

การจำแนกอาคารจากภาพถ่ายดาวเทียมโดยการถ่ายโอนความรู้

ปิยะเนตร์ ต่วนพะเอม¹ เอกรัฐ รัชกาญจน์^{2*}

^{1,2*} คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

*ผู้ประพันธ์บรรณกิจ อีเมล : ekarat@as.nida.ac.th

รับต้นฉบับ : 27 ตุลาคม 2564; รับบทความฉบับแก้ไข : 6 ธันวาคม 2564; ตอบรับบทความ : 2 มีนาคม 2565

เผยแพร่ออนไลน์ : 27 มิถุนายน 2565

บทคัดย่อ

เทคโนโลยีภาพถ่ายดาวเทียมมีความสำคัญในงานทางด้านวิเคราะห์พื้นที่ทางภูมิศาสตร์ เช่น Change Detection Real Estate Analysis เนื่องจากภาพมีรายละเอียดสูงและถูกจัดเก็บไว้อย่างครบถ้วน แต่เนื่องจากผู้เชี่ยวชาญทางด้านการวิเคราะห์พื้นที่ทางภูมิศาสตร์มีจำนวนน้อยเมื่อเทียบกับพื้นที่ทางภูมิศาสตร์ของประเทศไทย ทำให้การวิเคราะห์นั้นถูกพัฒนาเป็นระบบอัตโนมัติที่มีประสิทธิภาพสูง ในงานนี้ผู้วิจัยได้สร้างระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียม (Buildings Classification from Satellite Images) โดยการสร้างตัวแบบการถ่ายโอนความรู้ (Transfer Learning) ด้วยชุดข้อมูลที่สร้างขึ้นใหม่ 4CateSAT ที่ประกอบด้วยภาพอาคารสถานที่ คือ (1) สนามบิน (2) สนามกีฬา (สนามฟุตบอล) (3) โรงเรียน และ (4) วัดในประเทศไทย ตลอดจนเปรียบเทียบประสิทธิภาพของระบบหลังจากดำเนินการแก้ปัญหา Imbalance Dataset โดยการปรับแต่งข้อมูล (Data Augmentation) ทั้ง 2 แบบ ซึ่งผลการทดลองแสดงค่าความแม่นยำ (Accuracy) ของตัวแบบที่ดีที่สุดเมื่อทดสอบในชุดข้อมูลทดสอบ (Test Dataset) สูงถึง 96.88%

คำสำคัญ : ภาพถ่ายดาวเทียม สโมท การเรียนรู้เชิงลึก การจำแนกประเภทข้อมูลภาพถ่าย การถ่ายโอนความรู้

Buildings Classification from Satellite Images by Transfer Learning

Piyanate Touncha-em¹ Ekarat Rattagan^{2*}

^{1,2*}*Graduate School of Applied Statistics, National Institute of Development Administration, Bangkok, Thailand*

*Corresponding Author. E-mail address: ekarat@as.nida.ac.th

Received: 27 October 2021; Revised: 6 December 2021; Accepted: 2 March 2022

Published online: 27 June 2022

Abstract

Satellite imaging technology is essential for various applications such as real estate analysis, disaster monitoring, and more. However, as satellite data is big data, it is time-consuming and challenging for humans to analyze it, even a simple task such as detecting the types of buildings in a large area. In this paper, we develop a satellite imaging-analytics technique by applying the transfer learning algorithm to learn and classify different types of buildings in Thailand. The proposed model is learnt and tested on our created datasets, namely 4CateSAT, including the images of buildings including (1) airports, (2) stadiums (football fields), (3) schools, and (4) temples in Thailand. We also apply well-known algorithms to handle the imbalanced data, and the experimental results show that the accuracy of the best model is 96.88%.

Keywords: Satellite imagery, SMOTE, Deep learning, Image classification, Transfer learning

1) บทนำ

ปัจจุบันเทคโนโลยีมีความก้าวหน้าและพัฒนาอย่างรวดเร็ว ส่งผลให้การวิเคราะห์และใช้ประโยชน์จากสิ่งต่าง ๆ หันมาพึ่งเทคโนโลยีมากขึ้น สำหรับการวิเคราะห์พื้นที่ทางภูมิศาสตร์ก็ใช้ประโยชน์จากความก้าวหน้าตรงนี้เช่นกัน เช่น การจำแนกประเภทอาคาร การตรวจจับการใช้ที่ดิน การปรับปรุงแผนที่ทางภูมิศาสตร์

เนื่องจากนักภูมิสารสนเทศของประเทศไทย เช่น เจ้าหน้าที่ของกรมแผนที่ดิน เจ้าหน้าที่ของสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ มีจำนวนน้อยเมื่อเทียบกับพื้นที่ทางภูมิศาสตร์ของประเทศไทยที่มีขนาดใหญ่และการประยุกต์ใช้ประโยชน์จากภาพถ่ายดาวเทียมทางภูมิศาสตร์ที่ไม่รู้จบ ทำให้เทคโนโลยีในด้านการวิเคราะห์พื้นที่ทางภูมิศาสตร์ เกิดการพัฒนาอย่างต่อเนื่อง ประกอบกับซอฟต์แวร์คอมพิวเตอร์ที่เรียกว่าการเรียนรู้เชิงลึก (Deep Learning) ก็มีบทบาทมาก ดังนั้นการวิเคราะห์ภาพถ่ายดาวเทียมทางภูมิศาสตร์จึงถูกพัฒนาให้เป็นระบบอัตโนมัติที่มีประสิทธิภาพสูง แต่ใช้เวลาและบุคลากรในการวิเคราะห์น้อย

