การทำนายการจดจำภาพด้วยการเรียนรู้เชิงลึก

IMAGE MEMORABILITY PREDICTION USING DEEP LEARNING

รัฐพร คุณสมบัติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

การทำนายการจดจำภาพด้วยการเรียนรู้เชิงลึก

รัฐพร คุณสมบัติ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566

RATTAPORN KUNSOMBAT

A Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2023

สารนิพนธ์

เรื่อง

การทำนายการจดจำภาพด้วยการเรียนรู้เชิงลึก

ของ

รัฐพร คุณสมบัติ

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.นภา แซ่เบ๊)

..... ประธาน

(อาจารย์ ดร.นิดา ชาติวัฒนศิริ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	การทำนายการจดจำภาพด้วยการเรียนรู้เชิงลึก
ผู้วิจัย	รัฐพร คุณสมบัติ
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นภา แซ่เบ๊

ความสามารถในการจดจำภาพสามารถวัดได้จากพฤติกรรมและประสบการณ์ของแต่ละบุคคล โดยมุมมองทางจิตวิทยาความจำมาจากการกระตุ้นภายในสมองและการใช้ชีวิตประจำวัน ในงานวิจัยนี้มุ่งศึกษาการสร้างแบบจำลองเพื่อทำนายการจดจำภาพโดยใช้เทคนิคการเรียนรู้เชิงลึก (Deep Learning) โดยรูปแบบของแบบจำลองที่นำมาศึกษาประกอบด้วย 1) สถาปัตยกรรมแบบ ResNet50 ซึ่งเป็นโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network, CNN) 2) สถาปัตยกรรมแบบ ViT ซึ่งเป็นโครงข่ายประสาทเทียมแบบทรานฟอร์เมอร์ (Transformer) และ 3) การแบบจำลองผสมผสานที่ได้จาการนำทั้งสองโมเดลมาเชื่อมต่อกันแบบคู่ขนานโดยในการฝึกแบบจำลองแบ่งเป็น 3 แบบ ได้แก่ 1) การฝึกแบบจำลองจากแรกเริ่ม (Trained from scratch) 2) การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ(Pretrained model) และ 3) การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น,มาปรับแต่งเพิ่มเติม (Fine-tuning) เพื่อเปรียบเทียบประสิทธิภาพโมเดลในการจดจำภาพ โดยจากผลการทดลอง การนำแบบจำลอง ResNet50 ที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะมาใช้ฝึกในชุดข้อมูลที่มีการคละหมวดหมู่ ให้ผลการทดลองที่ดีที่สุดโดยมีค่าประสิทธิภาพการทำนายคะแนนการจดจำภาพดังนี้ คือ Mean Squared Error (MSE) 0.0001 Mean Absolute Error (MAE) 0.0082 R-square (R^2) 0.9947 และ Spearman Correlation Coefficient (Spearman's rho) 0.9896

คำสำคัญ : คะแนนการจดจำภาพ, โครงข่ายสังวัฒนาการ, โครงข่ายทรานสฟอรมเมอร์

Title	IMAGE MEMORABILITY PREDICTION USING DEEP LEARNING
Author	RATTAPORN KUNSOMBAT
Degree	MASTER OF SCIENCE
Academic Year	2023
Thesis Advisor	Assistant Professor Dr. Napa Sae-Bae

The ability to memorize images can be assessed based on the behavior and experiences of individuals, from a psychological perspective on memory, stemming from internal brain stimulation and daily life usage. This research focuses on predicting image memorization using deep learning techniques. In particular, this research employs three types of model architecture: (1) ResNet50 (a 50-layer convolutional neural network) which utilized; (2) ViT (Vision Transformer model); and (3) a hybrid model using ResNet50 and ViT in conjunction to predict memorability scores. These models were trained using three distinct approaches: (1) training from scratch; (2) utilizing the pretrained models; and (3) fine-tuning the pretrained model, in order to compare the performance of the models in image memorization. The result revealed that the pretrained ResNet50 model (without fine-tuning) yielded the best performance compared to other models, with 0.0001 Mean Squared Error (MSE), 0.0082 Mean Absolute Error (MAE), 0.9947 R-square (R^2) and a 0.9896 Spearman Correlation Coefficient (Spearman's rho).

Keyword : Image memorability, Convolutional Neural Network, Transformer

กิตติกรรมประกาศ

การจัดทำวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีจากการสนับสนุน ความรู้ ความช่วยเหลือ คำแนะนำ ตลอดจนแนวทางในการทำวิจัยและจัดทำสารนิพนธ์ของ ผศ. ดร. นภา แซ่เบ๊ อาจารย์ที่ปรึกษา และคณาจารย์ทุกท่านในหลักสูตรวิทยาการข้อมูล ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรต การสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

รัฐพร คุณสมบัติ

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

การจดจำภาพเป็นสิ่งสำคัญในชีวิตประจำวันของมนุษย์ทุกคน ซึ่งประสบการณ์ที่เป็นเอกลักษณ์และสร้างความทรงจำเป็นของตนเองมักเกิดขึ้นเมื่อพบเจอสิ่งต่าง ๆ บางสิ่งอาจจดจำได้ทันที ในขณะที่บางสิ่งก็อาจลืมได้ในระยะเวลาอันสั้น ความสามารถในการจดจำภาพสามารถวัดได้จากพฤติกรรมและประสบการณ์ของแต่ละบุคคล โดยมุมมองทางจิตวิทยาความจำมาจากการกระตุ้นภายในสมองวิทยาและการใช้ชีวิตประจำวัน ในความเป็นจริง มนุษย์ไม่สามารถที่จะคาดการณ์ได้ว่าจะสามารถจดจำสิ่งต่าง ๆ ได้ดีหรือไม่ งานวิจัยอื่น (Thomas & Thomas, 2023) พบว่าในแต่ละบุคคลมักจะมีความสอดคล้องกันในภาพที่สามารถจดจำได้ นั่นคือ ผู้คนมักจะจดจำและลืมภาพแบบเดียวกัน แม้ว่าจะมีประสบการณ์ที่แตกต่างกัน เช่น ใบหน้าและฉากพื้นหลัง เราสามารถสร้างแนวคิดเกี่ยวกับการจดจำโดยวัดผลจากความน่าจะเป็นที่แต่ละบุคคลจะจำภาพนั้นได้หลังจากที่พบเห็นหรือไม่ ดังนั้น แม้ว่ามนุษย์จะไม่สามารถคาดการณ์การจดจำของภาพได้ แต่ในงานวิจัยนั้น (Isola et al., 2014) นักวิจัยสามารถให้คำตอบว่ามนุษย์จะจดจำภาพใดได้บ้างโดยพิจารณาจากความสามารถในการจดจำที่วัดได้เพียงอย่างเดียว ซึ่งความทรงจำในความเป็นจริงนั้น การรู้ว่าภาพใดที่น่าจะถูกจดจำโดยไม่คำนึงถึงลักษณะของผู้สังเกตการณ์สามารถเป็นประโยชน์และต่อยอดในการทำงานอื่น ๆ ได้ ในทางหนึ่ง เราสามารถใช้ความรู้นี้เพื่อจัดการกับความทรงจำ มีการใช้งานด้านการจดจำภาพนี้ในงานสาขาต่าง ๆ มากมาย เช่น การศึกษา การโฆษณาและสื่อการแพทย์ เป็นต้น

จากเหตุผลที่กล่าวมาข้างต้น ผู้วิจัยเล็งถึงความสำคัญของการจดจำภาพในชีวิตประจำวันของมนุษย์ การจดจำนี้สามารถมีผลประโยชน์ได้ในหลายมิติต่าง ๆ ดังนั้น ผู้วิจัยได้ทำการสร้างแบบจำลองทำนายการจดจำภาพโดยใช้เทคนิคการเรียนรู้เชิงลึกสามรูปแบบ โดยรูปแบบที่หนึ่งอาศัยโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network, CNN) ที่เรียกว่า ResNet50 รูปแบบที่สองอาศัยโครงสร้างแบบจำลองในบริบททางภาษา (Transformer) ที่เรียกว่า Vision Transformer มาใช้ในการประมวลผลเพื่อคำนวณเวกเตอร์คุณลักษณะที่นำไปใช้ในการทำนายการจดจำและรูปแบบที่สามอาศัยการนำเวกเตอร์คุณลักษณะทั้งสองแบบจำลองมาเชื่อมต่อกันโดยในการฝึกแบบจำลองแบ่งเป็น 3 แบบ ได้แก่ 1) การฝึกสร้างแบบจำลองจากแรกเริ่มศูนย์หรือต้นแบบ (Trained From Scratch) 2) การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะของการเรียนรู้เชิงลึกก่อนหน้า

(Pretrained Model) และ 3) นำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติมกับแบบจำลองที่ได้ฝึกไว้ก่อนหน้า (Fine-Tuning) เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองในการจดจำภาพ และทำการเปรียบเทียบผลการทำนายของแบบจำลองทั้งสามรูปแบบ โดยการนำเวกเตอร์คุณลักษณะจากแบบจำลองทั้งสามประเภทมาการทำนายการจดจำภาพแบบถดถอย (Regression)

1.2 วัตถุประสงค์

เพื่อทำการพัฒนาต้นแบบของแบบจำลองเพื่อใช้ในการทำนายคะแนนการจดจำภาพโดยมีรายละเอียดดังนี้

1. อาศัยโครงสร้างแบบจำลอง ResNet50 และ Vision Transformer โดยฝึกแบบจำลองดังกล่าวใน 3 รูปแบบ ดังนี้

1.1 การฝึกสร้างแบบจำลองจากแรกเริ่มจากศูนย์ หรือต้นแบบ

1.2 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะของการเรียนรู้เชิงลึกก่อนหน้า

1.3 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาการปรับแต่งเพิ่มเติมแบบจำลองที่ได้ฝึกไว้ก่อนหน้า

2. ทดสอบประสิทธิภาพของแบบจำลองที่ฝึกในรูปแบบต่าง ๆ จากชุดข้อมูลแบบเฉพาะหมวดหมู่ (Individual Category) และชุดข้อมูลแบบคละหมวดหมู่ (Merged Category)

1.3 ขอบเขตงานวิจัย

1.3.1 ชุดข้อมูลที่ใช้ในงานวิจัย

ชุดข้อมูลชื่อ Memcat (Goetschalckx & Wagemans, 2019)เป็นชุดข้อมูลสาธารณะแหล่งที่มาจาก ImageNet, COCO, Open Images และ SUN จากงานวิจัยเกี่ยวกับชุดข้อมูลนี้พบว่ามีปัญหาในเรื่องของความหลากหลายของรูปภาพในชุดข้อมูล เพื่อแก้ไขปัญหานี้ ชุดข้อมูลดังกล่าวได้ถูกจัดระเบียบโดยแยกแยะรูปภาพออกเป็นหมวดหมู่ ซึ่งแบ่งเป็น 5 หมวดหมู่ ได้แก่ Animal, Food, Sports, Landscape และ Vehicle โดยมีจำนวนรูปภาพในแต่ละหมวดหมู่เป็น 2,000 รูป รวมทั้งหมด 10,000 รูปภาพ ซึ่งในชุดข้อมูลนี้ประกอบด้วยข้อมูล 2 ส่วนหลักๆ ได้แก่ ไฟล์รูปภาพ (jpeg) และไฟล์ข้อมูลที่เกี่ยวข้องกับรูปภาพ (csv) ดังนี้

1. รูปภาพ
2. ข้อมูลรูปภาพที่เกี่ยวข้อง ประกอบด้วย
3. ชื่อรูปภาพ
4. หมวดหมู่
5. หมวดหมู่ย่อย
6. ความกว้างและสูงของรูป
7. แหล่งที่มาของรูปภาพ
8. ป้ายกำกับภาพ จากแหล่งที่มา
9. จำนวนการเข้าดูรูป
10. การแจ้งเตือนที่ผิดพลาดจากการจดจำรูปภาพ
11. จำนวนผู้เข้าร่วมการจดจำภาพที่ผ่านเกณฑ์
12. คะแนนความจำภาพที่ไม่มีการแก้ไขจากการแจ้งเตือนความผิดพลาด
13. คะแนนความจำภาพที่มีการแก้ไขจากการแจ้งเตือนความผิดพลาด

1.3.2 ตัวแปรที่ศึกษา

ในการวิจัยนี้ตัวแปรที่ศึกษามีสองประเภท ได้แก่ ตัวแปรอิสระและตัวแปรตาม ดังนี้

- ตัวแปรอิสระ
    1.1 รูปภาพรวมทั้งหมด 10,000 รูป
    1.2 ข้อมูลหมวดหมู่ของรูปภาพแต่ละภาพ ประกอบด้วย
        1. Animal
        2. Food
        3. Sports
        4. Landscape
        5. Vehicle
- ตัวแปรตาม เป็นคะแนนการจดจำภาพแต่ละภาพ (โดยไม่มีการแก้ไขจากการแจ้งเตือนความผิดพลาด)

1.4 ขั้นตอนการดำเนินงานวิจัย

1. ศึกษาข้อมูลเกี่ยวกับการจดจำภาพ

2. ศึกษางานวิจัยที่เกี่ยวกับการทำนายการจดจำภาพ

3. ศึกษาข้อมูลงานวิจัยเกี่ยวกับ ประเภทโครงข่ายประสาทเทียมแบบสังวัตนาการ แล Transformer สำหรับใช้ในการทำนายการจดจำภาพ พร้อมทั้งเปรียบเทียบประเมินผลจากงานวิจัยของแบบจำลองแต่ละประเภท

4. วิเคราะห์และเตรียมชุดข้อมูลที่เหมาะสมสำหรับการทดลอง

5. สร้างและฝึกแบบจำลองด้วยชุดข้อมูลที่เลือกโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ ที่เรียกว่า ResNet50, แบบจำลองในบริบททางภาษา (Transformer) ที่เรียกว่า Vision Transformer และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน

6. ประเมินประสิทธิภาพของแบบจำลองทั้งหมดโดยใช้ชุดข้อมูลทดสอบที่เฉพาะเจาะจง เปรียบเทียบผลลัพธ์ระหว่างแบบจำลองวิเคราะห์ผลลัพธ์เกี่ยวกับความสามารถของแต่ละแบบจำลอง

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. แบบจำลองที่ได้สามารถนำไปใช้ในการศึกษาการจดจำภาพในมนุษย์

2. แบบจำลองที่ได้สามารถนำไปใช้ในการทำนายการจดจำแบรนด์ของผู้บริโภค เพื่อให้บริษัทสามารถสร้างกลยุทธ์การตลาดและโฆษณาที่เหมาะสม ช่วยเพิ่มประสิทธิภาพด้านการตลาดสินค้าและบริการ

3. แบบจำลองที่ได้สามารถนำไปพัฒนาเทคโนโลยีและการแพทย์ที่มีประโยชน์ เช่น การช่วยในการวินิจฉัยโรคที่เกี่ยวข้องกับสมอง การตรวจสอบภาวะสุขภาพจิต หรือนำไปใช้วิเคราะห์เพื่อการพัฒนาเทคโนโลยีที่ช่วยในการฟื้นฟูสมองหลังจากเกิดอุบัติเหตุที่เกี่ยวข้องกับสมอง

1.6 สมมติฐานในการวิจัย

การจดจำภาพได้ในแต่ละบุคคลมีความสอดคล้องกันจากประสบการณ์การพบเห็นวัตถุต่าง ๆ ร่วมกัน โดยแบบจำลองการเรียนรู้เชิงลึกที่ใช้ได้ดีในการทำนายคุณสมบัติต่าง ๆ ของภาพสามารถนำมาทำนายการจดจำภาพได้ และใช้การถ่ายโอนความรู้ของแบบจำลองในการทำนายประเภทวัตถุ เพื่อเพิ่มประสิทธิภาพการทำนายได้ดียิ่งขึ้น

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

2.1 การเรียนรู้เชิงลึก (Deep Learning)

2.1.1 โครงสร้างของโครงข่ายการเรียนรู้เชิงลึก

2.1.2 โครงสร้างของโครงข่ายการเรียนรู้เชิงลึกสำหรับการทำนายแบบถดถอย (Regression)

2.2 สถาปัตยกรรมโครงข่ายการเรียนรู้เชิงลึกที่ใช้ในการคำนวณเวกเตอร์คุณลักษณะ

2.2.1 โครงข่ายแบบสังวัตนาการ (Convolutional Neural Networks, CNN)

