (Graph Neural Network: GNN) อย่าง SCG-Net [38] ที่ได้คะแนน F1 ร้อยละ 92.0 บนชุดข้อมูล ISPRS Potsdam และยังมีจำนวนพารามิเตอร์น้อยกว่าแบบจำลองแบบ CNN อย่างมาก ซึ่งส่งผลให้แบบจำลองแบบกราฟใช้ทรัพยากรณ์ในการประมวลผลที่น้อยกว่าด้วยเช่นกัน แต่หากเทียบในเชิงของความแม่นยำแล้ว ในปัจจุบัน แบบจำลองแบบทรานฟอร์มเมอร์อย่าง ViT-Adapter-L [39] ยังลงให้ความแม่นยำที่มากที่สุดบนชุดข้อมูล Cityscapes dataset โดยอ้างอิงจากเว็บไซต์ paperswithcode [40] จึงเป็นหนึ่งในเหตุผลที่ผู้วิจัยเลือกที่จะใช้แบบจำลองแบบทรานฟอร์มเมอร์ในงานนี้

สำหรับในขั้นต้นของงานวิจัยนี้ ผู้วิจัยได้ทำการประยุกต์ใช้แบบจำลอง SegFormer [34] และแบบจำลอง DETR [41] มาพัฒนาต่อยอด

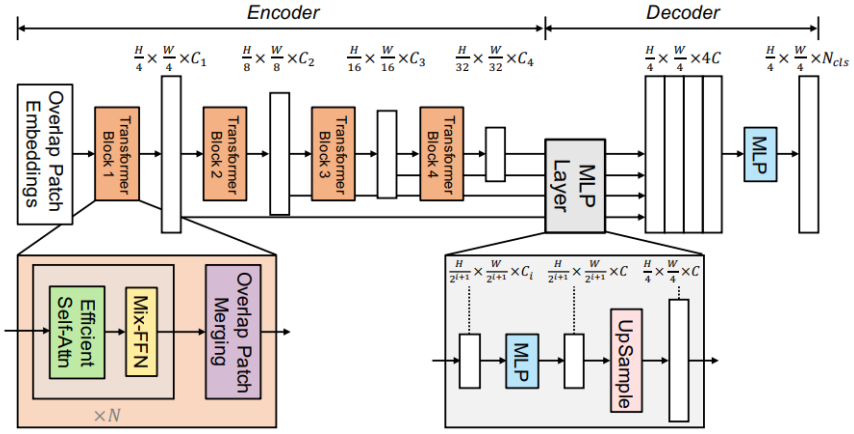
Chulalongkorn University

แบบจำลอง SegFormer เป็นแบบจำลองสำหรับการแบ่งภาพเชิงความหมายซึ่งถูกพัฒนาต่อยอดมาจาก SETR [35] โดยมีการเปลี่ยนการทำ patch embedding ที่จากเดิมที่ใช้ patch ขนาด 16x16 เปลี่ยนเป็น 4x4 ซึ่งส่งผลให้แบบจำลองมีประสิทธิภาพมากขึ้น และมีการใช้ตัวเข้ารหัสทรานฟอร์มเมอร์เชิงลำดับ (Hierarchical transformer encoder) ซึ่งเป็นตัวเข้ารหัสแบบผสม (Mix Transformer encoder: MiT) ซึ่งประกอบไปด้วย 1. ชั้นความสนใจตนเองอย่างมีประสิทธิภาพ (Efficient self-attention) ซึ่งมีความซับซ้อนในการคำนวณเพียง $O(n^2)$ น้อยกว่าเมื่อเทียบกับ self-attention ของ [26] ที่มีความซับซ้อนอยู่ที่ $O(n^2d)$

2. ชั้นโครงข่ายประสาทผสมแบบป้อมข้อมูลไปข้างหน้า (Mix-FFN) ใน SETR การเข้ารหัสตำแหน่ง (Positional encoding) จะมีข้อเสียตรงที่หากข้อมูลที่ใช้ทดสอบมีความละเอียดต่างไปจากตอนที่สอน จะทำให้ประสิทธิภาพของแบบจำลองตกลง ซึ่ง Mix-FFN จะทำ convolution 3x3 กับ FFN เพื่อลดปัญหาเหล่านี้