แต่ประสิทธิภาพของระบบอัตโนมัติ นั้น ขึ้นอยู่กับชุดข้อมูลที่มีความเหมาะสม ดังนั้นควรเลือกใช้หรือสร้างชุดข้อมูลใหม่อย่างระมัดระวัง และตรวจสอบอย่างถี่ถ้วน [1] จึงเป็นเหตุผลให้ทางผู้วิจัยสร้างชุดข้อมูลใหม่ที่เรียกว่า 4CateSAT เนื่องจากภาพถ่ายดาวเทียมของประเทศไทยในหมวดหมู่ที่ต้องการศึกษา ยังไม่มีและภาพถ่ายดาวเทียมบางประเภทไม่อนุญาตให้เข้าถึงหรือเข้าถึงได้แต่เสียค่าบริการ

ในงานวิจัยชิ้นนี้ ผู้วิจัยได้สร้างระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียม (Buildings Classification from Satellite Images) โดยจัดเป็นหมวดหมู่ดังนี้คือ สนามบิน สนามกีฬา (สนามฟุตบอล) โรงเรียน วัดในประเทศไทย ซึ่งชุดข้อมูลที่ใช้เป็นชุดข้อมูล 4CateSAT ซึ่งผู้วิจัยได้ทำการสร้างขึ้นใหม่ โดยบันทึกภาพถ่ายดาวเทียมความละเอียดสูงจากซอฟต์แวร์กูเกิล เอิร์ธโพร (Google Earth Pro) ที่อนุญาตให้เข้าถึงได้และไม่เสียค่าบริการ ตามหมวดหมู่ที่ต้องการศึกษา แต่เนื่องจากจำนวนข้อมูลในแต่ละประเภทยังมีขนาดไม่เท่ากัน (Imbalance Dataset) ทำให้ต้องแก้ปัญหาดังกล่าวโดยการเพิ่มจำนวนข้อมูล ซึ่งทางผู้วิจัยได้แบ่งวิธีการเพิ่มจำนวนข้อมูลออกเป็น 2 การทดลองดังนี้ (1) การปรับแต่งข้อมูล (Data Augmentation) โดยสร้างจากไลบรารี TensorFlow (2) การปรับแต่งข้อมูล (Data

Augmentation) โดยสร้างจากแพ็คเกจ Imblearn ที่ใช้สมิโธ (SMOTE : Synthetic Minority Over-sampling Technique) และดำเนินการสร้างระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียม (Buildings Classification from Satellite Images) โดยการสร้างตัวแบบการถ่ายโอนความรู้ (Transfer Learning) และเปรียบเทียบประสิทธิภาพของระบบในแต่ละการทดลอง ตลอดจนแสดงการสร้างภาพนิทัศน์จากข้อมูล (Data Visualization) ว่าตัวแบบที่ดีที่สุดจำแนกภาพว่าเป็นอาคารประเภทนั้น ๆ จากองค์ประกอบใดของภาพด้วย Grad-CAM (Gradient-weighted CAM)

2) วัตถุประสงค์ในการวิจัย

1. สร้างชุดข้อมูลภาพถ่ายดาวเทียมใหม่ 4CateSAT ให้เหมาะสมกับงาน
2. สร้างระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียมในประเทศไทยที่มีประสิทธิภาพ โดยจำแนกออกเป็น 4 หมวดหมู่ คือ สนามบิน สนามกีฬา (สนามฟุตบอล) โรงเรียน วัด ซึ่งอาจประยุกต์ใช้ระบบในเรื่องการตรวจจับการใช้ที่ดิน ปรับปรุงแผนที่ทางภูมิศาสตร์ เป็นต้น

3) เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ได้ทบทวนแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียม โดยมีเนื้อหาเกี่ยวข้องกับการกำหนดค่าจำกัดความในด้านต่าง ๆ รวมถึงชี้ให้เห็นถึงความสำคัญของการเรียนรู้เชิงลึกต่อการวิเคราะห์ภาพ การแก้ปัญหาจำนวนข้อมูลในแต่ละประเภทยังมีขนาดไม่เท่ากัน การสร้างภาพนิทัศน์จากข้อมูลเพื่อดูประสิทธิภาพของตัวแบบ ตลอดจนทบทวนการจำแนกประเภทข้อมูลภาพถ่ายดาวเทียม (Satellite Image Classification) ในวิธีการต่าง ๆ ซึ่งผู้วิจัยได้รวบรวมข้อมูลและทฤษฎีที่เป็นประโยชน์ ต่อการวิจัยครั้งนี้ สามารถแบ่งได้ดังนี้

3.1) ความสำคัญของการเรียนรู้เชิงลึก (Deep Learning) ต่อการวิเคราะห์ภาพ

การเรียนรู้เชิงลึกเป็นศาสตร์ย่อยของการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งเป็นการจำลองมาจากโครงสร้างของระบบประสาทในสิ่งมีชีวิต โดยการเรียนรู้ที่มีประสิทธิภาพเกิดจากการสร้างโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network : CNN) ซึ่งเป็นตัวแบบ