2.2.2 โครงข่ายแบบ Transformer

2.3 การฝึกแบบจำลองโครงข่ายการเรียนรู้เชิงลึก

2.3.1 การฝึกแบบจำลองจากแรกเริ่ม (Trained from Scratch)

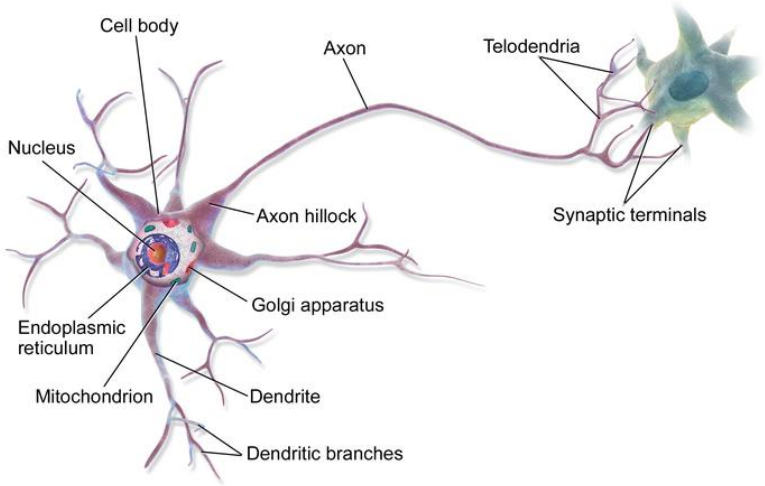
2.3.2 การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ (Pretrained Model)

2.3.3 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม (Fine-Tuning)

2.4 งานวิจัยที่เกี่ยวข้องกับการทำนายการจดจำภาพ

2.1 การเรียนรู้เชิงลึก

การเรียนรู้อัตโนมัติจำลองการทำงานของเซลล์ประสาทของมนุษย์ ซึ่งการเรียนรู้เชิงลึกเป็นการใช้ชั้นของเครือข่ายประสาทหลายชั้นที่รวมกัน ยิ่งมีจำนวนชั้นมากเครือข่ายจะมีโครงสร้างที่ซับซ้อนมากขึ้น ดังนั้น โครงข่ายการเรียนรู้เชิงลึก (LeCun Y, 2015) เป็นการเพิ่มจำนวนชั้นซ่อนที่ในเครือข่ายประสาท ซึ่งจะทำให้มีประสิทธิภาพในการทำงานเพิ่มขึ้นได้



ภาพประกอบ 1 แสดงโครงสร้างเซลล์ประสาทมนุษย์

ที่มา: (เซลล์ประสาท, 2023)

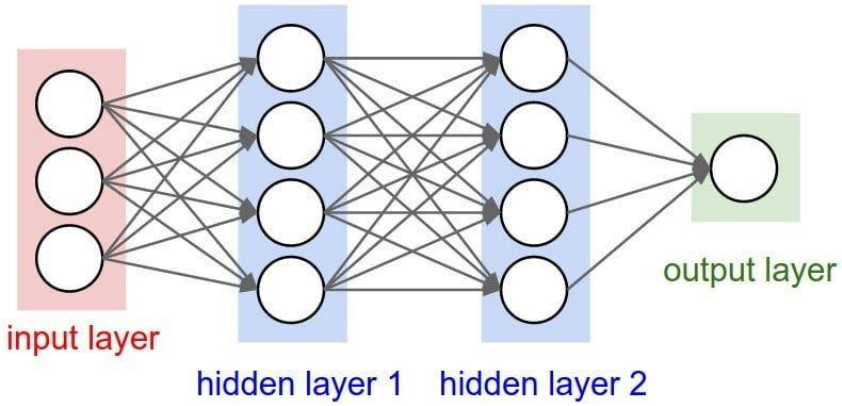
2.1.1 โครงสร้างของโครงข่ายการเรียนรู้เชิงลึก

1. โครงข่ายการเรียนรู้เชิงลึกประกอบด้วย 4 รูปแบบ ซึ่งแต่ละชั้นมีหน้าที่แตกต่างออกไปตามลักษณะของการประมวลผลข้อมูล โดยทั่วไปมีลักษณะดังภาพประกอบ 2 และมีรายละเอียดของชั้น (Layer) ในโครงข่ายการเรียนรู้เชิงลึก ประกอบด้วย

1.1. ชั้นข้อมูลเข้า (Input Layer) ทำหน้าที่รับข้อมูลนำเข้า (Input Data) เข้าสู่โครงข่ายประสาทเทียมโดยจะมีจำนวนโหนด (Nodes) ที่เท่ากับจำนวนคุณสมบัติหรือลักษณะของข้อมูลนำเข้า

1.2. ชั้นที่ซ่อน (Hidden Layer) ทำหน้าที่รับข้อมูลจากชั้นก่อนหน้าและส่งต่อไปยังชั้นถัดไปประกอบด้วยหลายๆ ชั้นซึ่งใช้ในการประมวลผลข้อมูลภายในโครงข่ายชั้นที่ซ่อนช่วยให้แบบจำลองเรียนรู้ลักษณะและความซับซ้อนของข้อมูล

1.3. ชั้นชั้นผลลัพธ์ (Output Layer) ทำหน้าที่แสดงผลลัพธ์หลังจากการประมวลผลเสร็จสิ้นจำนวนโหนดในชั้นนี้ขึ้นอยู่กับประเภทของงานที่ต้องการ เช่น การจำแนกประเภทแบบทวิภาค (Binary Classification) มีจำนวน 2 โหนด, การจำแนกประเภทแบบหลายคลาส (Multi-Class Classification) จำนวนโหนดขึ้นอยู่กับจำนวนคลาสที่แบ่ง หรือการทำนายแบบถดถอย มีจำนวน 1 โหนด เป็นต้น



ภาพประกอบ 2 แสดงโครงสร้างโครงข่ายการเรียนรู้เชิงลึก

ที่มา: (Deep Learning & Neural Networks, 2019)

2. ประเภทของชั้นในโครงข่ายการเรียนรู้เชิงลึกชั้นพิเศษอื่น ๆ ในงานการเรียนรู้เชิงลึก

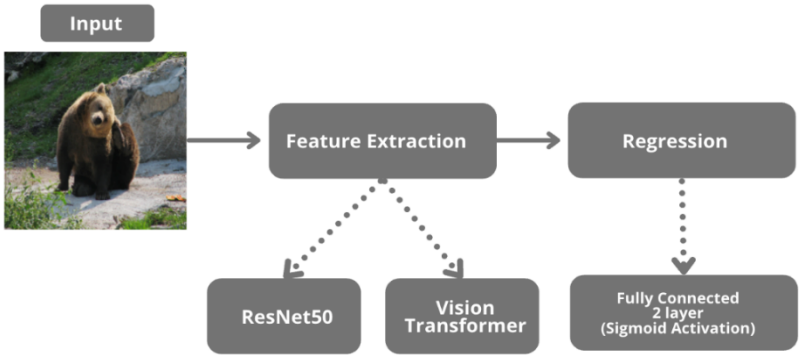
2.1 Dropout Layer ชั้นที่ช่วยลดโอกาสการเกิดกับข้อมูลที่ไม่เคยเห็นมาก่อนได้ (Overfitting) ในแบบจำลองโดยการสุ่มปิดการทำงานของบางโหนดในชั้นก่อนหน้าในขณะการฝึกแบบจำลอง

2.2 Batch Normalization Layer ชั้นที่ช่วยลดปัญหา Vangishing Gradient และ Exploding Gradient ช่วยให้การฝึกแบบจำลองเร็วขึ้นโดยการปรับปรุงและมีสัดส่วนของค่า Gradient ที่เหมาะสมในแต่ละชั้น

2.3 Pooling Layer ชั้นที่ใช้ในการลดขนาดของข้อมูล และยังช่วยลดการคำนวณที่ซับซ้อนในแบบจำลอง มีสองประเภทหลักคือ Max Pooling และ Average Pooling

2.1.2 โครงสร้างของโครงข่ายการเรียนรู้เชิงลึกสำหรับการทำนายแบบถดถอย

1. การทำนายแบบถดถอยในโครงข่ายการเรียนรู้เชิงลึกเป็นกระบวนการที่แบบจำลองนำเสนอผลเป็นค่าต่อเนื่อง (Continuous Value) โดยที่เป้าหมายคือการทำนายค่าต่อเนื่องของตัวแปรต้นทาง (Independent Variables) จากชุดข้อมูลที่ใช้ในการฝึก (Training Data) โดยใช้โครงข่ายการเรียนรู้เชิงลึกในการเรียนรู้ความสัมพันธ์ระหว่างตัวแปรต้นทางกับตัวแปรตามที่เป้าหมาย โดยมักใช้สำหรับงานที่ต้องการทำนายค่าต่อเนื่อง ซึ่งการเขียนโครงสร้างของโครงข่ายการเรียนรู้เชิงลึกสำหรับปัญหาแบบถดถอยในงานวิจัยมีลักษณะโครงสร้าง ดังภาพประกอบ 3



ภาพประกอบ 3 โครงสร้างของการเรียนรู้เชิงลึกสำหรับการทำนายแบบถดถอยในงานวิจัย

2. การสร้างโครงข่ายการเรียนรู้เชิงลึกสำหรับการทำนายแบบถดถอย ในการเรียนรู้เชิงลึก สามารถทำได้โดยใช้โครงข่ายประสาทเทียม (Neural Networks) ที่มีชั้นซ่อน (Hidden Layers) มากพอสมควรให้แบบจำลองสามารถเรียนรู้และคาดการณ์ความสัมพันธ์ที่ซับซ้อนระหว่างข้อมูลนำเข้าและผลลัพธ์ที่ต้องการ โดยในชั้น Output สุดท้ายของแบบจำลองจะมีเซลล์เดียวที่มีการแปลงเป็นค่าตัวเลขหรือค่าต่อเนื่องที่ต้องการทำนาย เช่น ใช้ฟังก์ชันเชิงเส้นหรือเชิงเน้น Activation Function เพื่อให้ได้ผลลัพธ์ที่เหมาะสมตามความต้องการของงานที่กำหนดชั้นสำหรับการทำนายแบบถดถอยในการเรียนรู้เชิงลึก ประกอบด้วย

- Fully Connected Layer (Dense Layer) ใช้ในการประมวลผลเชิงเส้นของคุณลักษณะหรือเวกเตอร์ที่ถูกสกัดมา เพื่อสร้างผลลัพธ์ที่เป็นตัวเลขหรือค่าที่ต้องการทำนาย

- Output Layer ชั้นสุดท้ายที่ใช้ในการสร้าง Output ของ Regression Head โดยมักเป็นชั้น Linear Layer ที่มีเซ็ตของช่องเป้าหมายในการทำนายค่าตัวเลขที่ต้องการ

- Activation Function ในบางกรณีอาจมีการใช้ Activation Function เพื่อปรับค่า

- Output ให้อยู่ในช่วงที่ต้องการ โดยในงานวิจัยนี้เลือกใช้ Sigmoid มีค่าระหว่าง 0-1

2.2 สถาปัตยกรรมโครงข่ายการเรียนรู้เชิงลึกที่ใช้ในการคำนวณเวกเตอร์คุณลักษณะ

การคำนวณเวกเตอร์คุณลักษณะในโครงข่ายการเรียนรู้เชิงลึก (Zoph et al., 2018) มักเกี่ยวข้องกับการใช้โครงข่ายประสาทเทียมเชิงลึก (Deep Neural Networks) ที่มีการฝึกอย่างมากขึ้นเพื่อสกัดคุณลักษณะที่สำคัญจากข้อมูลนำเข้า เวกเตอร์คุณลักษณะ (Feature Vectors) เป็นเวกเตอร์ที่แทนคุณลักษณะหรือลักษณะเฉพาะของข้อมูลนำเข้า ซึ่งเป็นข้อมูลที่ไม่สามารถนำเข้าแบบจำลองการเรียนรู้เชิงลึกได้โดยตรง (Mosavi et al., 2020)

เมื่อมีโครงข่ายประสาทเทียมที่ถูกฝึกด้วยชุดข้อมูลสามารถนำข้อมูลนำเข้าใหม่เข้าสู่โครงข่ายดังกล่าวเพื่อสร้างเวกเตอร์คุณลักษณะสำหรับข้อมูลดังกล่าวได้ เวกเตอร์คุณลักษณะนี้สามารถใช้ในการจำแนกหรือการประมวลผลต่อไปโดยสถาปัตยกรรมโครงข่ายการเรียนรู้เชิงลึกที่ใช้ในงานวิจัยนี้ มี 2 โครงข่าย ได้แก่ โครงข่ายแบบสังวัตนาการ และ โครงข่าย Transformer ดังรายละเอียดในข้อ 2.2.1 และ 2.2.2 ตามลำดับ

2.2.1 โครงข่ายแบบสังวัตนาการ (Convolutional Neural Networks, CNN)

โครงข่ายแบบสังวัตนาการ (Nash, 2015) เป็นสถาปัตยกรรมของการเรียนรู้เชิงลึกที่พัฒนามาเพื่อการประมวลผลภาพและข้อมูลที่มีลักษณะเป็นตาราง (Grid-like) โดยเฉพาะ ซึ่งได้รับความนิยมและใช้อย่างแพร่หลายในงานที่เกี่ยวข้องกับภาพ ซึ่งโครงข่ายแบบสังวัตนาการประกอบด้วยชั้นต่าง ๆ ที่มีลำดับการทำงานเฉพาะเพื่อสกัดคุณลักษณะ (Feature Extraction) โดยมีรายละเอียดดังนี้

1. Convolutional Layer ชั้นนี้ใช้ตัวกรอง (filter) ในการทำคอนโวลูชันกับข้อมูลนำเข้า เพื่อสกัดคุณลักษณะเด่น ๆ จากภาพ โดยตัวกรองภาพจะตรวจจับลักษณะต่าง ๆ ได้ เช่น ขอบ (edges), ลายเส้น (lines), และรูปร่าง (shapes) เป็นต้น

2. Activation Function หลังจากที่มีการคำนวณค่าของคอนโวลูชัน จะถูกนำผ่าน Activation Function เพื่อเพิ่มการไม่เชิงเส้นให้กับแบบจำลอง โดยทำในรูปแบบที่ซับซ้อนได้

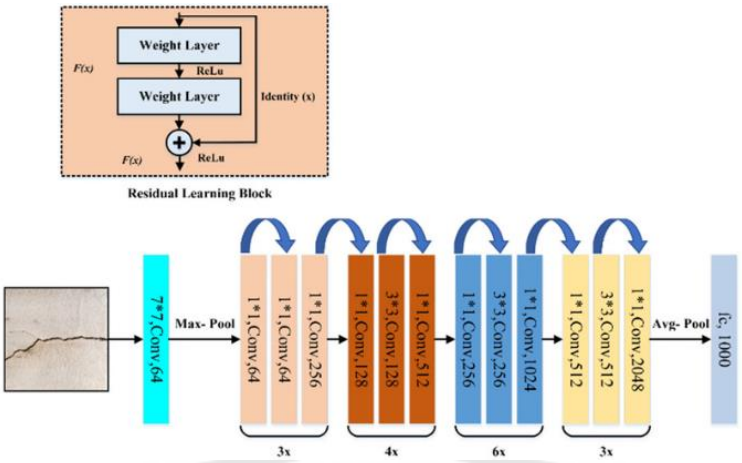
3. Pooling Layer ใช้สำหรับลดขนาดของข้อมูลที่ผ่านมาจากชั้นก่อนหน้า โดยที่ยังคงคุณลักษณะที่สำคัญ ทำให้ลดความซับซ้อนของแบบจำลองและลดการคำนวณได้

4. Fully Connected Layer เป็นชั้นที่แต่ละโหนดมีการเชื่อมต่อกันทั้งหมด เพื่อทำการแยกข้อมูล

โดยในงานวิจัยนี้ใช้โครงข่ายแบบสังวัฒนาการ ของแบบจำลอง ResNet50 ในการสกัดคุณลักษณะ เป็นกระบวนการที่มีการนำโครงสร้างของโครงข่ายการเรียนรู้เชิงลึกแบบ ResNet50 มาใช้ในการสกัดคุณลักษณะที่สำคัญจากข้อมูลนำเข้า มีรายละเอียด ดังนี้