3. การรวมภาพย่อยแบบซ้อนทับ (Overlapped Patch Merging) ใน SETR ตัวเข้ารหัสจะสามารถสกัดคุณลักษะเด่นได้เพียงระดับเดียว (Single feature representation) แต่ใน SegFormer ได้ใช้การรวมภาพย่อยใหม่โดยให้มีส่วนที่ทับกัน ทำให้สามารถสกัดลักษณะเด่นได้เพิ่มในอีกหลายระดับ (Multi-level feature) ซึ่งคล้ายกับการเข้ารหัสของ CNN

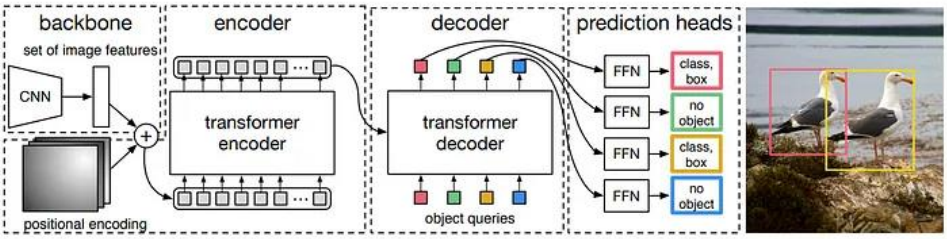
โดยที่ Hierarchical transformer encoder คือการนำ MiT มาต่อกันเรื่อยๆขึ้นอยู่กับความซับซ้อนของข้อมูล ดังที่แสดงในรูปที่ 32



รูปที่ 37 Block diagram แสดงโครงสร้างของแบบจำลอง SegFormer

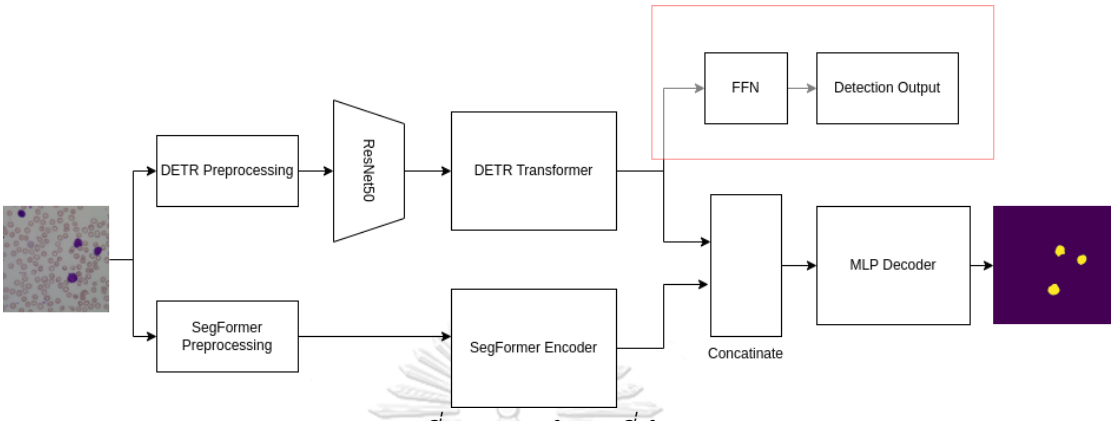
จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

แบบจำลอง DETR เป็นแบบจำลองสำหรับการตรวจจับวัตถุโดยใช้ทั้งตัวเข้ารหัสและถอดรหัสที่เป็นแบบจำลอง Transformer ผลลัพธ์ที่ได้จากตัวถอดรหัสของแบบจำลองชนิดนี้จะเป็นคุณลักษณะเด่นของข้อมูลแต่ละประเภท ดังรูปที่ 38



รูปที่ 38 Block diagram แสดงโครงสร้างของแบบจำลอง DETR

ผู้วิจัยได้นำคุณลักษณะเด่นที่สกัดได้จากตัวเข้ารหัสของ SegFormer มารวมกับคุณลักษณะเด่นจากตัวถอดรหัสของ DETR ดังรูปที่ 39