โครงข่ายประสาท (Neural Network) ขนาดใหญ่ที่มีหลายชั้น (Layers) โดยแต่ละชั้นจะเกิดการเรียนรู้ที่แตกต่างกัน และจะสกัดค่าคุณลักษณะ (Feature Extraction) อัตโนมัติ ซึ่งโครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการแสดงดังรูปที่ 1 ซึ่งสามารถพัฒนาให้เป็นตัวแบบที่มีประสิทธิภาพและนำมาใช้ในงานถ่ายโอนความรู้ (Transfer Learning) โดยมีองค์ประกอบดังนี้

- (1) Input Layer : รูปภาพสี RGB มีขนาดเป็นกว้าง*ยาว*ลึก
- (2) Convolutional Layer : อาศัยหลักการ Convolution คือ ผลคูณสเกลาร์ (Dot Product) ของเคอร์เนลหรือตัวกรอง (Filter) หรือน้ำหนัก (Weight) กับพื้นที่ส่วนย่อยในรูปภาพ เพื่อให้ได้ฟังก์ชันคุณลักษณะ (Feature Map)
- (3) Pooling Layer : การลดขนาดฟังก์ชันคุณลักษณะ โดยจะเลือกบางค่าออกมาจากในแต่ละพื้นที่ส่วนย่อย โดยการคำนวณที่นิยมมี 2 แบบ คือ Max Pooling (เลือกค่าที่สูงที่สุดจากพื้นที่ส่วนย่อยนั้น) และ Average Pooling (เลือกค่าเฉลี่ยของพื้นที่ส่วนย่อยนั้น)
- (4) Fully Connected Layer : ทำหน้าที่ในการเรียนรู้และจำแนกประเภทข้อมูลภาพ
- (5) Output Layer : แสดงประเภทข้อมูลภาพที่ได้จากการเรียนรู้

ในระหว่างการฝึก (Train) นั้น นิวรอนจะปรับปรุงค่าน้ำหนักโดยใช้อัลกอริทึมทางคณิตศาสตร์ที่เรียกว่าการหาค่าเหมาะสม (Optimization) เพื่อให้ตัวแบบเกิดการเรียนรู้มากที่สุดซึ่งหมายความว่าค่าปรับจะต่ำกว่าค่าผิดพลาด (Error) จะน้อยที่สุดหรือยอมรับได้

ทำให้เกิดตัวแบบที่มีประสิทธิภาพ จึงเกิดการถ่ายโอนความรู้ (Transfer Learning) [1] จากตัวแบบที่มีประสิทธิภาพนั้นมา

ประยุกต์ใช้กับตัวแบบอีกงานหนึ่งได้ ทำให้งานใหม่มีค่าความแม่นยำ (Accuracy) ที่ดีขึ้น และประหยัดเวลาในการฝึก

3.2) การแก้ปัญหาจำนวนข้อมูลในแต่ละประเภทมีขนาดไม่เท่ากัน (Imbalance Dataset)

3.2.1) การปรับแต่งข้อมูลแบบ TensorFlow เป็นเทคนิคการเพิ่มจำนวนข้อมูล โดยคุณลักษณะหลักยังคงเดิมแต่เกิดภาพที่มีลักษณะแตกต่างกันออกไป เช่น การพลิกภาพ (Flip) การเลื่อนภาพ (Shift) การหมุนภาพ (Rotation) และยังเป็นการเพิ่มความหลากหลายให้กับชุดข้อมูล ตัวแบบก็จะเรียนรู้ได้ดีขึ้นและลดปัญหา Overfit ลง [2]

3.2.2) การปรับแต่งข้อมูลแบบสโมท เป็นเทคนิคการสุ่มเพิ่มจำนวนข้อมูลกลุ่มน้อย ให้มีจำนวนใกล้เคียงกับข้อมูลกลุ่มหลัก (Over-sampling) แต่ข้อมูลใหม่นั้นมาจากการสังเคราะห์ โดยใช้อัลกอริทึมเพื่อนบ้านที่ใกล้ที่สุด (K-nearest neighbor) และดำเนินการตามสมการที่ (1) ซ้ำหลายครั้งจนกระทั่งจำนวนข้อมูลกลุ่มน้อยมีจำนวนใกล้เคียงกับข้อมูลกลุ่มหลัก [3]

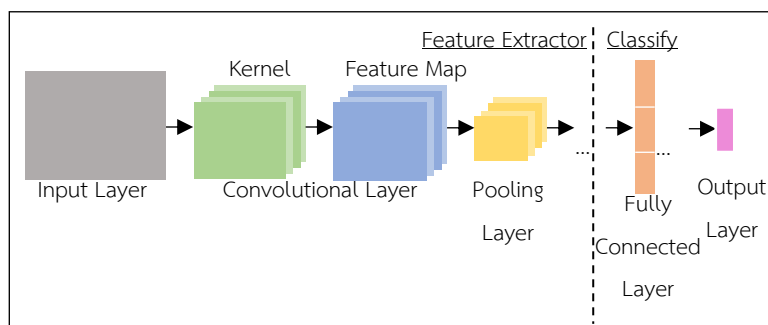
$$x_{new} = x_i + \lambda(x_{zi} - x_i) \quad (1)$$