<u>โครงสร้างของโครงข่ายการเรียนรู้เชิงลึกแบบ ResNet</u>

โครงข่ายการเรียนรู้เชิงลึกแบบ ResNet หรือ Residual Network มีการใช้ Skip Connections เพื่อแก้ปัญหาของการหา Gradient ซึ่งในแบบจำลองที่มีความลึกมาก ๆ ที่ทำให้เกิดปัญหาของการหา Gradient ซึ่งสามารถทำให้แบบจำลองที่มีความลึกมาก ๆสามารถฝึกได้ง่ายขึ้น ประกอบด้วยจำนวนชั้นที่ต่างกัน (He et al., 2016) แบบจำลอง ResNet มีหลายประเภทแบ่งตามจำนวนชั้น เช่น ResNet18, ResNet34, ResNet50, ResNet101 หรือ ResNet152 เป็นต้น หากจำนวนชั้นมากขึ้นนั้น จะมีความซับซ้อนมากขึ้น มีการใช้ทรัพยากรในการฝึกแบบจำลองและประมวลผลสูงขึ้นด้วย รวมทั้งมีโอกาสที่จะเกิดปัญหา Overfitting สูงขึ้นอีกด้วย



ภาพประกอบ 4 แสดงโครงสร้างของแบบจำลอง ResNet50

ที่มา: (wisdomml, 2023)

โครงสร้างของแบบจำลอง ResNet50 (Kaiming He, 2016) แสดงดังภาพประกอบ 4 และมีรายละเอียด ดังนี้

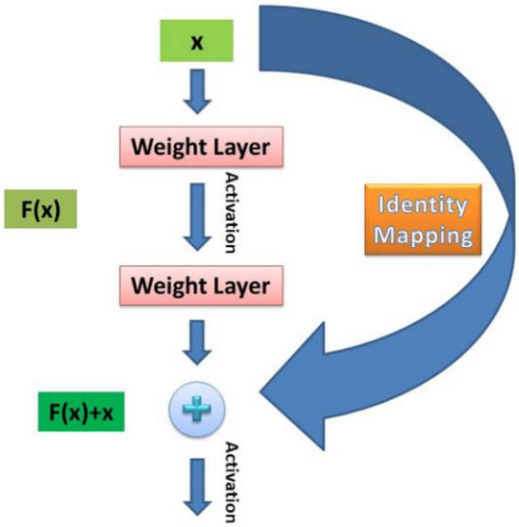
1. ชั้นนำข้อมูลเข้า (Input Layer) ชั้นที่รับข้อมูลภาพเข้าสู่แบบจำลอง

2. ชั้นที่ใช้ในการสกัดลักษณะเด่นของภาพ (Convolutional Layers) เป็นชั้นแรกของโครงข่ายที่ดำเนินการคอนโวลูชันบนภาพนำเข้า จากนั้นตามด้วยชั้น Max Pooling ที่ลดขนาดของเอาต์พุตของชั้นคอนโวลูชัน และผลลัพธ์ของชั้น Max Pooling จะถูกส่งผ่านชุดของ Residual Blocks อีกด้วย

3. Residual Blocks ชั้นที่ประกอบด้วย Convolutional Layers และ Skip Connections เพื่อรองรับการส่งข้อมูลไปยังชั้นถัดไป ซึ่งแต่ละ Residual Blocks ประกอบด้วยชั้นคอนโวลูชันสองชั้น ที่แต่ละชั้นตามด้วยชั้น Batch Normalization และ Activation Function ประเภท ReLU ผลลัพธ์ของชั้นคอนโวลูชันที่สอง จะถูกเพิ่มเข้าไปในข้อมูลนำเข้าของ Residual Blocks แล้วผ่าน Activation Function ประเภท ReLU อีกครั้ง โดยผลลัพธ์ของ Residual Blocks จะถูกส่งต่อไปยัง Block ถัดไป

แนวคิดของ Skip Connection ในชั้น Residual Blocks (He et al., 2016) การข้ามการเชื่อมต่อ ซึ่งเป็นคุณลักษณะสำคัญของ ResNet50 สามารถเก็บรักษาข้อมูลจากชั้นก่อนหน้า ซึ่ง

ช่วยให้โครงข่ายเรียนรู้การแสดงข้อมูลรับเข้าได้ดีขึ้นการเชื่อมต่อแบบข้ามจะดำเนินการโดยการเพิ่มเอาต์พุตของชั้นก่อนหน้าไปยังเอาต์พุตของชั้นถัดไป แสดงดังภาพประกอบ 5



ภาพประกอบ 5 แสดงการทำ Skip Connection ในชั้น Residual Blocks
ที่มา: (Giannopoulos et al., 2020)

2.2.2 โครงข่ายแบบ Transformer

โครงสร้าง Transformer เป็นแบบจำลองเชิงลำดับ (sequence model) ที่ถูกพัฒนาขึ้นมาโดย Google Research เพื่อการประมวลผลข้อมูลที่เป็นลำดับอย่างมหาศาล เช่น ประมวลผลภาษาธรรมชาติ (Natural Language Processing - NLP) และการแปลภาษา เป็นต้น โดย Transformer ได้รับความนิยมสูงสุดจากแบบจำลองการแปลภาษาหลายภาษา (multilingual translation) (Vaswani et al., 2017) โครงสร้างของ Transformer ประกอบด้วยส่วนประกอบหลัก ๆ ดังนี้

1. ตัวเข้ารหัส (Encoder) ประกอบด้วยหลายชั้นที่ใช้เปลี่ยนข้อมูลนำเข้าเป็นเวกเตอร์คุณลักษณะที่เกี่ยวข้องกับแต่ละคำหรืออีกสิ่งหนึ่งในลำดับข้อมูลนั้น ๆ โดยใช้สองเทคนิคหลักคือ Multi-Head Self-Attention และ Position-wise Feed-Forward Networks

2. ตัวถอดรหัส(Decoder) ประกอบด้วยหลายชั้นที่ใช้เปลี่ยนเวกเตอร์คุณลักษณะที่เกี่ยวข้องกับแต่ละคำหรือสิ่งใดสิ่งหนึ่งในลำดับข้อมูลเหล่านี้เป็นเวกเตอร์ข้อมูลข้าม

ฟีเจอร์ของตัวเข้ารหัส โดยใช้ Multi-Head Self-Attention, ตามด้วย Multi-Head Cross-Attention และ Position-wise Feed-Forward Networks

3. Multi-Head Self-Attention เป็นเทคนิคสำคัญที่ใช้ในการให้แบบจำลองสามารถสังเกตการณ์และความสัมพันธ์ระหว่างคำหรือสิ่งที่อยู่ในลำดับข้อมูลเพื่อทำนายข้อมูลออกมา โดยมีหลายหัวที่ใช้ความสัมพันธ์และลักษณะต่าง ๆ เพื่อสกัดคุณลักษณะที่สำคัญ

4. Multi-Head Cross-Attention เป็นเทคนิคที่ใช้ในการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลจากตัวเข้ารหัสกับข้อมูลจากตัวถอดรหัสเพื่อให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ระหว่างคำหรือสิ่งที่อยู่ในลำดับข้อมูลทั้งสอง

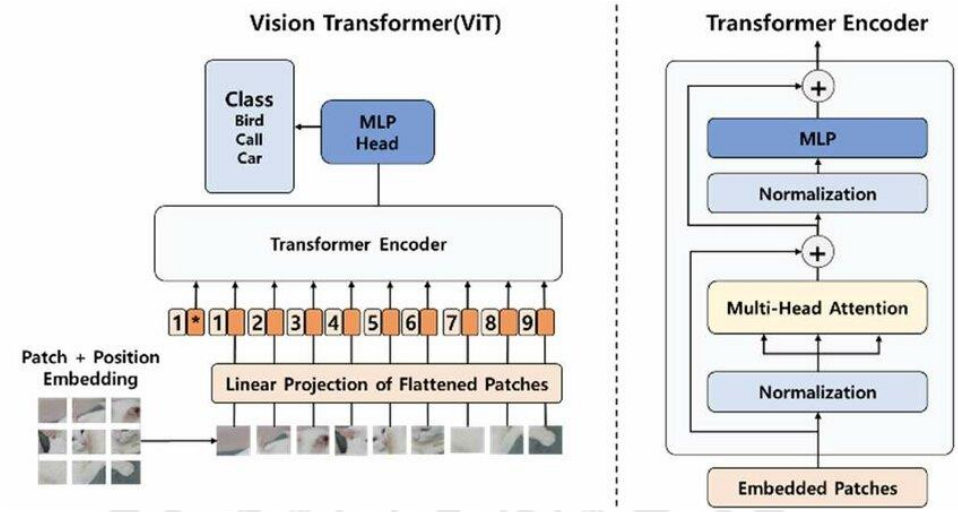
5. Position-wise Feed-Forward Networks เป็นเครือข่ายที่มีการเรียนรู้และแปลงข้อมูลที่สกัดมาจาก Multi-Head Self-Attention หรือ Multi-Head Cross-Attention เพื่อให้ได้เวกเตอร์คุณลักษณะที่เกี่ยวข้องกับแต่ละตำแหน่งในลำดับข้อมูล

โดยในงานวิจัยนี้ได้ในการใช้โครงข่ายTransformer ของแบบจำลอง Vision Transformer ในการสกัดคุณลักษณะเป็นกระบวนการที่มีการนำโครงสร้างของโครงข่ายการเรียนรู้เชิงลึกแบบ Vision Transformer เพื่อใช้ในการสกัดคุณลักษณะที่สำคัญออกมาจากข้อมูลนำเข้า มีรายละเอียด ดังนี้

<u>โครงสร้างของโครงข่ายการเรียนรู้เชิงลึกแบบ Vision Transformer</u>

Vision Transformer (ViT) เป็นแบบจำลองที่สร้างขึ้นเพื่อรองรับการประมวลผลภาพ โดยใช้โครงสร้าง Transformer ซึ่งประมวลผลข้อมูลแบบต่อเนื่อง (Sequential Data) โดยมีความสามารถในการจัดการกับลำดับของข้อมูลที่ไม่มีความเกี่ยวข้องกัน (Non-Sequential Data) อย่างมีประสิทธิภาพ แบบจำลองนี้ได้รับความนิยมมากในการประมวลผลภาษาธรรมชาติ (Natural Language Processing) และงานที่เกี่ยวข้องกับการทำนายลำดับของข้อมูล เช่น เพื่อแปลภาษา, สร้างคำบรรยายภาพ หรือ การแบ่งประเภทของข้อความ เป็นต้น รวมทั้งการนำมาใช้ในการประมวลผลภาพด้วยแบบจำลอง Vision Transformer (ViT) โดยเป้าหมายหลักของแบบจำลองคือการทำนายคุณลักษณะหรือลักษณะของภาพให้แม่นยำโดยใช้ตัวแปรเข้า (Input Embeddings) ที่เป็นพิกเซลของภาพแทนการใช้ Patch เหมือนกับการใช้ Convolutional Layer ในแบบจำลอง CNN (Convolutional Neural Network) และใช้หัว (Heads) สำหรับการคำนวณความสัมพันธ์ระหว่างตัวแปรเข้าเพื่อทำนายคุณลักษณะของภาพ นำ

ผลลัพธ์จากหัวแต่ละตัวมาผสมกันเพื่อให้ได้ผลลัพธ์สุดท้ายที่แม่นยำและเชื่อถือได้ (Alexey Dosovitskiy et al., 2021)



ภาพประกอบ 6 แสดงโครงสร้างของแบบจำลอง Vision Transformer
ที่มา: (Bang et al., 2023)

โครงสร้างหลักของ Vision Transformer ประกอบด้วย

1. Patch Embeddings ภาพถูกแบ่งเป็นช่องเล็ก ๆ ที่เรียกว่า Patches และแต่ละ Patch จะถูกแปลงเป็นเวกเตอร์ (Patch Embeddings) ก่อนถูกนำเข้าเป็นลำดับของข้อมูลในในบริบททางภาษา

2. Positional Encodings เพื่อให้แบบจำลองเข้าใจตำแหน่งของแต่ละ patch ในภาพ จึงต้องมี Positional Encodings เพื่อระบุตำแหน่งของแต่ละเวกเตอร์ในลำดับ

3. Transformer Encoder ข้อมูลที่ถูกแปลงจาก Patches Embeddings และ Positional Encodings จะถูกนำเข้าเป็นข้อมูลในแบบจำลอง Transformer โดยมีหลายๆ ชั้นของ Transformer Encoder ที่ใช้ในการประมวลผล

2.3 การฝึกแบบจำลองโครงข่ายการเรียนรู้เชิงลึก

2.3.1 การฝึกแบบจำลองจากแรกเริ่ม (Trained from Scratch)

ในกระบวนการฝึกแบบจำลองจากแรกเริ่ม (LeCun Y, 2015) จะเริ่มต้นด้วยแบบจำลองที่ยังไม่มีการฝึกอย่างใด ๆ บนชุดข้อมูลที่มีอยู่ แบบจำลองจะถูกฝึกตามชุดข้อมูลที่ใช้ในงานเฉพาะ

โดยใช้กระบวนการการปรับค่าพารามิเตอร์ (parameter) ต่าง ๆ ของแบบจำลองให้เข้ากับข้อมูล ซึ่งการฝึกใช้เวลาและทรัพยากรการคำนวณมาก เนื่องจากแบบจำลองต้องเรียนรู้คุณลักษณะของข้อมูลตั้งแต่ต้น

2.3.2 การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ (Pretrained Model)

การใช้แบบจำลองการเรียนรู้เชิงลึกที่ถูกฝึกสอนด้วยข้อมูลจำนวนมากและสกัดคุณลักษณะจากแบบจำลอง เพื่อนำมาปรับใช้ในงานที่มีความคล้ายคลึงกันโดยใช้การถ่ายโอนการเรียนรู้ (Transfer Learning) (Alex Krizhevsky, 2012) เป็นเทคนิคที่นิยมอย่างแพร่หลายในวงกว้างการถ่ายโอนการเรียนรู้ช่วยให้สามารถใช้ประโยชน์จากความรู้และความสามารถของแบบจำลองที่ถูกฝึกสอนไว้ก่อนหน้านี้ เพื่อใช้ในงานที่มีลักษณะและความต้องการที่คล้ายคลึงกันและเพื่อประสิทธิภาพที่ดียิ่งขึ้น

2.3.3 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม (Fine-Tuning)

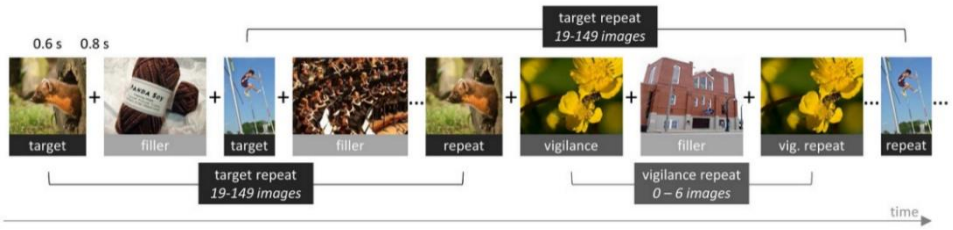
การปรับแต่งแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น (Fine-Tuning) (Yosinski, 2014) เป็นกระบวนการที่นำแบบจำลองที่ถูกฝึกมาจากชุดข้อมูลหนึ่งมาใช้ในงานอื่น ๆ โดยการปรับแต่งหรือเรียนรู้คุณลักษณะของข้อมูลในงานใหม่ กระบวนการนี้ช่วยให้แบบจำลองมีประสิทธิภาพในการทำงานใหม่โดยไม่ต้องเริ่มต้นฝึกแบบจำลองใหม่แบบจำลองจากแรกเริ่ม ซึ่งมักจะเป็นวิธีที่มักจะใช้เป็นส่วนมากในการปรับใช้แบบจำลองในงานที่มีการใช้แบบจำลองในหลาย ๆ งานได้ดีโดยไม่ต้องฝึกจากแรกเริ่มใหม่

2.4 งานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยนี้ได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้อง อาทิ ชุดข้อมูลที่ใช้ในการทำนายการจดจำภาพ ด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ตลอดไปจนถึงการใช้เทคนิคอื่น ๆ เพิ่มเติม โดยมีรายละเอียด ดังนี้