รูปที่ 39 แบบจำลองที่นำเสนอ

โดยผู้วิจัยสร้างแบบจำลอง DETR และ SegFormer จาก API ของ Hugging Face โดยใช้เส้นทาง “facebook/detr-resnet-50” สำหรับสร้าง DETR และ “nvidia/mit-b0” สำหรับสร้าง SegFormer และทำการสอนแบบจำลองโดยการแช่พารามิเตอร์ของ ResNet50

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

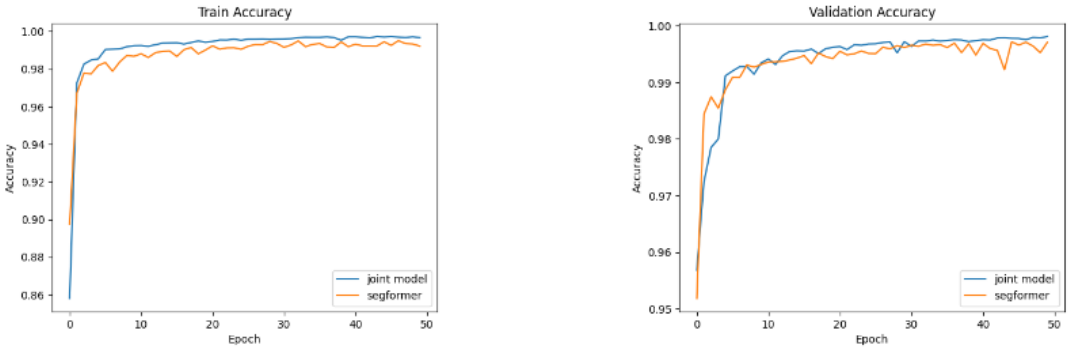
บทที่ 4 ผลการศึกษาและการอภิปรายผล

4.1 การพัฒนาแบบจำลองการแบ่งภาพเชิงความหมายโดยวิธีการเรียนรู้เชิงลึกเพื่อแบ่งภาพเซลล์ลิมโฟบลาสท์ในภาพฟิล์มเลือด

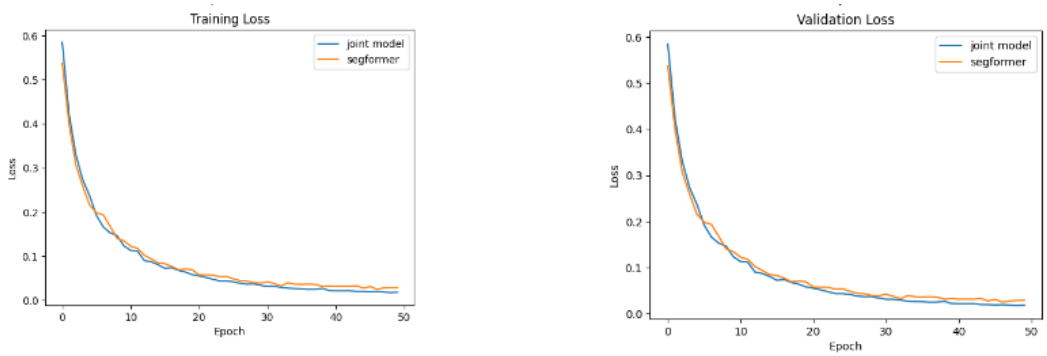
ในการสอนแบบจำลอง ผู้วิจัยได้ทำการสอนบนสภาพแวดล้อม (Environment) และกำหนดค่าพารามิเตอร์ดังตารางที่ 3 โดยได้ทำการเปรียบเทียบกับแบบจำลอง SegFormer

ตารางที่ 3 สภาพแวดล้อมและค่าพารามิเตอร์ต่างๆ

GPU	Nvidia RTX3060 VRAM 12 GB
Training Precision	FP-16-mixed precision [42]
Image_Size	512 x 512
Batch_Size	4
Epochs	50 with save best model
Optimizer	AdamW (lr=6e-5)
Loss	Cross Entropy Loss