โดยที่ x_{new} หมายถึง ข้อมูลใหม่ที่จะถูกสร้างขึ้นโดยพิจารณาจากเพื่อนบ้านใกล้เคียง k แห่ง

x_i หมายถึง ตัวอย่างข้อมูลในกลุ่มน้อยที่ถูกเลือก (จากการสุ่ม)

x_{zi} หมายถึง เพื่อนบ้านที่ใกล้ที่สุดที่ถูกเลือก

λ หมายถึง ค่า 1 ค่าที่อยู่ระหว่าง 0 ถึง 1 (จากการสุ่ม)



รูปที่ 1 : โครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network : CNN)

3.3) การสร้างภาพนิทัศน์จากข้อมูล (Data Visualization)

Grad-CAM (Gradient-weighted CAM) [4] เป็น CNN Visualization ใช้เพื่อบอกว่าตัวแบบจำแนกภาพว่าเป็นประเภทนั้น ๆ จากองค์ประกอบใดของภาพ โดยแสดงผ่านแผนภูมิความร้อน (Heatmap) ซึ่งการทำงานของ Grad-CAM คือ หาเกรเดียนต์ของค่าทำนายก่อนที่จะทำการ Normalize ใน Softmax ของหมวดหมู่นั้น ๆ กับ Feature Map Activation (ผลลัพธ์ของ Conv Layer ชั้นสุดท้าย) แต่ละแผ่น นำค่าเกรเดียนต์ที่ได้มาหา Global Average Pooling ทีละค่าจะได้ค่าแอลฟา และหาผลรวมเชิงเส้น (Linear Combination) ระหว่างค่าแอลฟาและ Feature Map Activation และผ่านฟังก์ชัน ReLU ก็จะได้ Grad-CAM Heatmap ซึ่งการทำงานของ Grad-CAM แสดงดังรูปที่ 2

3.4) การจำแนกประเภทข้อมูลภาพถ่ายดาวเทียม (Satellite Image Classification)

ในงานวิจัย [6] ได้ทำการจำแนกประเภทข้อมูลภาพถ่ายดาวเทียมโดยการเรียนรู้เชิงลึก ซึ่งใช้ชุดข้อมูล Functional Map of the World ของ IARPA ประกอบด้วยภาพถ่ายดาวเทียม 63 หมวดหมู่ (1 ล้านภาพ) มาพร้อมกับข้อมูลรายละเอียดของแต่ละภาพ

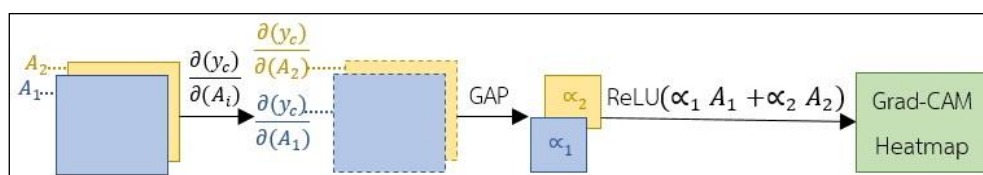
สำหรับการฝึกโครงข่ายประสาทเทียมแบบสังวัตนาการฝึก โดยการสร้างตัวแบบถ่ายโอนความรู้จากตัวแบบที่มีประสิทธิภาพ คือ VGG16 InceptionV3 ResNet50 ResNet152 DenseNet161 ทำการปรับแต่งข้อมูลเพียงการพลิกภาพ 90° 180° 270° เพราะพบว่าการเลื่อนภาพ และการหมุนภาพไม่ได้ทำให้ผลลัพธ์ดีขึ้น อีกทั้งยังได้ทดลองฝึกตัวแบบเองแต่ตัวแบบที่ได้จากการถ่ายโอนความรู้มีค่าความแม่นยำ (Accuracy) และประหยัดเวลาในการฝึกมากกว่า ตลอดจนทดลองปรับจูน (Fine-tuning) เข้าไปในชั้นที่ลึกขึ้นแต่ผลลัพธ์ก็ไม่ได้ดีขึ้น จากตัวแบบที่สร้างได้มีค่าความ

แม่นยำ (Accuracy) สูงถึง 83% ในชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset)

อีกงานวิจัยที่น่าสนใจคือการจำแนกประเภทที่ดิน (Land Use and Land Cover Classification) โดยการเรียนรู้เชิงลึก [7] ซึ่งใช้ชุดข้อมูลที่ทำกรสร้างใหม่ EuroSAT โดยเลือกข้อมูลมาจากภาพถ่ายดาวเทียม Sentinel-2 ประกอบด้วยภาพถ่ายดาวเทียม 10 หมวดหมู่ (27,000 ภาพ) โดยแบ่งเป็นชุดข้อมูลฝึก (Train Dataset) : ชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) = 80 : 20 และได้ศึกษาจนค้นพบว่าการสร้างตัวแบบที่ได้จากการถ่ายโอนความรู้ เหมาะกับชุดข้อมูลที่เป็นภาพถ่ายดาวเทียม