ชุดข้อมูลสำหรับที่ใช้การทำนายการจดจำภาพนั้นจาก (Khosla et al., 2015) ได้กล่าวถึงชุดข้อมูล “LaMeM” ซึ่งเป็นชุดข้อมูลความจำภาพที่มีคำอธิบายประกอบเชิงความหมายที่ใหญ่ที่สุด ประกอบด้วย 60,000 ภาพจากแหล่งที่มาที่หลากหลายโดยใช้ โครงข่ายประสาทเทียม (CNN) และคุณสมบัติเชิงลึกที่ปรับแต่งอย่างละเอียดเพื่อประเมินความสามารถในการจดจำของ

ภาพ มีประสิทธิภาพเหนือกว่าคุณสมบัติอื่น ๆ และอันดับ Spearman เท่ากับ 0.64 ซึ่งมีความสอดคล้องกับมนุษย์ วิจัยนี้ให้เห็นถึงการประมาณความสามารถในการจดจำภาพที่ดีในชั้นต่าง ๆ ความสามารถในการจดจำตำแหน่งและคุณสมบัติของภาพ มีการปรับรายละเอียดในส่วนต่าง ๆ ของแบบจำลอง เพื่อประเมินความสามารถในการจดจำของภาพซึ่งมีประสิทธิภาพดีกว่า ในแบบจำลองอื่น ๆ การวิจัยอาศัยคำอธิบายประกอบเชิงความหมายเป็นหลักเพื่อสร้างชุดข้อมูล LaMem ซึ่งอาจมีส่วนในความสามารถในการจดจำในแต่ละบุคคลได้ โดยการวิจัยไม่ได้สำรวจผลกระทบของปัจจัยบริบทเช่น คำบรรยายภาพหรือข้อความโดยรอบต่อความสามารถในการจดจำภาพ ในขณะเดียวกัน (Goetschalckx & Wagemans, 2019)ได้นำเสนอชุดข้อมูล "MemCat" ที่ประกอบด้วยภาพตามหมวดหมู่ 5 หมวดหลัก จำนวนรวม 10,000 ภาพ ที่เกี่ยวข้องกับการจดจำในมุมมองกว้างกว่าชุดข้อมูลอื่น ๆ โดยทั้ง 5 หมวดหมู่ นี้ได้แก่ Animal, Food, Sport, Vehicle, Landscape และแบ่งออกเป็นหมวดหมู่ย่อยซึ่งภาพถูกรวบรวมจากชุดภาพต้นฉบับที่มีอยู่ซึ่งมีคำอธิบายประกอบเชิงความหมาย ในชุดข้อมูลนี้มีการศึกษาปัจจัยความแปรปรวนและยังกระตุ้นควบคุมเชิงความหมายเกี่ยวกับความสัมพันธ์ของการจดจำ การรวบรวมคะแนนการจดจำภาพจากชุดข้อมูลในงานวิจัย ได้ทำผ่านแพลตฟอร์มที่ชื่อว่า Amazon Mechanical Turk โดยผู้เข้าร่วมทำแบบการจดจำภาพได้ค่าตอบแทนเป็นเงิน และผู้เข้าร่วมมีจำนวนที่เพียงพอที่เข้าร่วมในการทำแบบการจดจำ สามารถมั่นใจในข้อมูลที่เชื่อถือได้ ซึ่งในการเข้าร่วมครั้งนี้ผู้เล่นจะเห็นภาพแสดงขึ้นตามลำดับ เป็นเวลา 0.6 วินาที ช่องว่างระหว่างภาพแสดงเป็นเวลา 0.8 วินาที ซึ่งคะแนนการจดจำภาพจากงานวิจัยนี้ ได้ถูกเก็บรวบรวมจากชุดข้อมูลดังกล่าวโดยใช้งานการทดสอบความจำแบบการตรวจสอบซ้ำ โดยแต่ละคะแนนความจำเกิดขึ้นจากการตอบสนองของผู้เข้าร่วมโดยเฉลี่ยประมาณ 99 คน ซึ่งคะแนนความจำที่เก็บรวบรวมนั้นแสดงความสอดคล้องกันในหมู่ผู้เข้าทำแบบทดสอบนี้ภาพ วิธีการคำนวณคะแนนนั้นใช้มาตรการต่างๆ เช่น อัตราการโดนจับและอัตราการโดนจับที่ถูกแก้ไข เพื่อให้สามารถตอบสนองต่อข้อผิดพลาดเทียบเท่าได้



ภาพประกอบ 7 กระบวนการตรวจจับภาพซ้ำในงานวิจัย "MemCat: a new Category-based image set quantified on memorability"

ที่มา : (Goetschalckx & Wagemans, 2019)

ทั้งนี้ (Isola et al., 2014) มุ่งเน้นไปที่การทำนายความสามารถในการจดจำของภาพโดยมีการวิเคราะห์คุณสมบัติของภาพเพื่อนำไปสู่การจดจำภาพ ในงานวิจัยนี้กล่าวถึงการจดจำภาพสามารถทำได้โดยการใช้เทคนิค Vision Computer และให้ข้อมูลในเชิงลึกเกี่ยวกับคุณสมบัติของการจดจำภาพ ซึ่งสามารถประยุกต์ใช้กับงานหลายด้าน เช่น การสร้างภาพ การถ่ายภาพ การศึกษา และUser Interface Design ในงานวิจัยพบว่าเทคนิค Computer Vision สามารถใช้ในการทำนายความจำของภาพ ทำความเข้าใจเกี่ยวกับความจำ ออกแบบระบบที่มีความคล้ายคลึงกับการจดจำของมนุษย์ ซึ่งการวิเคราะห์คุณสมบัติและคุณลักษณะของภาพต่าง ๆ เป็นปัจจัยหลัก ๆ ที่ทำให้ภาพน่าจดจำ โดยการจดจำแบ่งออกเป็น 3 รูปแบบได้แก่ การรับรู้, ระยะสั้น และระยะยาว ดังนั้นในงานวิจัยต้องการวัดความสามารถการจดจำของภาพและแนวโน้มของภาพที่มีการจดจำดีที่สุด ซึ่งกระบวนการในการวัดผลการจดจำเบื้องต้นนั้นแบ่งออกเป็น 2 วิธี ได้แก่ การสอบถามผู้สังเกตการณ์ว่าเคยเห็นภาพนี้มาก่อนหรือไม่ (วิธีการตรวจจับซ้ำ) และทางเลือกบังคับอีกสองทางเลือก (วิธีการสิ่งใดสิ่งหนึ่ง) โดยในงานวิจัยนี้เลือกวิธีการตรวจจับซ้ำ ซึ่งการตรวจจับภาพรูปแบบนี้ช่วยให้สามารถทดสอบความผิดปกติของภาพที่กำหนดได้อย่างชัดเจน งานวิจัยนี้จึงให้นิยาม “ความทรงจำ” โดยการวิเคราะห์แต่ละภาพว่าผู้เข้าร่วมสามารถตรวจจับรูปซ้ำได้อย่างถูกต้องหรือไม่

ภาพประกอบ 8 แสดงตัวอย่างชุดข้อมูลของการจดจำภาพระดับสูง (ด้านขวา) และการจดจำภาพระดับต่ำ (ด้านซ้าย)

ที่มา: (Isola et al., 2014)

นอกจากนี้ยังมีบทความที่เกี่ยวข้องกับการนำเทคนิคการเรียนรู้เชิงลึกมาใช้ในการทำนายการจดจำภาพ (Vaswani et al., 2017) ที่ได้ นำเสนอโครงสร้าง Transformer สำหรับการประมวลผลข้อความที่ไม่ใช้ Recurrent Networks เช่น LSTM หรือ GRU แต่ใช้เทคนิค Attention เพื่อเรียนรู้ความสัมพันธ์ระหว่างคำในประโยค โครงสร้าง Transformer ประกอบด้วยส่วน Encoder และ Decoder โดยส่วน Encoder เป็นที่ใช้เรียนรู้เพื่อเข้าใจความสัมพันธ์ระหว่างคำในประโยค ส่วน Decoder จะใช้เพื่อสร้างประโยคที่ถูกแปลและควบคุมโดยคำแนะนำจากส่วน Encoder งานวิจัยนี้เป็นหนึ่งในงานที่เกี่ยวข้องกับภาษาธรรมชาติและการประมวลผลข้อความที่สำคัญอย่างกว้างขวาง และ (Alexey Dosovitskiy et al., 2021) ได้มีการสำรวจการประยุกต์ใช้โครงสร้างในบริบททางภาษา ที่ใช้กันอย่างแพร่หลายในการประมวลผลภาษาธรรมชาติ มีการใช้งานที่จำกัดในการประมวลผลรูปภาพ โดยวิธีการเดิมในการประมวลผลรูปภาพนั้น มักจะรวม Attention กับ Convolutional Networks หรือแทนบางส่วนของ Convolutional Networks ด้วย Attention โดยใช้ในบริบททางภาษา ที่ใช้งานโดยตรงกับลำดับของ Image Patches สำหรับงานการจำแนกภาพ เรียกแบบจำลองว่า Vision Transformer (ViT) ที่ทำการฝึกรูปแบบการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะด้วยข้อมูลจำนวนมากและโอนย้ายไปยัง Benchmarks ซึ่งการจำแนกภาพต่าง ๆ ได้ผลลัพธ์เป็นอย่างดี เมื่อเปรียบเทียบกับ Convolutional Networks ที่เป็น State-of-the-art และยังใช้ทรัพยากรคำนวณน้อยกว่าอีกด้วย

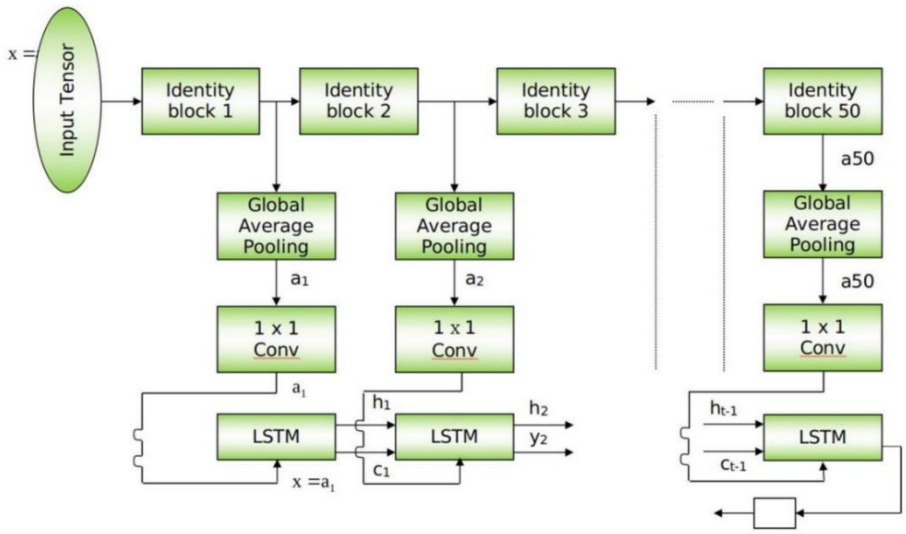
เทคนิคเฉพาะสำหรับการทำนายการจดจำภาพ (Dubey et al., 2015) ได้ศึกษาและวิเคราะห์ความจำของวัตถุในภาพ และสำรวจความสัมพันธ์ระหว่างความจำของวัตถุและภาพ งานวิจัยนี้ได้รวบรวมข้อมูลจริงเพื่อเข้าใจปัจจัยที่มีผลต่อความจำของวัตถุ เช่น ประเภทของวัตถุ

และความชัดของภาพและยังสำรวจความสัมพันธ์ของการจดจำระหว่างภาพกับวัตถุ โดยใช้แบบจำลองการเรียนรู้เชิงลึก ConvNet ที่ได้รับการฝึกบนชุดข้อมูล ImageNet ถูกนำมาใช้เพื่อทำนายความสามารถในการจดจำวัตถุในภาพ ซึ่งคุณสมบัติรูปภาพ ยังใช้ SIFT และ HOG เพื่อทำนายความสามารถในการจดจำวัตถุ จากนั้นได้รวบรวมชุดข้อมูลเกณฑ์สำหรับการทำนายความจำของวัตถุโดยอัตโนมัติ ซึ่งงานวิจัยนี้ได้พบว่าค่าความจำต่อรูปภาพที่สูงที่สุดนั้นมีความสัมพันธ์อย่างมีนัยสำคัญกับความจำของภาพ กล่าวคือ วัตถุที่สำคัญที่สุดในภาพมีบทบาทสำคัญในการกำหนดความจำโดยรวม เช่น เครื่องมือ, เฟอร์นิเจอร์, ธรรมชาติ และคน เป็นต้น

ภาพประกอบ 9 แสดงตัวอย่างคะแนนการจดจำของวัตถุภายในภาพ

ที่มา: (Dubey et al., 2015)

(Arockia Praveen et al., 2021) ได้เสนอสถาปัตยกรรมการเรียนรู้เชิงลึกแบบใหม่ที่เรียกว่า ResMem-Net ซึ่งเป็นผสมผสานระหว่าง LSTM และ CNN โดยแบบจำลอง ResMem-Net มาจากโครงสร้างพื้นฐานของ ResNet (Residual Network) ซึ่งเป็นแบบจำลองการเรียนรู้เชิงลึกที่มีความซับซ้อนและมีประสิทธิภาพสูงในการประมวลผลภาพ เพื่อนำมาใช้ในการประมาณค่าความจำของภาพ โดยการปรับเปลี่ยนโครงสร้างและการเรียนรู้ที่เน้นไปที่การจำและการเรียนรู้จากข้อมูลที่เกี่ยวข้องกับความจำของภาพ ResMem-Net ใช้ข้อมูลจาก Hidden Layers ของ CNN เพื่อคำนวณคะแนนความสามารถในการจดจำของภาพ ซึ่งแบบจำลองนี้ได้รับการฝึกอบรมและประเมินผลโดยใช้ชุดข้อมูล Large-scale Image Memorability (LaMem) ผลลัพธ์พบว่ามีความสัมพันธ์อันดับ เท่ากับ 0.679 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เท่ากับ 0.011 ซึ่งมีประสิทธิภาพที่ดี



ภาพประกอบ 10 แสดงโครงสร้างแบบจำลอง ResMem-Net

ที่มา: (Arockia Praveen et al., 2021)

(Thomas & Thomas, 2023) กล่าวถึงการทำนายความสามารถในการจดจำของภาพจากคุณสมบัติภายในของภาพ มากกว่าประสบการณ์หรือลักษณะเฉพาะของผู้สังเกตแต่ละคน และในขณะที่แบบจำลอง ResNet ได้ถูกใช้อย่างแพร่หลายในการทำนายความสามารถในการจดจำภาพ อีกทั้งได้ศึกษาสร้างแบบจำลอง เรียกว่า ViTMem ซึ่งเป็นการจดจำแบบใหม่ที่ใช้ Vision Transformer และเปรียบเทียบประสิทธิภาพกับ ResMem ที่ใช้ ResNet โดยประเมินสิทธิภาพแบบจำลองจากค่าความสูญเสีย Mean Squared Error (MSE) และค่าสหสัมพันธ์ Spearman's rho (ρ)

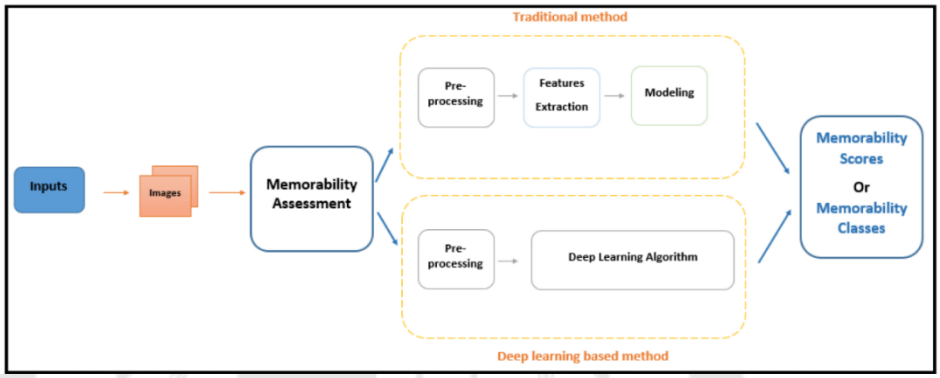
ตาราง 1 แสดงการเปรียบเทียบแบบจำลอง ResMem กับ แบบจำลอง ViTMem