รูปที่ 40 เปรียบเทียบความแม่นต่อชุดข้อมูลฝึก และความแม่นต่อชุดข้อมูลทดสอบระหว่างสอนของแบบจำลองที่นำเสนอ และ SegFormer



รูปที่ 41 เปรียบเทียบ loss ต่อชุดข้อมูลฝึก และ loss ต่อชุดข้อมูลทดสอบระหว่างสอนของแบบจำลองที่นำเสนอ และ SegFormer

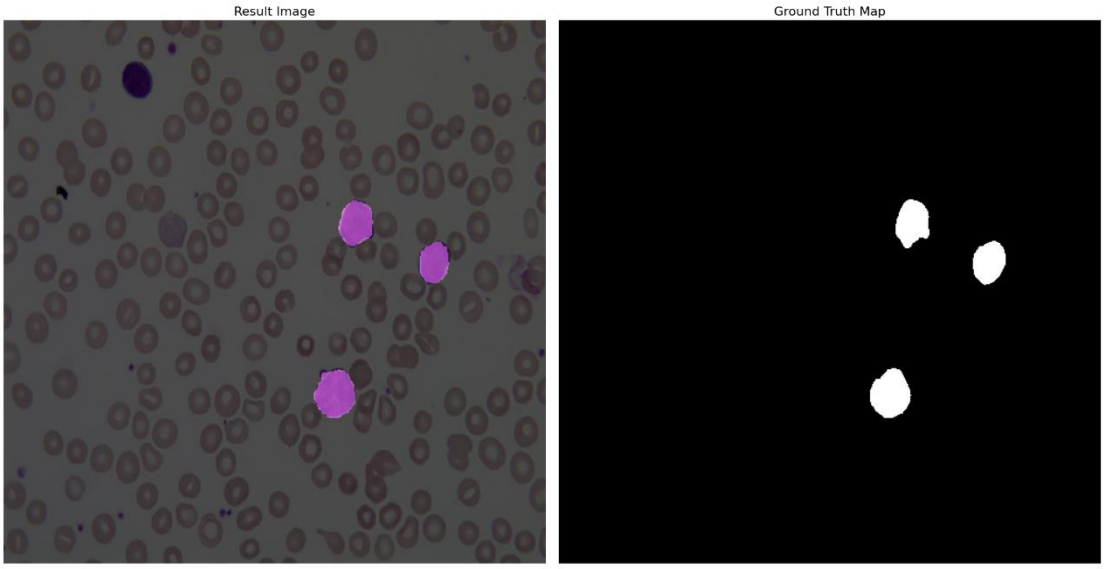
ตารางที่ 4 ผลลัพธ์โดยเฉลี่ย

	Proposed Model	SegFormer
mIoU	0.9104	0.8801
mAP50	0.8143	0.8216
mTPR (mean sensitivity)	0.8686	0.8463

ตารางที่ 5 ผลลัพธ์แบบแยกประเภท

	Proposed Model		SegFormer	
Metric	Background	Lymphoblastic Cell	Background	Lymphoblastic Cell
IoU	0.9947	0.8351	0.9934	0.8014
AP50	0.995	0.6636	0.987	0.6562
TPR	0.989	0.7482	0.993	0.6996

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 42 ตัวอย่างผลลัพธ์ของแบบจำลองการเรียนรู้เชิงลึกในการแยกเซลล์ลิมโฟบลาสต์

4.2 การพัฒนาแบบจำลองการแบ่งภาพเชิงความหมายโดยวิธีการเรียนรู้เชิงลึกเพื่อแบ่งภาพเซลล์เม็ดเลือดแดง และเซลล์เม็ดเลือดขาวในภาพฟิล์มเลือด

ในการสอนแบบจำลอง ผู้วิจัยได้ทำการสอนบนสภาพแวดล้อม (Environment) และกำหนดค่าพารามิเตอร์ดังตารางที่ 6