สำหรับการฝึกสรุปได้ว่าการใช้ภาพถ่ายดาวเทียมเป็นภาพสี RGB ดีกว่า CI และ SWIR อีกทั้งสรุปได้ยิ่งกว่าการฝึกด้วย deepCNN (GoogleNet และ ResNet50) มีประสิทธิภาพดีกว่า shallowCNN และ shallowCNN มีประสิทธิภาพดีกว่า BoVW (SIFT+SVM) จากตัวแบบที่สร้างได้มีค่าความแม่นยำ (Accuracy) สูงถึง 98.57%

ระหว่างการฝึก เพื่อปรับปรุงค่าน้ำหนักมีไฮเปอร์พารามิเตอร์ที่เป็นที่นิยมที่สุดคือ Optimizer ที่ชื่อ Adaptive Moment Estimation (Adam) เพราะจะปรับ อัตราการเรียนรู้ (Learning Rate) ให้อัตโนมัติ ลดปัญหา Decaying Learning Rate ของเกรเดียนต์ซึ่งช่วยให้ตัวแบบไม่หยุดการเรียนรู้ และลู่เข้าได้เร็วกว่า Gradient Descent อีกทั้งพารามิเตอร์ไม่แกว่ง เพราะให้ความสำคัญในเรื่องการพุ่งไปยังทิศทางที่ใกล้จุดกลางมากที่สุดก่อน [8] และ ไฮเปอร์พารามิเตอร์อีกตัวที่น่าสนใจคือ kernel_initializer (วิธีการกำหนดน้ำหนักกลุ่มเริ่มต้นของชั้น Keras) ที่ชื่อ HeNormal เนื่องจากมีความเหมาะสมหากฟังก์ชันกระตุ้น (Activation Function) เป็น ReLU เพราะจะช่วยให้ตัวแบบไม่หยุดการเรียนรู้และเรียนรู้ได้รวดเร็วมากขึ้น [9]

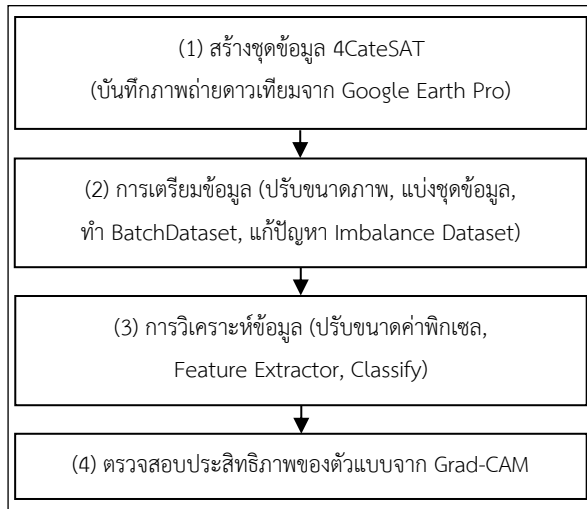


รูปที่ 2 : การทำงานของ Grad-CAM (Gradient-weighted CAM) [5]

4) วิธีดำเนินการวิจัย

ในงานวิจัยชิ้นนี้ ผู้วิจัยได้กำหนดวิธีดำเนินการวิจัยไว้ 4 ส่วน แสดงดังต่อไปนี้

4.1) แผนภาพการดำเนินงาน



รูปที่ 3 : แผนภาพการดำเนินงาน

4.2) ชุดข้อมูล (Dataset)

รูปที่ 3 ข้อ (1) แสดงให้เห็นการสร้างชุดข้อมูล 4CateSAT ซึ่งผู้วิจัยได้ทำการสร้างขึ้นใหม่ โดยบันทึกภาพถ่ายดาวเทียมความละเอียดสูงจากซอฟต์แวร์กูเกิล เอิร์ธโพร (Google Earth Pro) เป็นเครื่องมือที่มีประสิทธิภาพสำหรับการดูข้อมูลทางภูมิศาสตร์ โดยมีแหล่งที่มาจากบริษัทดาวเทียมหลายแห่ง (เช่น U.S. public domain, คีย์โฮล, กูเกิลแมพ, ซิตี้เซตส์ (CitySets)) ซึ่งภาพเหล่านี้ถูกรวบรวมกันเป็นภาพโมเสกที่ถ่ายจากหลายระยะเวลา ซึ่งชุดข้อมูลชุดนี้ประกอบด้วย ภาพถ่ายดาวเทียมในหมวดหมู่สนามบิน สนามกีฬา (สนามฟุตบอล) โรงเรียน วัด ซึ่งอยู่ในประเทศไทย จำนวน 1,402 ภาพ โดยได้ดำเนินการเก็บข้อมูลเป็นระยะเวลาประมาณ 2 เดือน ระหว่างเดือน กุมภาพันธ์-มีนาคม 2564 โดยชุดข้อมูลภาพถ่ายทางดาวเทียมในแต่ละประเภท ผู้วิจัยจะเลือกเฉพาะภาพที่มีคุณลักษณะเด่น ๆ ตามที่ผู้วิจัยกำหนดดังแสดงในตารางที่ 1 ดังนั้นหากเป็นภาพที่ไม่มีคุณลักษณะเด่นตามที่ผู้วิจัยกำหนดอาจจำแนกประเภทคลาดเคลื่อน สำหรับตัวอย่างภาพในชุดข้อมูล 4CateSAT จะแสดงในรูปที่ 4 [10]