Model	MSE Loss	MSE
ResMem	0.009	0.67
ViTMem	0.005	0.77

ที่มา: (Thomas & Thomas, 2023)

(Lahrache & Ouazzani, 2022) มุ่งเน้นสำรวจวิธีการและเทคนิคต่าง ๆ รวมไปถึงชุดข้อมูลที่ใช้ในการทำนายและวิเคราะห์ความสามารถในการจดจำภาพกล่าวถึงชุดข้อมูลและการประเมินที่ใช้สำหรับการประมาณความสามารถในการจำประสิทธิภาพ ตลอดจนโอกาสในการ

ปรับปรุงการทำนายการจดจำของภาพ พร้อมทั้งทบทวนการศึกษาความจำภาพโดยใช้ทั้งวิธีการแบบดั้งเดิมและวิธีการเรียนรู้เชิงลึกวิธีการแบบดั้งเดิมเกี่ยวข้องกับการใช้คุณสมบัติต่าง ๆ อาทิ สีพื้นผิว และรูปร่างในขณะที่ใช้การเรียนรู้เชิงลึกโดยใช้โครงข่ายแบบสังวัฒนาการเพื่อดึงลักษณะที่ซ่อนอยู่จากข้อมูลการศึกษาวิจัยเพื่อเปรียบเทียบประสิทธิภาพของแนวทางที่แตกต่างกันเหล่านี้ และเน้นถึงความก้าวหน้าที่เกิดขึ้นจากวิธีการเรียนรู้เชิงลึกในการปรับปรุงการคาดการณ์การจดจำภาพ



ภาพประกอบ 11 แสดงขั้นตอนทั่วไปของการจดจำภาพ

ที่มา : (Lahrache & Ouazzani, 2022)

จากการรวบรวมงานวิจัยการจดจำภาพมีการใช้วิธีการแบบดั้งเดิมเช่นการใช้ความหมายของฉากวัตถุลักษณะภาพพื้นฐานและคุณสมบัติทั่วไปสำหรับการทำนายความจำของภาพ และวิธีการที่ใช้การเรียนรู้เชิงลึกโดยดึงคุณลักษณะจากภาพโดยใช้โครงข่ายแบบสังวัฒนากการ และแสดงให้เห็นถึงความพัฒนาในการปรับปรุงการทำนายความสามารถในการจำได้

ตาราง 2 สรุปงานวิจัยที่ศึกษาทั้งหมด

งานวิจัย	แบบจำลอง	ชุดข้อมูล	Performance	
			MSE	Spearman's
(Khosla et al., 2015)	CNN	LaMem	-	0.64
(Isola et al., 2014)	SVR	ข้อมูลที่ติดป้ายกำกับเป็นบุคคลจำนวน 5,000 ภาพ	-	0.51
(Dubey et al., 2015)	Conv-net	ImageNet	-	0.7
(Arockia Praveen et al., 2021)	ResNet	LaMem	0.011	0.679
(Isola et al., 2014)	ViT	LaMem, FIGRIM	0.05	0.77

บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

3.1 กระบวนการทำงานของแบบจำลอง

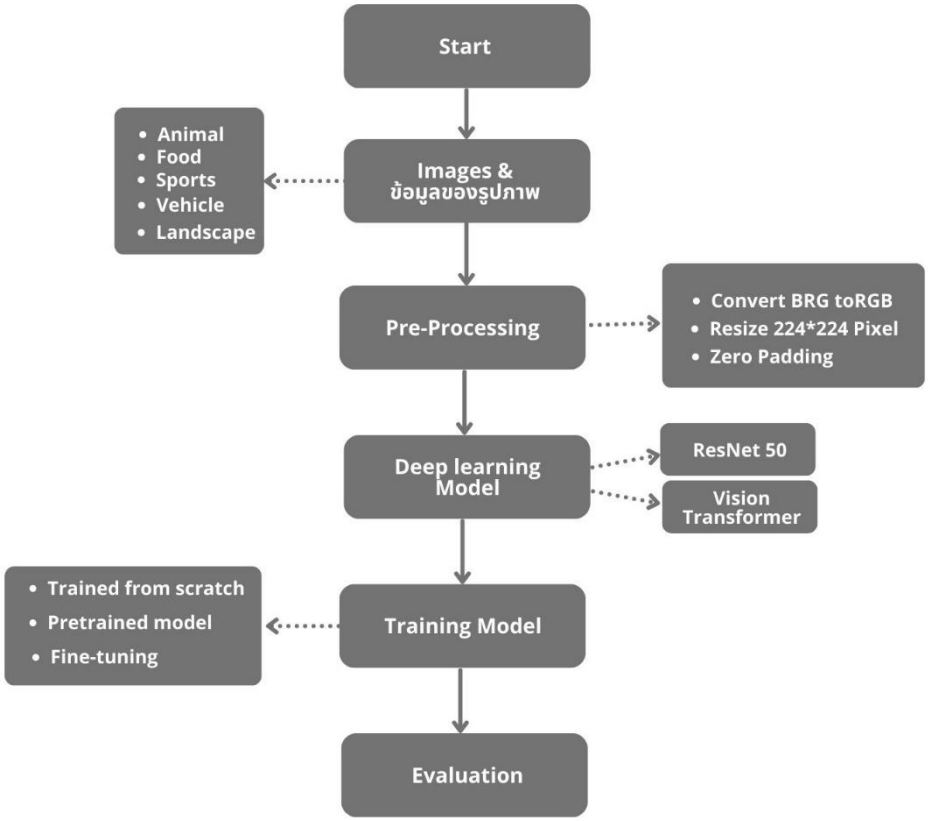
3.2 ชุดข้อมูลที่ใช้ในการทดลอง

3.3 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง

3.4 การสร้างแบบจำลอง

3.5 การประเมินผลแบบจำลอง

3.1 กระบวนการทำงานของแบบจำลอง



ภาพประกอบ 12 กระบวนการทำงานของแบบจำลอง

จากภาพภาพประกอบที่ 12 แสดงกระบวนการสร้างแบบจำลอง (Model Building Process) เริ่มต้นด้วยขั้นตอนการนำเข้าข้อมูลและจัดการข้อมูลก่อนเข้ากระบวนการสร้างแบบจำลอง และแบ่งข้อมูลออกเป็นข้อมูลฝึกสอน (Training Set) ชุดข้อมูลในการประเมินประสิทธิภาพของแบบจำลอง (Validation Set) และชุดข้อมูลทดสอบ (Test Set) เพื่อสร้างและทดสอบประสิทธิภาพแบบจำลอง ซึ่งสร้างแบบจำลองโดยใช้เทคนิคการเรียนรู้เชิงลึก โดยโครงสร้างจาก 2 แบบจำลอง ได้แก่ โครงข่ายแบบสังวัฒนาการ และโครงข่าย Transformer ซึ่งอาศัยการสกัดคุณลักษณะจากแบบจำลอง ResNet50 และแบบจำลอง Vision Transformer เพื่อการทำนายคะแนนการจดจำภาพในรูปแบบการทำนายแบบถดถอย

3.2 **ชุดข้อมูลที่ใช้ในการทดลอง**

ชุดข้อมูลที่ใช้สำหรับงานวิจัยการทำนายการจดจำภาพนั้น มีหลากหลายชุดข้อมูล โดยมีรายละเอียดดังนี้

ตาราง 3 สรุปและเปรียบเทียบชุดข้อมูลที่มีอยู่ของความสามารถในการจดจำภาพ

ชุดข้อมูล	จำนวน(รูป)	ป้ายกำกับ	รายละเอียดของชุดข้อมูล	ค่าความสัมพันธ์
LaMem	60000	6000	รวบรวมชุดข้อมูลรูปภาพหลากหลายรูปแบบเพื่อการประมวลผลทางรูป	0.68
MemCat	10000	10000	ชุดข้อมูลแบ่งออกเป็น 5 หมวดหมู่	0.78
FIGRIM	9428	1754	การจดจำภาพของมนุษย์ โดยใช้ภาพที่ถ่ายจาก Flickr	0.74

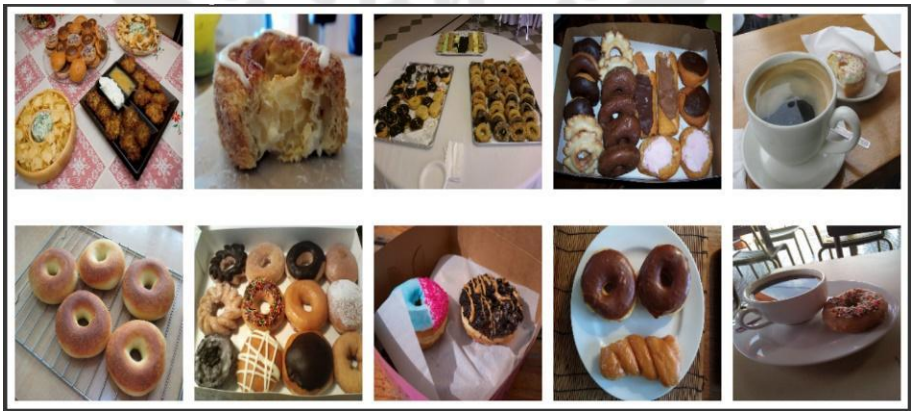
ที่มา : (Lahrache & Ouazzani, 2022)

จากตารางที่ 3 สรุปชุดข้อมูลเพื่อใช้ในการจดจำภาพดังกล่าวมีการเปรียบเทียบระหว่างชุดข้อมูลโดยมีรายละเอียดในการเปรียบเทียบ พบว่านอกจากชุดข้อมูล MemCat นั้นไม่มีชุดข้อมูลอื่นๆที่ทำการจำแนกคุณลักษณะแบ่งออกเป็นหมวดหมู่ อีกทั้งแยกป้ายกำกับจากรูปภาพในแต่ละหมวดหมู่ และมีค่าความสัมพันธ์ในระดับสูง ดังนั้น ผู้วิจัยจึงเลือกใช้ชุดข้อมูล “MemCat” เพื่อทำการทดลอง

MemCat (Goetschalckx & Wagemans, 2019) ประกอบด้วยชุดข้อมูลรูปภาพ และชุดข้อมูลที่เกี่ยวข้อกับรูปภาพ ซึ่งเป็นการรวมกันของ 4 ชุดข้อมูล ImageNet, COCO, SUNและV4 เนื่องจากชุดข้อมูลทั้ง 4 ชุดนี้ เป็นชุดข้อมูลขนาดใหญ่ มีความพร้อมใช้งานของคำอธิบายประกอบเชิงความหมาย และความพร้อมใช้งานของคำอธิบายประกอบแบบ Bounding Box

ชุดข้อมูลรูปภาพ Memcat แบ่งเป็น 5 หมวดหมู่ 10,000 รูป หมวดละ 2,000 รูป ได้แก่ Animal , Food, Sports , Landscape และ Vehicle

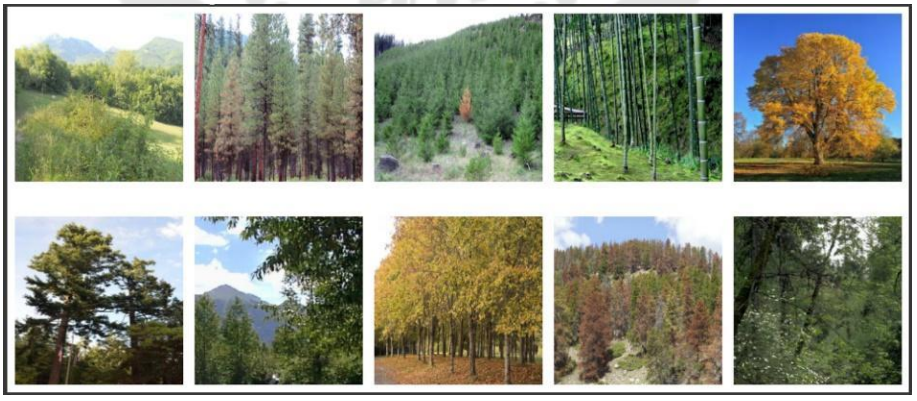
ภาพประกอบ 13 ภาพตัวอย่างชุดข้อมูลหมวด Animal



ภาพประกอบ 14 ภาพตัวอย่างชุดข้อมูลหมวด Food

ภาพประกอบ 15 ภาพตัวอย่างชุดข้อมูลหมวด Sports

ภาพประกอบ 16 ภาพตัวอย่างชุดข้อมูลหมวด Vehicle



ภาพประกอบ 17 ภาพตัวอย่างชุดข้อมูลหมวด Vehicle

3.3 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง

3.3.1 ชุดข้อมูล

เริ่มต้นด้วยการโหลดข้อมูลไฟล์รูปภาพ (jpeg) และไฟล์ข้อมูลที่เกี่ยวข้องกับรูปภาพ (CSV) จากนั้นทำการรวมไฟล์รูปกับไฟล์ข้อมูลรูปเข้าด้วยกัน

image_file	category
/content/animal/000000003481.jpg	animal
/content/animal/000000005745.jpg	animal
/content/animal/000000011552.jpg	animal
/content/animal/000000027439.jpg	animal
/content/animal/000000055601.jpg	animal

ภาพประกอบ 18 ตัวอย่างผลลัพธ์จากการรวมไฟล์รูปและข้อมูลที่เกี่ยวข้อง

3.3.2 การเตรียมรูปภาพ

หลังจากทำการรวมไฟล์รูปกับไฟล์ข้อมูลรูปภาพ เพื่อความเหมาะสม ต้องทำการปรับรูปภาพก่อนนำเข้าแบบจำลอง มีรายละเอียด ดังนี้

1. Zero Padding เพิ่มค่าศูนย์ (zero values) ลงไปรอบๆ ของข้อมูลเดิม เพื่อให้ข้อมูลที่เข้ามามีขนาดเท่ากันกับข้อมูลที่ออกมาหลังจากการส่งผ่านชั้นการประมวลผลต่างๆ ในแบบจำลอง

2. ปรับภาพในชุดข้อมูลปรับขนาดรูปภาพเป็น 224*224 พิกเซล เป็นขนาดภาพมาตรฐานช่วยลดภาระการประมวลผลของแบบจำลอง deep learning โดยลดเวลาที่ใช้ในการฝึกและทดสอบ และทำให้สามารถทำการประมวลผลบนข้อมูลภาพได้รวดเร็วมากขึ้น

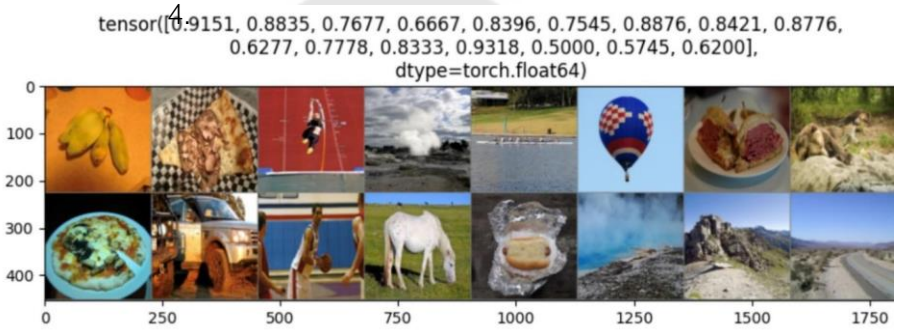
3. แปลงรูปภาพจาก BGR (Blue-Green-Red) เป็น RGB (Red-Green-Blue) เป็นกระบวนการสำคัญที่ช่วยให้ภาพมีความเหมาะสมสำหรับการประมวลผลและแสดงผล เพื่อแสดงผลภาพ RGB ช่วยให้รูปภาพแสดงผลอย่างถูกต้องและสวยงาม

3.3.3 การแบ่งข้อมูลเพื่อทำการทดลอง

รูปภาพและข้อมูลที่เกี่ยวข้อง อยู่ในรูปของไฟล์ csv แบ่งเป็น 2 รูปแบบ คือ รูปแบบ 5 หมวดหมู่ หมวดหมู่ละ 2,000 รูป และรูปแบบรวมทั้ง 5 หมวดหมู่ ทั้งหมด 10,000 รูป โดยแบ่งตามอัตราส่วนดังนี้

1. Train set: 90%
2. Validation Set: 5%
3. Test Set: 5%

4.