ตารางที่ 6 สภาพแวดล้อมและค่าพารามิเตอร์ต่างๆ

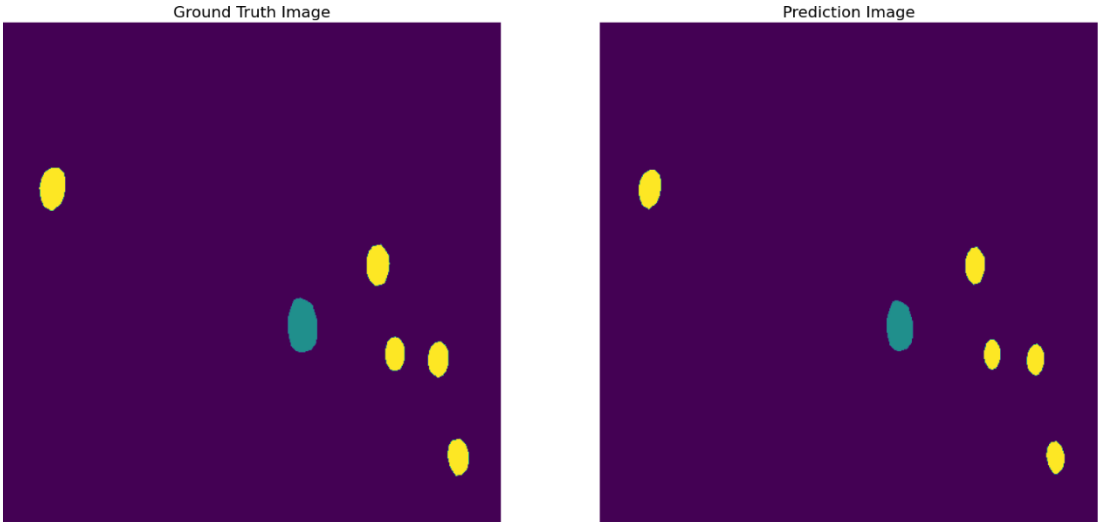
GPU	Nvidia RTX3060 VRAM 12 GB
Training Precision	FP-16-mixed precision [42]
Image_Size	512 x 512
Batch_Size	4
Epochs	100 with save best model
Optimizer	AdamW (lr=6e-5)
Loss	Cross Entropy Loss

ตารางที่ 7 ผลลัพธ์โดยเฉลี่ย

	Proposed Model
mIoU	0.9304
mAP50	0.9153
mTPR (mean sensitivity)	0.9341

ตารางที่ 8 ผลลัพธ์แบบแยกประเภท

	Proposed Model		
Metric	Background	Red Blood Cells	White Blood Cells
IoU	0.9947	0.8835	0.8516
AP50	0.9513	0.8057	0.7099
TPR	0.9842	0.8569	0.8026



รูปที่ 43 ผลการแยกเซลล์เม็ดเลือดแดงและเซลล์เม็ดเลือดขาว

4.3 การอภิปรายผลการศึกษา

การพัฒนาแบบจำลองการแบ่งภาพเชิงความหมายโดยการนำคุณลักษณะเด่นที่สกัดได้จาก SegFormer และ DETR มารวมกันแล้วสอนด้วยชุดข้อมูลที่ผ่านการขยายตามวิธีที่นำเสนอ ทำให้แบบจำลองมีประสิทธิภาพที่สูงขึ้นเมื่อเทียบกับ SegFormer ในการแบ่งภาพเซลล์ลิมโฟบลาสต์ โดยในรูปที่ 40 และ 41 เป็นการเปรียบเทียบทั้งสองแบบจำลอง จะเห็นได้ว่า แบบจำลองที่นำเสนอมีความแม่นที่สูงกว่าทั้งในชุดข้อมูลสอน และชุดข้อมูลทดสอบระหว่างสอน อีกทั้งยังมี loss ที่คำนวณจาก cross entropy loss ที่น้อยกว่าด้วยเช่นกัน และในชุดข้อมูลทดสอบ แบบจำลองที่นำเสนอได้ mIoU ที่ 0.9104, mAP ที่ 0.8143, และ mean sensitivity ที่ 0.9341 ซึ่งสูงกว่า SegFormer ที่ได้ mIoU ที่ 0.8801, mAP ที่ 0.8216, และ mean sensitivity ที่ 0.8463 โดยทั้งสอบแบบจำลองได้ทำการปรับค่าพารามิเตอร์บนสภาพแวดล้อมเดียวกันดังตารางที่ 3