ตารางที่ 1 : คุณลักษณะเด่นของภาพในแต่ละหมวดหมู่

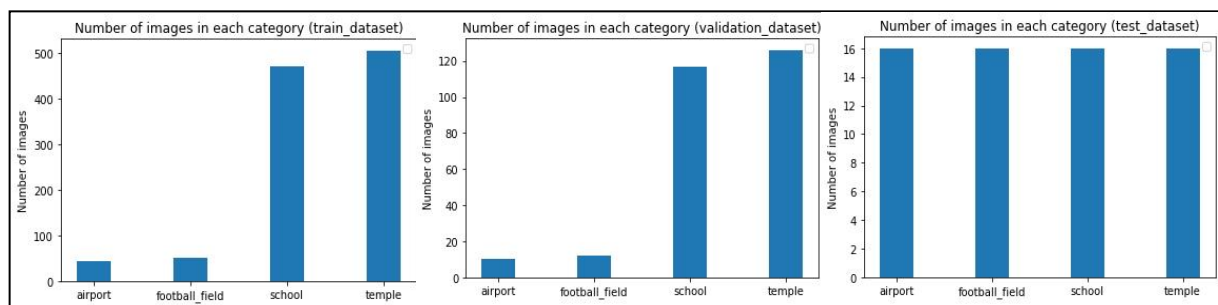
หมวดหมู่	คุณลักษณะเด่นของภาพ
สนามบิน	ลานบิน (Runway) เครื่องบิน
สนามกีฬา (สนามฟุตบอล)	สนามฟุตบอล
โรงเรียน	อาคาร สนามกีฬาต่าง ๆ
วัด	เจดีย์ หลังคาสีส้มเป็นกลุ่ม ๆ

4.3) การเตรียมข้อมูล

จากรูปที่ 3 ข้อ (2) เนื่องจากภาพถ่ายดาวเทียมในชุดข้อมูล 4CateSAT มีขนาดประมาณ 4800×2600 ผู้วิจัยได้ปรับเป็นขนาด 256×256 (ขนาดภาพที่ใหญ่ที่สุดที่โปรแกรมจะสามารถรัน SMOTE ได้) และดำเนินการแบ่งชุดข้อมูลโดยแยกชุดข้อมูลทดสอบ (Test Dataset) ออกมา 64 ภาพ แล้วจึงแบ่งเป็นชุดข้อมูลฝึก (Train Dataset) : ชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) = $80\%(1,073 \text{ ภาพ}) : 20\%(265 \text{ ภาพ})$ นำเข้าตัวแบบเป็น BatchDataset ทีละ 32 ภาพ ซึ่งจำนวนภาพในแต่ละหมวดหมู่ของชุดข้อมูลฝึก (Train Dataset) ชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) และชุดข้อมูลทดสอบ (Test Dataset) แสดงในรูปที่ 5 แต่เนื่องจากจำนวนข้อมูลในแต่ละประเภทยังมีขนาดไม่เท่ากัน (Imbalance Dataset) ทำให้ต้องแก้ปัญหาดังกล่าวโดยการเพิ่มจำนวนข้อมูลซึ่งทางผู้วิจัยได้แบ่งวิธีการเพิ่มจำนวนข้อมูลออกเป็น 2 การทดลองดังนี้ (1) การปรับแต่งข้อมูลแบบ TensorFlow เลือกใช้ Flip('horizontal') และ Rotation(0.2) จากไลบรารีสำเร็จรูป TensorFlow โดยคุณลักษณะหลักของภาพยังคงเดิมแต่เกิดภาพที่มีลักษณะแตกต่างกันออกไป ซึ่งตัวอย่างภาพหลังจากทำการปรับแต่งข้อมูลของการทดลองที่ 1 แสดงในรูปที่ 6 (ภาพซ้าย) (2) การปรับแต่งข้อมูลแบบสโมท สร้างจากแพ็คเกจสำเร็จรูป Imblearn ที่ใช้สโมท โดยหลักการ Over-sampling แต่ข้อมูลใหม่นั้นมาจากการสังเคราะห์ สำหรับตัวอย่างภาพหลังจากทำการปรับแต่งข้อมูลของการทดลองที่ 2 แสดงในรูปที่ 6 (ภาพขวา) และจำนวนภาพในแต่ละหมวดหมู่ของชุดข้อมูลฝึก (Train Dataset) และชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) หลังจากดำเนินการตามการทดลองที่ 2 แสดงในรูปที่ 7



รูปที่ 4 : ตัวอย่างภาพในชุดข้อมูล (Dataset) 4CateSAT ที่บันทึกภาพจากซอฟต์แวร์กูเกิล เอิร์ธโพร (Google Earth Pro)
แหล่งที่มา : Google Earth, Image © 2021 Maxar Technologies (ภาพบนซ้าย-ขวา, ภาพล่างขวา) [10], Google Earth, Image © 2021 CNES/Airbus (ภาพล่างซ้าย) [10]