ภาพประกอบ 19 ตัวอย่างภาพในการฝึกแบบจำลอง

3.4 การสร้างแบบจำลอง

งานวิจัยนี้ผู้วิจัยได้ทำการวิเคราะห์โดยใช้ Google Colab - GPU และดำเนินการสกัดคุณลักษณะโดยใช้โครงข่ายแบบสังวัฒนาการและโครงข่าย Transformer ซึ่งใช้การสกัดคุณลักษณะจากภาพด้วยแบบจำลอง ResNet50 และ Vision Transformer โดยในการสกัดคุณลักษณะทั้งสองแบบดังกล่าวจะได้ ในชั้นสุดท้ายเป็นการทำนายแบบถดถอยเพื่อการจดจำภาพ ในงานวิจัยนี้ได้ใช้เทคนิคการเรียนรู้เชิงลึก 3 แบบ คือ แบบจำลอง ResNet50, Vision Transformer, และการรวมเวกเตอร์คุณลักษณะจากทั้ง 2 แบบจำลอง โดยได้เลือกใช้ไลบรารีของ PyTorch สร้างแบบจำลองเนื่องจากมีความยืดหยุ่นสำหรับการเรียนรู้เชิงลึก โดยได้ทำการฝึกแบบจำลองทั้งหมด 3 แบบ ดังนี้

3.4.1 การสร้างแบบจำลองจากศูนย์หรือต้นแบบ

สร้างทั้งหมด 3 แบบจำลอง ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน โดยกำหนดพารามิเตอร์ (Parameters) ดังนี้

1. Epochs: 50
2. Batch Size: 32
3. Activation Function: Sigmoid
4. Learning Rate: $1e^{-4}$
5. Optimizer: Adam

รายละเอียดของแต่ละชั้นของแบบจำลอง มีดังนี้

3.4.1.1 แบบจำลอง Vision Transformer

ชั้น Input Embedding จะถูกแบ่งออกเป็นขนาด 16x16 Flatten Patches โดยในที่นี้เราจะใช้ Convolutional (Conv2d) แทนการใช้ Linear Layer ซึ่งสามารถทดแทนกันได้ โดยจะต้องกำหนด kernel size และ stride ให้มีค่าเท่ากับ 16 (Patches Size ที่กำหนด) และทำการ Flatten ผลลัพธ์ออกมาโดยใช้คำสั่ง Rearrange ของไลบรารี่ Einops ที่ส่วนของ Transformer Encoder หลังจากนั้น Patch ที่แปลงแล้วจะถูกนำเข้าไปใน Transformer Encoder ซึ่งประกอบด้วยหลายชั้นของ Transformer blocks โดยแต่ละบล็อกจะประกอบไปด้วยการทำ Self-Attention และ Feedforward Neural Network ในการประมวลผลภาพ การใช้ Transformer ในภาพทำให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ระหว่างพิกเซลในภาพได้และในขั้นตอนสุดท้ายของแบบจำลองทำการรวมขั้นตอนทั้งหมดประกอบเข้าด้วยกันก่อนจะนำไป training ทั้ง PatchEmbedding และ TransformerEncoder

3.4.1.2 แบบจำลอง ResNet50

Convolution block เป็นส่วนย่อยที่สุด ที่อยู่ในแต่ละ Residual Block ประกอบไปด้วย Convolutional Layer และ Batch Normalize ภายใน Residual block จะมีการเรียกใช้ Convolution block (ConvBlock) ที่ได้ทำการสร้างก่อนหน้าโดยจะมีจำนวน 3 ConvBlock ในทุก ๆ Residual block จะมีการกำหนด output_chanel, kernel_size, Stride และ Padding

ส่วนของ 50 layer ส่วน Max Pooling (MaxPooling2D) มี 5 ชั้น ใช้พื้นที่วินโดว์ขนาด 3x3 และ Fully Connected Layers มี 2 ชั้น ที่ปลายสุดของแบบจำลอง ชั้นแรกมีจำนวน

input features (in_features) ทั้งหมด 2048 และมีจำนวน Output Features (out_features) ทั้งหมด 512 ส่วนชั้นที่สอง มีจำนวน Input Features ทั้งหมด 512 และจำนวน Output Features ทั้งหมด 1 โหนด

3.4.1.3 แบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50

ทำการเชื่อมต่อแบบจำลองระหว่าง Resnet กับ Transformer เข้าด้วยกัน โดยกำหนดให้มีการนำเข้า 2048 และ 768 ตามลำดับ ดังนั้นจะนำเข้าสู่แบบจำลองที่รวมเวกเตอร์คุณลักษณะของทั้งสองแบบจำลอง เท่ากับ 2816

3.4.2 การฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ

ฝึกแบบจำลองทั้งหมด 3 แบบจำลอง ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน โดยกำหนดพารามิเตอร์ (Parameters) เพื่อทดสอบแบบจำลอง ดังนี้

1. Epochs: 30
2. Batch Size: 32
3. Activation Function: Sigmoid
4. Learning Rate: $1e^{-4}$
5. Optimizer: Adam

รายละเอียดของแต่ละชั้นของแบบจำลอง มีดังนี้

3.4.2.1 แบบจำลอง Vision Transformer

ใช้แบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองคำนวณเวกเตอร์คุณลักษณะด้วย “vit_base_patch16_224_miil.in21k” การฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะแบบจำลองจะใช้ชุดข้อมูล ImageNet-21K และประกอบด้วยชั้นต่าง ๆ ที่ปรับแต่งให้เหมาะสำหรับงานการประมวลผลภาพ เช่น ชั้น Multi-head Self-attention, Position-wise Feedforward และ Layer Normalization

3.4.2.2 แบบจำลอง ResNet50

นำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะด้วย “resnet50.a1_in1k” กับชุดข้อมูล ImageNet ที่มีภาพจำนวนมากและคลาสหลากหลายอย่างจำนวน 1,000 คลาส โดยใช้กระบวนการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะเพื่อให้แบบจำลองเรียนรู้และเข้าใจลักษณะของภาพแต่ละคลาสในชุดข้อมูล ImageNet

3.4.2.3 แบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน

การต่อแบบจำลองระหว่าง Resnet กับ Transformer เข้าด้วยกันและเข้าสู่ชั้น Regression Head เพื่อเปรียบเทียบกับสองแบบจำลองก่อนหน้า

3.4.3 การฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม

ทำการปรับแต่งพารามิเตอร์ของแบบจำลองด้วยชุดข้อมูลเดิม เพื่อให้แบบจำลองมีประสิทธิภาพมากยิ่งขึ้น โดยโหลดน้ำหนัก (Load Weight) จากการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะของทั้ง 3 แบบจำลอง และ Unfreeze Weight ในชั้นสุดท้าย ซึ่งในไลบารี่ PyTorch สามารถทำได้โดยการตั้งค่า requires_grad ของพารามิเตอร์ในชั้นสุดท้ายที่ต้องการให้เรียนรู้ใหม่เป็น True และนำมารวมกับส่วน Regression Head และทำการฝึกแบบจำลอง

3.5 การประเมินผลแบบจำลอง

หลังจากฝึกทั้ง 3 รูปแบบ ส่วนสุดท้ายจะเป็นการทำนายแบบถดถอยซึ่งจัดอยู่ในกลุ่มของ Supervise Learning ซึ่งมีผลลัพธ์มีค่าเป็นตัวเลขต่อเนื่องกันหลายค่า โดยวิธีการทางสถิติจะใช้เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ประกอบด้วย ตัวแปรประมาณการหรือตัวแปรต้น (Predictor, Independent Variable, X) และตัวแปรตอบสนองหรือตัวแปรตาม (Response ,Dependent Variable, Y) และนำมามาวิเคราะห์ว่าปัจจัยหรือเป็นเหตุผลของซึ่งกันและกันหรือไม่ อย่างไร แทนด้วยสมการทางคณิตศาสตร์ ดังนี้

$$y = f(x)$$

หรือ

$$y = ax + b$$

โดยที่ x แทนข้อมูลนำเข้า (Input)

y แทนข้อมูลผลลัพธ์ที่ได้ (Output)

a แทนค่าคงที่ของสมการถดถอย หรือค่าจุดตัด (Intercept) แกน y ของสมการ

b ค่าสัมประสิทธิ์ การถดถอย (Regression Coefficient) ของ x

การทำนายแบบถดถอย สำหรับวิจัยฉบับนี้ได้วัดประสิทธิภาพทั้งหมด 4 ตัวชี้วัด ดังนี้

1. Mean Squared Error (MSE) เป็นค่าได้จากความคลาดเคลื่อนในการประมาณค่าของแบบจำลอง โดยคำนวณจากความแตกต่างของค่าทำนายได้กับค่าจริง ในรูปแบบยกกำลังสองของความแตกต่างแต่ละค่าและนำมาหาค่าเฉลี่ย ดังสมการ

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Yi - \widehat{Yi})^2$$

โดยที่ Yi แทนค่าจริง (Actual Value) ของ Samples Test ที่ i

$\widehat{Yi}$ แทนค่าที่ประมาณได้ หรือค่า Predict ของ Samples Test ที่ i

n แทน จำนวน Samples ทั้งหมด

2. Mean Absolute Error (MAE) หรือ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เป็นค่าเฉลี่ยของความแตกต่างสมบูรณ์ระหว่างค่าทำนายและค่าจริง หากค่า MAE นั้นมีค่าน้อย แสดงว่าค่าทำนายนั้นมีค่าใกล้เคียงกับค่าจริง ดังสมการ

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|yi - \hat{y}|$$

โดยที่

Yi แทนค่าจริง (Actual Value) ของ Samples Test ที่ i

$\widehat{Yi}$ แทนค่าที่ประมาณได้ หรือค่า Predict ของ Samples Test ที่ i

n แทนจำนวน Samples ทั้งหมด

3. R-Squared (R^2) เป็นค่าที่วัดความเหมาะสมของแบบจำลองในการอธิบายข้อมูล โดยคำนวณเปรียบเทียบความแปรปรวนระหว่างค่าทำนายกับค่าจริง ดังสมการ

- ค่าที่เข้าใกล้ 0 หมายถึง แบบจำลองนั้นไม่สามารถอธิบายข้อมูลใด ๆ ทั้งสิ้น
- ค่าที่เข้าใกล้ 1 หมายถึง แบบจำลองอธิบายข้อมูลได้เพียงพอทุกประการ
- ค่าระหว่าง 0 ถึง 1 หมายถึง ความเหมาะสมของแบบจำลองในการอธิบายข้อมูล โดยค่าที่มากขึ้นแสดงการอธิบายข้อมูลดีขึ้น

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

โดยที่

SS_{res} แทนค่าความแปรปรวนของความคลาดเคลื่อนที่อธิบายได้โดยแบบจำลอง (Sum of Squares of Residuals)

SS_{tot} แทนค่าความแปรปรวนของข้อมูลเฉพาะ (Total Sum of Squares)

4. Spearman's rho (Spearman correlation coefficient) เป็นค่าที่เกี่ยวข้องกับการคำนวณความสัมพันธ์ระหว่างตัวแปรโดยใช้ค่าความสัมพันธ์ที่ดีของสเปียร์แมน (Spearman correlation coefficient) ซึ่งบ่งชี้ถึงความสัมพันธ์ระหว่างลำดับของข้อมูล โดยทั่วไปแล้วค่าของ Spearman correlation coefficient จะอยู่ในช่วง -1 ถึง 1 โดยค่าบวกแสดงถึงความสัมพันธ์ที่เชิงบวก ค่าลบแสดงถึงความสัมพันธ์ที่เชิงลบ และค่าเป็นศูนย์แสดงถึงขาดความสัมพันธ์ การใช้ Spearman correlation coefficient เป็นที่นิยมในการวิเคราะห์ข้อมูลที่มีการจัดลำดับ เช่น การวิเคราะห์ข้อมูลที่มีการจัดอันดับของคะแนนหรือการจัดอันดับของผลการทดลอง เป็นต้น ดังสมการ

$$\rho = 1 - \frac{6\Sigma\, \mathrm{d}_i^2}{n(n^2-1)}$$

โดยที่

ρ แทนค่าSpearman Correlation Coefficient

d_i^2 แทนค่าความต่างระหว่างความลำบากที่ปรากฏในอันดับของข้อมูลที่เปรียบเทียบกันระหว่างคู่ของตัวแปร

n แทนจำนวนข้อมูล

บทที่ 4
ผลการดำเนินงานวิจัย

ในการวิจัยการทำนายการจดจำภาพด้วยเทคนิคการเรียนรู้เชิงลึก ดำเนินการสกัดคุณลักษณะจากแบบจำลอง Vision Transformer จากโครงสร้าง Transformer และ แบบจำลอง ResNet50 จากโครงสร้าง Convolutional Neural Network ซึ่งทั้งสองแบบจำลองมีวิธีการสกัดที่แตกต่างกันในการสกัดคุณลักษณะของรูปภาพ โดยในงานวิจัยใช้ข้อมูล 5 หมวดหมู่ ได้แก่ Animal, Food, Sports, Landscape และ Vehicle ผู้วิจัยได้ดำเนินการวิจัยโดยศึกษาตามขั้นตอนต่าง ๆ ตลอดจนวัดประสิทธิภาพ การวิจัยนี้ได้กำหนดการฝึกแบบจำลองไว้ได้ ดังนี้

4.1 การฝึกแบบจำลองจากแรกเริ่ม

4.2 การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ

4.3 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม

4.1 การฝึกแบบจำลองจากแรกเริ่ม

การเปรียบเทียบแบบจำลองโดยการฝึกแบบจำลองจากแรกเริ่ม นั้น ใช้เวลาในการสร้างแบบจำลองเป็นเวลานานผู้วิจัยจึงนำชุดข้อมูลแบบคละหมวดหมู่ ใช้แบบจำลองทั้งหมด 3 แบบจำลอง ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน เพื่อเปรียบเทียบผลลัพธ์กับการฝึกรูปแบบอื่น ๆ ซึ่งมีผลลัพธ์ ดังนี้

ตาราง 4 แสดงผลการทดลองการฝึกแบบจำลองจากแรกเริ่ม

Category	Model	MSE	R-square	MAE
Merge	Vision Transformer	1.1983	-69.0230	1.0840
	ResNet50	0.3530	-19.5287	0.5707
	Vision Transformer + ResNet50	0.7155	-40.8800	0.8010

จากตารางที่ 4 ผลการทดลอง พบว่าแบบจำลอง ResNet50 ผลลัพธ์ดีที่สุดโดยมีค่า MSE (Mean Squared Error) เท่ากับ 0.353 และ MAE (Mean Absolute Error) เท่ากับ 0.5707 และผลลัพธ์ในลำดับถัดมาเป็นแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน ได้ผลลัพธ์ค่า MSE เท่ากับ 0.7155 และ MAE เท่ากับ 0.801 ตามลำดับ สุดท้ายคือ แบบจำลอง Vision Transformer ได้ผลลัพธ์ค่า MSE เท่ากับ 1.1983 และ MAE เท่ากับ 1.084 อย่างไรก็ตาม ค่า R-square ของแบบจำลองทั้งสามอยู่ในระดับต่ำ แสดงให้เห็นว่าแบบจำลองไม่สามารถอธิบายข้อมูลได้อย่างเพียงพอ ซึ่งยังมีความจำเป็นต้องการปรับปรุงเพิ่มเติมเพื่อให้สามารถอธิบายข้อมูลได้อย่างเหมาะสมยิ่งขึ้นในการฝึกรูปแบบอื่น ๆ

4.2 การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ

การเปรียบเทียบโดยการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะทำการฝึกกับชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ แล้วนำผลลัพธ์ที่ได้มาคำนวณเพื่อหาค่าเฉลี่ยของแบบจำลอง เพื่อเปรียบเทียบและผู้วิจัยได้ทำการทดลองทั้งหมด 3 แบบจำลอง ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 ดังตาราง 5

ตาราง 5 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น ในชุดข้อมูลชุดข้อมูลฝึกสอนและชุดข้อมูลที่ใช้ในการประเมินผลของประสิทธิภาพของแบบจำลอง Vision Transformer