ส่วนของแบบจำลองแบบภาพเซลล์เม็ดเลือดแดง และเซลล์เม็ดเลือดขาว แบบจำลองที่นำเสนอได้ mIoU ที่ 0.9304, mAP ที่ 0.9153, และ mean sensitivity ที่ 0.9341

4.4 แนวทางการพัฒนา

ในการเพิ่มประสิทธิภาพของแบบจำลองที่นำเสนอ อาจทำได้โดยการสอนโดยใช้ชุดข้อมูลจริงที่ใหญ่มากขึ้น เพื่อให้แบบจำลองได้เรียนรู้รูปแบบของเซลล์ที่หลากหลายมากขึ้น ซึ่งทำได้โดยการสอนโดยใช้ชุดข้อมูลที่คล้ายกัน แล้วจึงนำไปปรับพารามิเตอร์ใหม่ด้วยชุดข้อมูลที่ต้องการ

การทำการสังเคราะห์ภาพใหม่จากการใช้แบบจำลองการเรียนรู้เชิงลึกอย่าง Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) หรือ Diffusion model เพื่อขยายชุดข้อมูล จะทำให้ได้ชุดข้อมูลใหม่ที่มีการกระจายตัวของข้อมูลที่ใกล้เคียงทำต้นฉบับ ซึ่งอาจจะทำให้ได้ชุดข้อมูลที่หลากหลายขึ้น

การแยกสอนแบบจำลอง DETR ก่อนที่จะนำมารวมกับ SegFormer อาจจะทำให้แบบจำลอง DETR มีประสิทธิภาพในการสกัดคุณลักษณะเด่นที่ดียิ่งขึ้น เนื่องมาจากพารามิเตอร์ของแบบจำลองที่ต้องการปรับมีจำนวนลดลง แต่ข้อมูลที่ใช้สอนมีเท่าเดิม

พัฒนาแบบจำลองสำหรับการทำการแบ่งภาพตัวอย่าง (Instance segmentation) จะให้ผลลัพธ์ที่สามารถนำมาวัดผลและแปลผลด้วย mAP และ sensitivity ที่ดีกว่า เนื่องจากการแบ่งภาพเชิงความหมายเป็นการแบ่งภาพโดยรวมของแต่ละประเภท และการแบ่งภาพตัวอย่างจะเป็นการแบ่งที่ตัววัตถุที่สนใจ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

[1] R. B. Hegde, K. Prasad, H. Hebbar, and I. Sandhya, "Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review," *Biocybernetics and Biomedical Engineering,* vol. 38, no. 3, pp. 467-480, 2018.

[2] A. Tareef, Y. Song, D. Feng, M. Chen, and W. Cai, "Automated multi-stage segmentation of white blood cells via optimizing color processing," in *2017 IEEE 14th international symposium on Biomedical imaging (ISBI 2017)*, 2017: IEEE, pp. 565-568.

[3] G. K. Chadha, A. Srivastava, A. Singh, R. Gupta, and D. Singla, "An automated method for counting red blood cells using image processing," *Procedia Computer Science,* vol. 167, pp. 769-778, 2020.

[4] A. M. P. G. Vale, A. M. G. Guerreiro, A. D. Dória Neto, G. B. Cavalvanti Junior, V. C. L. T. d. S. Leitão, and A. M. Martins, "Automatic segmentation and classification of blood components in microscopic images using a fuzzy approach," *Revista Brasileira de Engenharia Biomédica,* vol. 30, pp. 341-354, 2014.

[5] M. Imran Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware CNN and extreme machine learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 49-55.

[6] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: deep contour-aware networks for accurate gland segmentation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 2487-2496.

[7] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in *2011 18th IEEE international conference on image processing*, 2011: IEEE, pp. 2045-2048.

[8] T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, and K.-W. Kang, "Blood cell images segmentation using deep learning semantic segmentation," in *2018 IEEE international conference on electronics and communication engineering (ICECE)*, 2018: IEEE, pp. 13-16.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 12, pp. 2481-2495, 2017.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[11] S. Fujita and X.-H. Han, "Cell detection and segmentation in microscopy images with improved mask R-CNN," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[12] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "An automated blood cells counting and classification framework using mask R-CNN deep learning model," in *2019 31st international conference on microelectronics (ICM)*, 2019: IEEE, pp. 300-303.