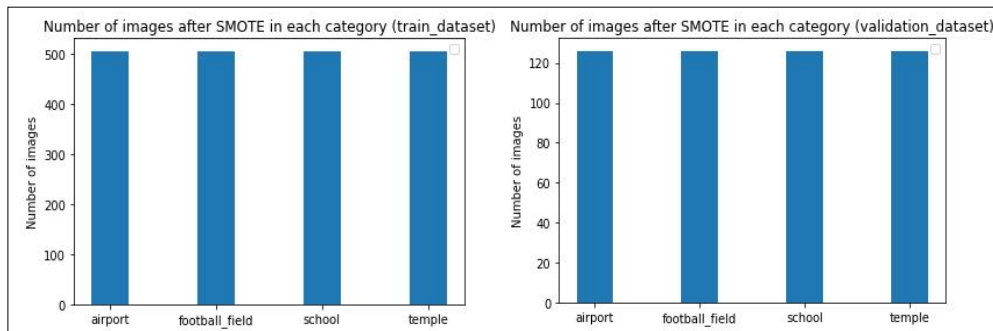
รูปที่ 5 : จำนวนภาพในแต่ละหมวดหมู่ของชุดข้อมูลฝึก (Train Dataset) ชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) และชุดข้อมูลทดสอบ (Test Dataset)



รูปที่ 6 : ตัวอย่างภาพหลังจากทำการปรับแต่งข้อมูล (Data Augmentation) ของการทดลองที่ 1 (ภาพซ้าย)

ตัวอย่างภาพหลังจากทำการปรับแต่งข้อมูล (Data Augmentation) ของการทดลองที่ 2 (ภาพขวา)

แหล่งที่มา : Google Earth, Image © 2021 CNES / Airbus (ภาพซ้าย) [10] และ Google Earth (ภาพขวา) [10]



รูปที่ 7 : จำนวนภาพในแต่ละหมวดหมู่ของชุดข้อมูลฝึก (Train Dataset) และชุดข้อมูลตรวจสอบความถูกต้อง (Validation Dataset) หลังจากดำเนินการตามการทดลองที่ 2

4.4) การวิเคราะห์ข้อมูล

ในส่วนรูปที่ 3 ข้อ (3) สร้างตัวแบบจากการถ่ายโอนความรู้ โดย Pre-Trained Model ที่ใช้คือ VGG16 [11], InceptionV3 [12], ResNet152V2 [13], DenseNet201 [14], Xception [15] ขั้นตอนการวิเคราะห์ข้อมูลมีดังนี้

4.4.1) ปรับขนาดค่าพิกเซลตามที่แต่ละ Pre-Trained Model กำหนด ([0,1] หรือ [-1,1])

4.4.2) ทำการสกัดค่าคุณลักษณะ (Feature extraction) เพื่อแปลงภาพถ่ายทางดาวเทียมภาพสี RGB ขนาด 256*256*3 ให้เป็นเวกเตอร์ที่ผ่านการสกัดค่าคุณลักษณะซึ่งมีขนาดแตกต่างกันไปตามแต่ละ Pre-Trained Model โดยการสร้างโมเดลพื้นฐาน (Base Model) จากคอนเนตต์ (Convnets) ที่ผ่านการฝึกอบรมมาแล้ว (Pre-Trained) โดยเลือกชั้น (Layers) ของ Pre-Trained Model และตรึง Convolutional Base เพื่อใช้เป็นตัวสกัดค่าคุณลักษณะเพราะป้องกันไม่ให้เกิดการโอเวอร์ฟิต น้ำหนักในชั้นที่กำหนดระหว่างการฝึกซึ่งจะแปลงเป็นเวกเตอร์โดยหาค่าเฉลี่ยของแต่ละตำแหน่งเชิงพื้นที่ โดยใช้ชั้น GlobalAveragePooling ซึ่งเวกเตอร์ที่ได้จะมีขนาดแตกต่างกันไปตามแต่ละ Pre-Trained Model (เนื่องจากเวกเตอร์ที่ผ่านการสกัดค่าคุณลักษณะในแต่ละ Pre-Trained Model มีขนาดไม่เท่ากัน ส่งผลให้ ถ้าเราเพิ่ม Fully Connected Layer เข้าไปที่ตัวแบบส่วนบน จะมี Hidden Units ไม่เท่ากัน ซึ่งอาจส่งผลต่อการฝึกเพราะอยู่นอกเหนือการควบคุม)

4.4.3) ทำการจำแนก (Classify) เพื่อแปลงคุณลักษณะเหล่านี้เป็นการคาดคะเนว่าภาพนี้เป็นหมวดหมู่ใด และเพื่อให้ตัวแบบมีการเรียนรู้ที่เพิ่มมากขึ้นจึงเพิ่ม Fully Connected Layers อีก 3 ชั้น Dropout 1 ชั้น และเปลี่ยน Prediction Layer ซึ่งทั้ง 3 ส่วนแสดงดังนี้

Dense(256, activation='relu',kernel_initializer=HeNormal())
Dense(512, activation='relu',kernel_initializer=HeNormal())
Dense(1920, activation='relu',kernel_initializer=HeNormal())
Dropout(0.2) เพื่อลด Overfit Dense(4, activation='softmax')
(มี Hidden Units = 4 ตามหมวดหมู่ที่เรากำหนด)

และมีการปรับจูนไฮเปอร์พารามิเตอร์ ดังนี้ Learning Rate = 0.0001, Optimizer = Adam, Loss = Sparse Categorical Crossentropy, Metrics = Accuracy, Epochs = 100 (ซึ่งเวลาที่ใช้ในการฝึกทดลองครั้งหนึ่งเมื่อ kernel_initializer = HeNormal) สำหรับภาพรวมของการวิเคราะห์ข้อมูลแสดงในรูปที่ 8