Category	Dataset	MSE	R-square	MAE
animal	Train	0.0047	0.7994	0.0409
	Validation	0.0002	0.9380	0.0130
food	Train	0.0001	0.8849	0.0090
	Validation	0.0001	0.9512	0.0087
landscape	Train	0.0004	0.9790	0.0202
	Validation	0.0001	0.9890	0.0087

ตาราง 5 (ต่อ)

Category	Dataset	MSE	R-square	MAE
sport	Train	0.0002	0.9580	0.0136
	Validation	0.0001	0.9860	0.0080
vehicle	Train	0.0004	0.9452	0.0170
	Validation	0.0006	0.9150	0.0191
Average	**Train**	**0.0012**	**0.9133**	**0.0202**
	Validation	**0.0002**	**0.9558**	**0.0115**
Merge	**Train**	**0.0037**	**0.8310**	**0.0500**
	Validation	**0.0001**	**0.9918**	**0.0078**

จากตาราง 5 ผลการทดลองในการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลสำหรับประเมินประสิทธิภาพของแบบจำลอง โดยแบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลอง Vision Transformer ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียดค่าตัวชี้วัด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.00370 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.00010 ค่า MSE อยู่ในระดับต่ำแสดงความแม่นยำ
- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.83100 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.99180 ซึ่งแสดงให้เห็นว่าแบบจำลองสามารถอธิบายข้อมูลได้ดีและมีความเชื่อถือได้ในการทำนายข้อมูลจริง
- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.05000 และในชุดข้อมูลในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.00784

ตาราง 6 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น ในชุดข้อมูลชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง ResNet50

Category	Dataset	MSE	R-square	MAE
animal	Train	0.0001	0.9630	0.0099
	Validation	0.0001	0.9760	0.0089
food	Train	0.0002	0.9247	0.0088
	Validation	0.0011	0.5298	0.0328
landscape	Train	0.0001	0.9890	0.0085
	Validation	0.0001	0.9962	0.0095
sport	Train	0.0001	0.9936	0.0050
	Validation	0.0010	0.7648	0.0298
vehicle	Train	0.0007	0.7800	0.0250
	Validation	0.0002	0.9250	0.0129
Average	**Train**	**0.0002**	**0.9301**	**0.0114**
	Validation	**0.0005**	**0.8384**	**0.0188**
Merge	**Train**	**0.0001**	**0.9919**	**0.0098**
	Validation	**0.0007**	**0.9745**	**0.0081**

จากตาราง 6 ผลการทดลองในการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง แบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลอง ResNet50 ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียดค่าตัวชี้วัด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0001 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.0007 ซึ่งค่า MSE อยู่ในระดับต่ำ ดังนั้นความแม่นยำของแบบจำลองในการทำนายอยู่ในระดับที่แม่นยำ

- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.9919 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.9745 ซึ่งพบว่าแบบจำลองสามารถอธิบายข้อมูลได้ดีและมีความเชื่อถือได้ในการทำนายข้อมูลจริง
- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0098 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.0081

ตาราง 7 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน

Category	Model	MSE	R-square	MAE
animal	Train	0.0007	0.9056	0.0221
	Validation	0.0005	0.9082	0.0186
food	Train	0.0004	0.9226	0.0149
	Validation	0.0009	0.9248	0.0181
landscape	Train	0.0020	0.8818	0.0404
	Validation	0.0003	0.9758	0.0113
sport	Train	0.0006	0.9159	0.0235
	Validation	0.0001	0.9598	0.0089
vehicle	Train	0.0004	0.9141	0.0169
	Validation	0.0001	0.9601	0.0079
Average	**Train**	**0.0008**	**0.9080**	**0.0235**
	Validation	**0.0004**	**0.9457**	**0.0130**
Merge	**Train**	**0.0006**	**0.8570**	**0.0171**
	Validation	**0.0002**	**0.9930**	**0.0136**

จากตาราง 7 ผลการทดลองของแบบจำลองในการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง โดยแบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่ และ ชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกันนั้น ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียดค่าตัวชี้วัด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.000591 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.000207
- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.857 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.993
- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0171 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.0136

โดยผลการทดลองการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ ในชุดข้อมูลทดสอบของแบบจำลอง Vision Transformer แบบจำลอง ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน ซึ่งผู้วิจัยได้วัดประสิทธิภาพ แสดงดังตาราง 8,9 และ 10 ตามลำดับ

ตาราง 8 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น ในชุดข้อมูลทดสอบของแบบจำลอง Vision Transformer

Category	MSE	R-square	MAE	Spearman's
animal	0.0001	0.9300	0.0129	0.9850
food	0.0001	0.9360	0.0087	0.6363
landscape	0.0004	0.9296	0.0192	0.9890
sport	0.0002	0.9319	0.0218	0.6401
vehicle	0.0003	0.5037	0.0131	0.8327
Average	**0.0002**	**0.8698**	**0.0151**	**0.6529**
Merge	**0.0003**	**0.9870**	**0.0076**	**0.9716**

จากตาราง 8 การทดลองด้วยแบบจำลอง Vision Transformer โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.8698 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0151 และค่า Spearman's เฉลี่ยอยู่ที่ 0.6529
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0003 และ R-square เฉลี่ยอยู่ที่ 0.987 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0076 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9716

ตาราง 9 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น ในชุดข้อมูลทดสอบของแบบจำลอง ResNet50

Category	MSE	R-square	MAE	Spearman's
animal	0.0002	0.9355	0.0084	0.9774
food	0.0003	0.9350	0.0108	0.5475
landscape	0.0001	0.9795	0.0068	0.8981
sport	0.0004	0.9265	0.0370	0.7968
vehicle	0.0005	0.9845	0.0150	0.8611
Average	**0.0003**	**0.9522**	**0.0156**	**0.8162**
Merge	**0.0002**	**0.9945**	**0.0065**	**0.9839**

จากตาราง 9 การทดลองด้วยแบบจำลอง ResNet50 โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0003 และ R-square เฉลี่ยอยู่ที่ 0.9522 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0156 และค่า Spearman's เฉลี่ยอยู่ที่ 0.8162
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.9945 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0065 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9839

ตาราง 10 แสดงผลการทดลองแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่น ในชุดข้อมูลทดสอบของแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน

Category	MSE	R-square	MAE	Spearman's
animal	0.0002	0.9708	0.0113	0.8503
food	0.0001	0.8760	0.0091	0.9083
landscape	0.0001	0.9880	0.0075	0.9642
sport	0.0001	0.9405	0.0100	0.8571
vehicle	0.0001	0.9879	0.0078	0.9642
Average	**0.0002**	**0.9526**	**0.0091**	**0.9088**
Merge	**0.0001**	**0.9360**	**0.0101**	**0.9761**

จากตาราง 10 การทดลองด้วยแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.95264 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0091 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9088
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0001 และ R-square เฉลี่ยอยู่ที่ 0.936 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0101 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9761

จากผลการทดลองการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่า MSE, R-square และMAE และ Spearman's ระหว่างแบบจำลอง Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน ได้แสดงให้เห็นว่าในภาพรวมนั้นแบบจำลองทั้งสามรูปแบบทำงานได้ดีกับชุดข้อมูลแบบคละหมวดหมู่

4.3 การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม (Fine-Tuning)

การเปรียบเทียบการนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติมทำการฝึกกับชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ แล้วนำผลลัพธ์ที่ได้มาคำนวณเพื่อหาค่าเฉลี่ยของแบบจำลองและเปรียบเทียบกัน โดยผู้วิจัยได้ทำการทดลองทั้งหมด 3 แบบจำลอง ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 ดังนี้

ตาราง 11 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง Vision Transformer

Category	Dataset	MSE	R-square	MAE
animal	Train	0.0007	0.9710	0.0190
	Validation	0.0001	0.9630	0.0107
food	Train	0.0001	0.9050	0.0075
	Validation	0.0002	0.9200	0.0097
landscape	Train	0.0002	0.9898	0.0120
	Validation	0.0002	0.9850	0.0100
sport	Train	0.0002	0.9716	0.0102
	Validation	0.0001	0.9861	0.0071
vehicle	Train	0.0009	0.8915	0.0206
	Validation	0.0003	0.9460	0.0154
Average	**Train**	**0.0004**	**0.9458**	**0.0139**
	Validation	**0.0002**	**0.9600**	**0.0106**
Merge	**Train**	**0.0002**	**0.9930**	**0.0094**
	Validation	**0.0003**	**0.9760**	**0.0140**

จากตาราง 11 ผลการทดลองการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลสำหรับประเมินประสิทธิภาพของแบบจำลอง โดยแบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลอง Vision Transformer ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.00016 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.00029
- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.99300 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.97600 ซึ่งแสดงให้เห็นว่าแบบจำลองสามารถอธิบายข้อมูลได้ดีและมีความเชื่อถือได้ในการทำนายข้อมูลจริง
- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.00939 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.01400

ตาราง 12 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง ResNet50

Category	Dataset	MSE	R-square	MAE
animal	Train	0.0007	0.7588	0.0252
	Validation	0.0003	0.9451	0.0143
food	Train	0.0001	0.9345	0.0100
	Validation	0.0003	0.8572	0.0179
landscape	Train	0.0001	0.9898	0.0090
	Validation	0.0001	0.9949	0.0111
sport	Train	0.0009	0.9467	0.0238
	Validation	0.0003	0.9267	0.0156
vehicle	Train	0.0004	0.8801	0.0160
	Validation	0.0001	0.9470	0.0103

ตาราง 13 (ต่อ)

Category	Dataset	MSE	R-square	MAE
Average	Train	0.0004	0.9020	0.0168
	Validation	0.0002	0.9342	0.0138
Merge	Train	0.0001	0.9916	0.0097
	Validation	0.0003	0.9869	0.0072

จากตาราง 12 ผลการทดลองของแบบจำลองในการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง โดยแบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าในภาพรวมแบบจำลอง ResNet50 ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0001 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.0003
- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.9916 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.9869 ซึ่งแสดงให้เห็นว่าแบบจำลองสามารถอธิบายข้อมูลได้ดีและมีความเชื่อถือได้ในการทำนายข้อมูลจริง
- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0097 และ ในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.0072

ตาราง 14 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของ แบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน

Category	Dataset	MSE	R-square	MAE
animal	Train	0.0014	0.8262	0.0275
	Validation	0.0001	0.9760	0.0081

ตาราง 15 (ต่อ)

Category	Dataset	MSE	R-square	MAE
food	Train	0.0001	0.9771	0.0101
	Validation	0.0001	0.9915	0.0059
landscape	Train	0.0002	0.9909	0.0087
	Validation	0.0064	0.4560	0.0401
sport	Train	0.0006	0.9218	0.0195
	Validation	0.0001	0.9643	0.0096
vehicle	Train	0.0004	0.9188	0.0192
	Validation	0.0005	0.7949	0.0125
Average	**Train**	**0.0005**	**0.9270**	**0.0170**
	Validation	**0.0014**	**0.8365**	**0.0153**
Merge	**Train**	**0.0003**	**0.9332**	**0.0143**
	Validation	**0.0028**	**0.9191**	**0.0476**

จากตาราง 13 ผลการทดลองของแบบจำลองในการทำนายจากชุดข้อมูลฝึกสอน และชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง โดยแบ่งเป็นการทำนายจากชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน ฝึกได้ดีกับชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- MSE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0003 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองเท่ากับ 0.0028
- R-square ในชุดข้อมูลฝึกสอน เท่ากับ 0.9332 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.9191 ซึ่งแสดงให้เห็นว่าแบบจำลองสามารถอธิบายข้อมูลได้ดีและมีความเชื่อถือได้ในการทำนายข้อมูลจริง

- MAE ในชุดข้อมูลฝึกสอน เท่ากับ 0.0143 และในชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลอง เท่ากับ 0.0476

ผลการทดลองการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลทดสอบของแบบจำลอง Vision Transformer, แบบจำลอง ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน โดยได้วัดประสิทธิภาพต่าง ๆ แสดงดังตาราง 14, 15 และ 16 ตามลำดับ

ตาราง 16 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลทดสอบของแบบจำลอง Vision Transformer

Category	MSE	R-square	MAE	Spearman's
animal	0.0001	0.9560	0.0097	0.9850
food	0.0002	0.9200	0.0095	0.9580
landscape	0.0010	0.9780	0.0100	0.9350
sport	0.0001	0.9832	0.0095	0.9850
vehicle	0.0001	0.8080	0.0097	0.9366
Average	**0.0003**	**0.9290**	**0.0097**	**0.8519**
Merge	**0.0006**	**0.9864**	**0.0082**	**0.9517**

จากตาราง 14 การทดลองด้วยแบบจำลอง Vision Transformer โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0003 และ R-square เฉลี่ยอยู่ที่ 0.9290 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0096 และค่า Spearman's เฉลี่ยอยู่ที่ 0.8519
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0006 และ R-square เฉลี่ยอยู่ที่ 0.9864 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0082 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9517

ตาราง 17 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลทดสอบของแบบจำลอง ResNet50

Category	MSE	R-square	MAE	Spearman's
animal	0.0003	0.9074	0.0116	0.5568
food	0.0001	0.9320	0.0118	0.9968
landscape	0.0001	0.9784	0.0080	0.9922
sport	0.0001	0.9961	0.0085	0.9901
vehicle	0.0001	0.9956	0.0070	0.9407
Average	**0.0002**	**0.9619**	**0.0094**	**0.8953**
Merge	**0.0001**	**0.9947**	**0.0082**	**0.9896**

จากตาราง 15 การทดลองด้วยแบบจำลองที่ใช้ในการทำนาย พบว่าในการใช้แบบจำลอง ResNet50 โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่ และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.9619 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0094 และค่า Spearman's เฉลี่ยอยู่ที่ 0.8953
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0001 และ R-square เฉลี่ยอยู่ที่ 0.9947 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0082 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9896

ตาราง 18 แสดงผลการทดลองฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ในชุดข้อมูลทดสอบของแบบจำลองด้วยการรวมเวกเตอร์คุณลักษณะจากทั้ง 2 แบบจำลอง

Category	MSE	R-square	MAE	Spearman's
animal	0.0002	0.9690	0.0133	0.8024
food	0.0001	0.8769	0.0073	0.8288
landscape	0.0005	0.9403	0.0212	0.7364
sport	0.0001	0.9448	0.0099	0.8928
vehicle	0.0002	0.9785	0.0097	0.8214
Average	**0.0002**	**0.9419**	**0.0123**	**0.8163**
Merge	**0.0002**	**0.9279**	**0.0105**	**0.9523**

จากตาราง 16 การทดลองด้วยแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน โดยแบ่งการทำนายเป็นชุดข้อมูลแบบเฉพาะหมวดหมู่และชุดข้อมูลแบบคละหมวดหมู่ พบว่าแบบจำลองทำงานได้ดีในชุดข้อมูลแบบชุดข้อมูลแบบคละหมวดหมู่ โดยมีรายละเอียด ดังนี้

- ชุดข้อมูลแบบเฉพาะหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.9419 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0123และค่า Spearman's เฉลี่ยอยู่ที่ 0.8163
- ชุดข้อมูลแบบคละหมวดหมู่ ค่า MSE เฉลี่ยอยู่ที่ 0.0002 และ R-square เฉลี่ยอยู่ที่ 0.9279 ในขณะที่ค่า MAE เฉลี่ยอยู่ที่ 0.0105 และค่า Spearman's เฉลี่ยอยู่ที่ 0.9523

บทที่ 5
อภิปรายผล สรุปผลการวิจัย และข้อเสนอแนะ

การทำนายการจดจำภาพสามารถนำไปพัฒนาต่อยอดในงานด้านต่าง ๆ ไม่ว่าจะเป็นด้านการศึกษา การตลาดและโฆษณา เนื่องจากปัจจุบันด้านการตลาดมีการเติบโตอย่างต่อเนื่องในทุกปี ประกอบกับมีการแข่งขันทางธุรกิจออนไลน์สูงขึ้น จึงเป็นที่มาของงานวิจัยนี้ และได้ศึกษาการเรียนรู้เชิงลึก 3 รูปแบบ เพื่อสร้างแบบจำลองในทำนายการจดจำภาพ อีกทั้งเพื่อเปรียบเทียบประสิทธิภาพจากแบบจำลอง ที่สร้าง และสรุปผลการวิจัย โดยสามารถแบ่งหัวข้อได้ ดังนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้เป็นการทดลองการทำนายการจดจำภาพ จากคะแนนการจดจำซึ่งมีค่าอยู่ระหว่าง 0-1 โดยทำการทดลองใช้แบบจำลองทั้งหมด 3 แบบ ได้แก่ Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน ซึ่งการทดลองแบ่งออกเป็น 3 รูปแบบ ดังนี้ 1) การฝึกแบบจำลองจากแรกเริ่ม 2) การนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ 3) การนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม

จากผลการดำเนินการวิจัย พบว่าแบบจำลอง ResNet50 โดยใช้การฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม มีประสิทธิภาพดีที่สุดเมื่อเทียบกับแบบจำลองอื่น ๆ ซึ่งการฝึกแบบจำลองจากแรกเริ่ม มีประสิทธิภาพต่ำที่สุด และสามารถสรุปผลได้ดังนี้

1. การฝึกโดยนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ มีประสิทธิภาพดีกว่าการฝึกแบบจำลองจากแรกเริ่ม และการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม

2. ประสิทธิภาพของแบบจำลองการวัดผลออกมาในทิศทางเดียวกันทั้งในการฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ และการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม

3. จากสมมติฐานว่าการจดจำภาพได้ในแต่ละบุคคลมีความสอดคล้องกัน ในแต่ละบุคคลโดยการจดจำภาพนั้น พบว่าแบบจำลอง Vision Transformer สามารถทำนายการจดจำได้ดีเทียบเท่าแบบจำลองโครงข่ายแบบสังวัฒนาการ ทั้งนี้การเลือกใช้ความเหมาะสมของข้อมูล, ชนิดของข้อมูล และจำนวนของข้อมูล อีกทั้งจากการวิจัยพบว่าการจดจำภาพนั้นเป็นส่วนหนึ่งของประสบการณ์ที่แต่ละบุคคลพบเจอของวัตถุหรือ สามารถจดจำองค์ประกอบที่อยู่ภายในภาพได้ (Dubey et al., 2015)

4. สำหรับชุดข้อมูลทดสอบ พบว่าแบบจำลอง Vision Transformer, ResNet50 และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกันสามารถทำงานได้ดีกับชุดข้อมูลแบบคละหมวดหมู่

5.2 อภิปรายผลการวิจัย

ในงานวิจัยนี้มุ่งเน้นการทำนายคะแนนการจดจำภาพ ซึ่งมีการวิเคราะห์คุณสมบัติของภาพเพื่อนำไปสู่การจดจำภาพ (Isola et al., 2014) ได้ศึกษาการวัดผลการจดจำด้วยวิธีเลือกภาพที่จดจำได้จากผู้สังเกตการณ์ จึงนิยามได้ว่า “ความทรงจำ” โดยการวิเคราะห์แต่ละภาพจากผู้สังเกตการณ์ว่าสามารถตรวจจับรูปซ้ำได้อย่างถูกต้องหรือไม่ ซึ่งการวิจัยการทำนายการจดจำภาพมีผลจากการทดลองการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ และการนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติมของแบบจำลองทั้ง 3 แบบ ดังตารางที่ 17 และ 18

ตาราง 17 แสดงผลการทดลองการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะของแบบจำลองทั้ง 3 แบบ

Category	MSE	R-square	MAE	Spearman's
Vision Transformer				
Average 5 Categories	0.0003 (0.0012)	0.8487 (0.9132)	0.0159 (0.0216)	0.6422
Merge All Categories	**0.0003 (0.0037)**	**0.9870 (0.8310)**	**0.0076 (0.0500)**	**0.9716**
ResNet50				
Average 5 Categories	0.0003 (0.0002)	0.9522 (0.9301)	0.0156 (0.0114)	0.8162
Merge All Categories	**0.0002 (0.0001)**	**0.9945 (0.9919)**	**0.0065 (0.0098)**	**0.9839**
Vision Transformer + ResNet50				
Average 5 Categories	0.0002 (0.0008)	0.95264 (0.9080)	0.0091 (0.0235)	0.9088
Merge All Categories	**0.0001 (0.0005)**	**0.936 (0.8570)**	**0.0101 (0.0171)**	**0.9761**

หมายเหตุ ( ) เป็นการรายงานประสิทธิภาพของแบบจำลองบนชุดข้อมูลสำหรับฝึก

ตาราง 18 แสดงผลการทดลองการนำฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติมของแบบจำลองทั้ง 3 แบบ

Category	MSE	R-square	MAE	Spearman's
Vision Transformer				
Average 5 Categories	0.0003 (0.0004)	0.92904 (0.9457)	0.00968 (0.0138)	0.8519
Merge All Categories	**0.0006 (0.0001)**	**0.9864 (0.9930)**	**0.0082 (0.0093)**	**0.9517**
ResNet50				
Average 5 Categories	0.0002 (0.0004)	0.9619 (0.9020)	0.0094 (0.0168)	0.8953
Merge All Categories	**0.0001 (0.0001)**	**0.9947 (0.9916)**	**0.0082 (0.0097)**	**0.9896**
Vision Transformer + ResNet50				
Average 5 Categories	0.0002 (0.0005)	0.9419 (0.9270)	0.0123 (0.0170)	0.8163
Merge All Categories	**0.0002 (0.0003)**	**0.9279 (0.9332)**	**0.0105 (0.0143)**	**0.9523**

หมายเหตุ ( ) เป็นการรายงานประสิทธิภาพของแบบจำลองบนชุดข้อมูลสำหรับฝึก

จากตาราง 17 และ 18 แสดงเปรียบเทียบค่า MSE, R-square, MAE และ Spearman's ระหว่างแบบจำลอง Vision Transformer, ResNet50, และแบบจำลองที่รวมเวกเตอร์คุณลักษณะทั้งสอง Vision Transformer และ ResNet50 เข้าด้วยกัน แสดงให้เห็นว่าประสิทธิภาพของแบบจำลองในชุดข้อมูลทดสอบสูงกว่าชุดข้อมูลแบบฝึก โดยในงานวิจัยนี้ผลการทดลองประสิทธิภาพของชุดข้อมูลทดสอบสูงกว่าชุดข้อมูลแบบฝึก มีดังนี้

1. การทดลองแบบจำลอง Vision Transformer ในชุดข้อมูลแบบแยกหมวดหมู่ ได้แก่ค่า MSE และ MAE สำหรับในชุดข้อมูลแบบคละหมวดหมู่ ได้แก่ค่า MAE

2. การทดลองแบบจำลอง ResNet50 ในชุดข้อมูลแบบแยกหมวดหมู่และชุดข้อมูลแบบคละหมวดหมู่ ได้แก่ ค่าของ MSE, R-Square และ MAE

3. การเชื่อมต่อกันระหว่างสองแบบจำลองในชุดข้อมูลแบบแยกหมวดหมู่ ได้แก่ ค่า MSE, R-Square และ MAE สำหรับในชุดข้อมูลแบบคละหมวดหมู่ ได้แก่ ค่า MSE และ MAE

เนื่องจากผลลัพธ์ที่ได้ในงานวิจัยชุดข้อมูลฝึกและชุดข้อมูลทดสอบนั้นไม่แตกต่างกันโดยสิ้นเชิง งานวิจัยนี้จึงไม่ได้ทำการแยกชุดข้อมูลในการฝึกใหม่ เนื่องจากกรณีที่ผลลัพธ์ในชุดข้อมูลทดสอบสูงกว่าชุดข้อมูลฝึกนั้นสามารถเกิดขึ้นได้ อย่างไรก็ดีควรมีการทดสอบเพิ่มเติมโดยการสุ่มเลือกตัวอย่างสำหรับชุดข้อมูลฝึกและชุดข้อมูลทดสอบ เพื่อเพิ่มความน่าเชื่อถือยิ่งขึ้นของผลการทดลอง ซึ่งแบบจำลองทั้ง 3 แบบ สามารถทำงานได้ดีในชุดข้อมูลคละหมวดหมู่ ประสิทธิภาพของแบบจำลองที่ใช้ในการทำนายคะแนนการจดจำนั้น มีความใกล้เคียงกันระหว่างแบบจำลอง โดยแบบจำลองที่มีประสิทธิภาพสูงสุด คือ ResNet50 โดยการฝึกในรูปแบบการนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ เนื่องจากเปรียบเทียบในภาพรวมจากการฝึกแบบจำลอง เช่น ค่า MSE ของแบบจำลอง Vision Transformer จากการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ดีกว่าการฝึกที่มาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะแต่เมื่อนำมาทดสอบกลับมีประสิทธิภาพที่แย่ลง สาเหตุอาจเกิดจากการ Fine-Tuning ที่ทำให้เกิดการปัญหา Overfitting

ทั้งนี้ในงานการจดจำภาพนี้ส่วนใหญ่ใช้เทคนิคการเรียนรู้เชิงลึก ดังนั้นในงานวิจัยนี้จึงใช้เทคนิคการเรียนรู้เชิงลึกเพื่อให้ได้ผลลัพธ์จากชุดข้อมูลทดสอบแล้วเปรียบเทียบกับงานวิจัยอื่น ๆ ในลักษณะที่ใกล้เคียงกันและเกี่ยวข้องกับการจดจำภาพ ดังตาราง 19

ตาราง 19 แสดงสรุปผลแบบจำลองและเปรียบเทียบกับงานวิจัยอื่น

Model	Dataset	MSE	Spearman's
ResMem (Arockia Praveen et al., 2021)	LaMem	0.009	0.67
ViTMem Final Model (Thomas & Thomas, 2023)	LaMem+MemCat	0.006	0.77
Vision Transformer	MemCat	0.0006	0.9517
ResNet50	MemCat	0.0001	0.9896
ResNet50+Vision Transformer	MemCat	0.0002	0.9523

จากตาราง 19 แสดงประสิทธิภาพของแบบจำลองที่ใช้ในการทำนายคะแนนการจดจำนั้น พบว่ามีความใกล้เคียงกันระหว่างแบบจำลอง ซึ่งในงานวิจัยนี้แบบจำลองที่มีประสิทธิภาพสูงสุด นั่นคือ ResNet50 เนื่องจากแบบจำลอง Vision Transformer มีความซับซ้อนของแบบจำลองสูง ขึ้นอยู่กับการเลือกใช้ความเหมาะสมของข้อมูล, ชนิดของข้อมูล และจำนวนของข้อมูลในการฝึก

ในงานวิจัยนี้มีประสิทธิภาพของแบบจำลองสูงกว่างานวิจัยอื่น ๆ ซึ่งอาจเกิดได้จากการเว้นขอบภาพ (Padding) หลังจากการนำเข้าแบบจำลองเรียบร้อยแล้ว ทำให้อัตราส่วนของภาพยังคงเดิม แต่งานวิจัยอื่น ๆ ไม่ได้มีการเว้นขอบภาพ (Padding) อีกทั้งยังมีการเพิ่มข้อมูล (Data Augmentation) เช่น การหมุน, การตัดภาพ หรือ การซูม เป็นต้น ทำให้ผลลัพธ์ค่าคะแนนการจดจำภาพในงายวิจัยอื่นต่ำกว่างานวิจัยนี้

5.3 ข้อเสนอแนะ

1. ในงานวิจัยนี้มีการใช้ชุดข้อมูลแบบเฉพาะหมวดหมู่ หากใช้ข้อมูลที่มีจำนวนมากขึ้นอาจทำให้แบบจำลองสามารถทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น เนื่องจากแบบจำลองค่อนข้างมีความซับซ้อน และได้ผลดีกับชุดข้อมูลที่มีขนาดใหญ่มาก ๆ

2. สามารถทำเทคนิคอื่น ๆ เพิ่มเติม เช่น การตรวจจับวัตถุอื่น ๆ ที่มีลักษณะเด่นรองลงมา เพื่อหานำมาคำนวณค่าความจดจำได้แม่นยำยิ่งขึ้นเนื่องจากในชุดข้อมูลมีการติดป้ายกำกับเฉพาะวัตถุที่มีลักษณะเด่นในภาพ

3. สามารถทำงานได้ดีทั้งในการฝึกโดยนำแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาใช้เป็นแบบจำลองเพื่อการคำนวณเวกเตอร์คุณลักษณะ และการฝึกแบบจำลองที่ฝึกมาจากชุดข้อมูลอื่นมาปรับแต่งเพิ่มเติม ด้วยชุดข้อมูลแบบคละหมวดหมู่

บรรณานุกรม

Alex Krizhevsky, I. S., Geoffrey E. Hinton. (2012). ImageNet Classification with Deep Convolutional Neural Networks. 25. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (Curran Associates, Inc.)

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, & Houlsby, N. (2021). *An Image is Worth* 16*x*16 *Words: Transformers for Image Recognition at Scale* The International Conference on Learning Representations (ICLR) 2021,

Arockia Praveen, Abdulfattah Noorwali, Duraimurugan Samiayya, Mohammad Zubair Khan, Durai Raj Vincent P M, Ali Kashif Bashir, & Alagupandi, V. (2021). ResMem-Net: memory based deep CNN for image memorability estimation. *PeerJ Computer Science*. https://doi.org/https://doi.org/10.7717/peerj-cs.767

Bang, J. H., Park, S. W., Kim, J. Y., Park, J., Huh, J. H., Jung, S. H., & Sim, C. B. (2023). CA-CMT: Coordinate Attention for Optimizing CMT Networks. *IEEE Access*, 11, 76691-76702. https://doi.org/10.1109/ACCESS.2023.3297206

Deep Learning & Neural Networks. (2019). Subbrain. https://www.sub-brain.com/uncategorized/deep-learning-neural-networks/

Dubey, R., Peterson, J., Khosla, A., Yang, M. H., & Ghanem, B. (2015, 7-13 Dec. 2015). What Makes an Object Memorable? 2015 IEEE International Conference on Computer Vision (ICCV),

Giannopoulos, M., Aidini, A., Pentari, A., Fotiadou, K., & Tsakalides, P. (2020). Classification of Compressed Remote Sensing Multispectral Images via Convolutional Neural Networks. *Journal of Imaging*, 6(4), 24. https://www.mdpi.com/2313-433X/6/4/24

Goetschalckx, L., & Wagemans, J. (2019). MemCat: a new category-based image set quantified on memorability. *PeerJ*, 7, e8169. https://doi.org/10.7717/peerj.8169

He, K., Zhang, X., Ren, S., & Sun, J. (2016, 27-30 June 2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What Makes a Photograph Memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469-1482. https://doi.org/10.1109/TPAMI.2013.200

Kaiming He, X. Z., Shaoqing Ren, Jian Sun. (2016). <Deep Residual Learning for Image Recognition .pdf>. *IEEE Conference*. https://doi.org/10.1109/CVPR.2016.90

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015, 7-13 Dec. 2015). Understanding and Predicting Image Memorability at a Large Scale. 2015 IEEE International Conference on Computer Vision (ICCV),

Lahrache, S., & Ouazzani, R. E. (2022, 3-4 March 2022). A Survey on Image Memorability Prediction: From Traditional to Deep Learning Models. 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET),

LeCun Y, B. Y., Hinton G. (2015). Deep learning. 436–444. https://doi.org/10.1038/nature14539. PMID: 26017442.

Mosavi, A., Ardabili, S., & Várkonyi-Kóczy, A. R. (2020). List of Deep Learning Models. In *Engineering for Sustainable Future* (202-214). https://doi.org/10.1007/978-3-030-36841-8_20

Nash, K. O. S. a. R. (2015). An Introduction to Convolutional Neural Networks. *ArXiv e-prints*. https://doi.org/https://doi.org/10.48550/arXiv.1511.08458

Thomas, H., & Thomas, E. (2023). Image Memorability Prediction with Vision Transformers. *ArXiv*, 1, 1-6.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems,

wisdomml. (2023). *Understanding ResNet-50 in Depth: Architecture, Skip Connections, and Advantages Over Other Networks*. https://wisdomml.in/understanding-resnet-

50-in-depth-architecture-skip-connections-and-advantages-over-other-networks/

Yosinski, J. e. a. (2014). How transferable are features in deep neural networks? *ArXiv*.

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018, 18-23 June 2018). Learning Transferable Architectures for Scalable Image Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),

เซลล์ประสาท. (2023). วิกิพีเดีย สารานุกรมเสรี. https://th.wikipedia.org/w/index.php?title=&oldid=10714232

ประวัติผู้เขียน