[13] D. R. Loh, W. X. Yong, J. Yapeter, K. Subburaj, and R. Chandramohanadas, "A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN," *Computerized Medical Imaging and Graphics,* vol. 88, p. 101845, 2021.

[14] M. P. Paing, A. Sento, T. H. Bui, and C. Pintavirooj, "Instance Segmentation of Multiple Myeloma Cells Using Deep-Wise Data Augmentation and Mask R-CNN," *Entropy,* vol. 24, no. 1, p. 134, 2022.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.

[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[18] B. A. Hamilton. "2018 Data Science Bowl." https://www.kaggle.com/c/data-science-bowl-2018/ (accessed.

[19] N. C. Institute. "CANCER IMAGING ARCHIVE." https://www.cancerimagingarchive.net/ (accessed.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[20] A. Gupta *et al.*, "SegPC-2021: A challenge & dataset on segmentation of Multiple Myeloma plasma cells from microscopic images," *Medical Image Analysis,* p. 102677, 2022.

[21] L. Zhang, T. Wen, and J. Shi, "Deep image blending," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 231-240.

[22] J. Peng, Z. Nan, L. Xu, J. Xin, and N. Zheng, "A deep model for joint object detection and semantic segmentation in traffic scenes," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020: IEEE, pp. 1-8.

[23] Z. Nan *et al.*, "A joint object detection and semantic segmentation model with cross-attention and inner-attention mechanisms," *Neurocomputing,* vol. 463, pp. 212-225, 2021.

[24] L. V. G. M. Everingham, C.K.I. Williams, J. Winn, and A. Zisserman. "The {PASCAL} {V}isual {O}bject {C}lasses {C}hallenge 2012 {(VOC2012)} {R}esults." http://host.robots.ox.ac.uk/pascal/VOC/voc2012/ (accessed.

[25] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science,* vol. 2, no. 3, pp. 1-21, 2021.

[26] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.

[27] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020.

[28] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?," *Advances in Neural Information Processing Systems,* vol. 34, pp. 12116-12128, 2021.

[29] K. Patel, A. M. Bur, F. Li, and G. Wang, "Aggregating global features into local vision transformer," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022: IEEE, pp. 1141-1147.

[30] Y. Deng *et al.*, "StyTr2: Image Style Transfer with Transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11326-11336.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

[31] W. Li, S. Wen, K. Shi, Y. Yang, and T. Huang, "Neural architecture search with a lightweight transformer for text-to-image synthesis," *IEEE Transactions on Network Science and Engineering,* vol. 9, no. 3, pp. 1567-1576, 2022.

[32] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, and C. Schmid, "Vivit: A video vision transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6836-6846.

[33] J. Beal, E. Kim, E. Tzeng, D. H. Park, A. Zhai, and D. Kislyuk, "Toward transformer-based object detection," *arXiv preprint arXiv:2012.09958,* 2020.

[34] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems,* vol. 34, pp. 12077-12090, 2021.

[35] S. Zheng *et al.*, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 6881-6890.

[36] OpenCV. "Computer Vision Annotation Tool (CVAT)." https://github.com/opencv/cvat (accessed.

[37] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014: Springer, pp. 740-755.

[38] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, "Scg-net: self-constructing graph neural networks for semantic segmentation," *arXiv preprint arXiv:2009.01599,* 2020.

[39] Z. Chen *et al.*, "Vision Transformer Adapter for Dense Predictions," *arXiv preprint arXiv:2205.08534,* 2022.

[40] paperswithcode. "Semantic Segmentation on Cityscapes test." https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes (accessed.

[41] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*, 2020: Springer, pp. 213-229.

[42] P. Micikevicius *et al.*, "Mixed precision training," *arXiv preprint arXiv:1710.03740,*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2017.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล ภูมิพัฒน์ เจริญธนานุวัฒน์

วุฒิการศึกษา Bachelor of Electrical Engineering, Kasetsart University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University