5) ผลการวิจัย

5.1) Experimental Setup

ในการทดลองผู้วิจัยได้ใช้ Google Colab Pro System specs: GPUs = [K80, P100, T4], CPUs: 2 x vCPU และ RAM: 24GB ในการรันโมเดล และได้ใช้ Open Source Deep Learning Software Libraries ที่ชื่อ TensorFlow (Version 2.6.0) และ Keras (Version 2.6.0) เข้ามาช่วย

5.2) ประสิทธิภาพของแต่ละตัวแบบ

ตารางที่ 2 แสดงประสิทธิภาพของตัวแบบที่ได้จากการถ่ายโอนความรู้จากตัวแบบ DenseNet201 มีค่าความแม่นยำ (Accuracy) ทดสอบในชุดข้อมูลทดสอบ (Test Dataset) สูงถึง 96.88%

5.3) ประสิทธิภาพตัวแบบ

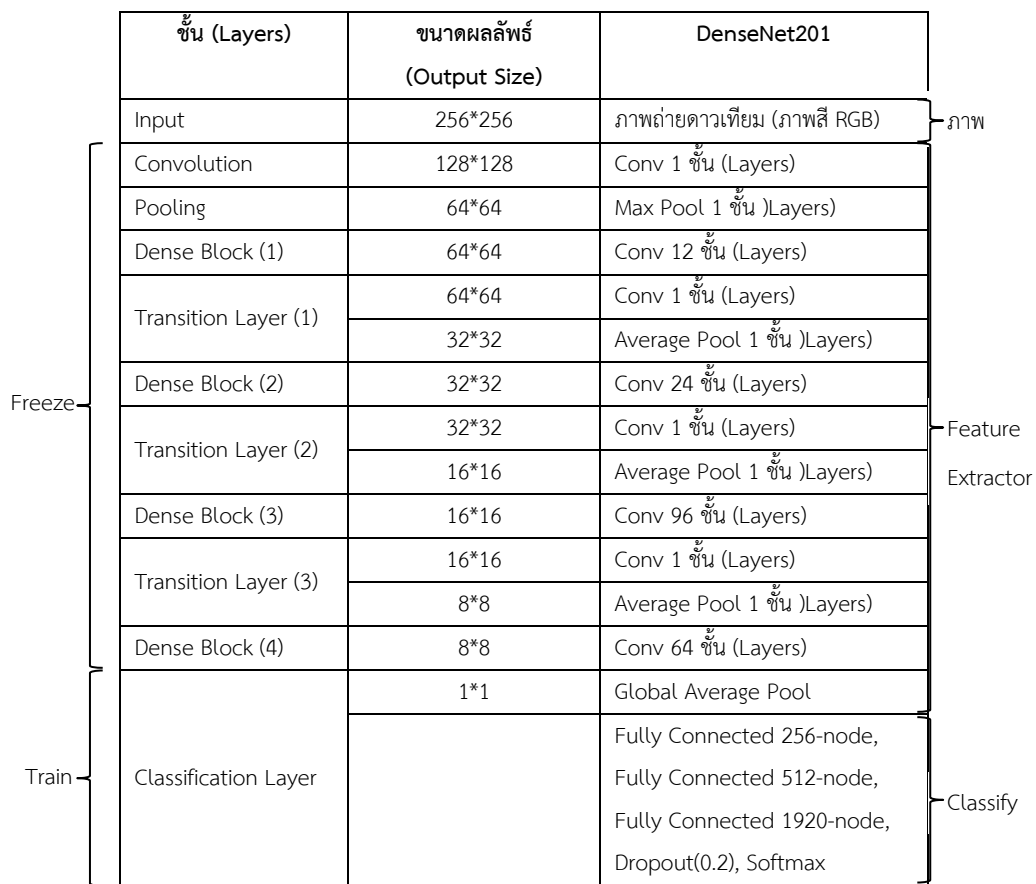
ตารางที่ 3 แสดงประสิทธิภาพตัวแบบ สรุปได้ว่า จะเลือกใช้ตัวแบบที่ได้จากการถ่ายโอนความรู้มาจากตัวแบบ DenseNet

201 ของการทดลองที่ 2 (การปรับแต่งข้อมูลแบบสโม่ท) เพราะค่า Loss ค่อนข้างไม่ผันผวน ซึ่งหมายความว่าค่าพารามิเตอร์ที่ใช้ในการเรียนรู้ค่อนข้างคงที่ จึงใช้เวลาในการฝึกต่อรอบน้อยกว่า

5.4) ประสิทธิภาพตัวแบบที่ได้จากการถ่ายโอนความรู้มาจากรูปแบบ DenseNet201 ของการทดลองที่ 2

รูปที่ 9 แสดงผลการจำแนกอาคารประเภทต่าง ๆ และการวัดประสิทธิภาพตัวแบบของงานวิจัยชิ้นนี้จะพิจารณาจาก

Confusion Matrix และ ROC Curve ดังแสดงในรูปที่ 10 รูปที่ 11 และตารางที่ 4 จากผลที่ได้สามารถสรุปได้ว่า ตัวแบบสามารถจำแนกได้ดีในหมวดหมู่สนามบินและวัด แต่ยังจำแนกคลาดเคลื่อน (Misclassification) ในหมวดหมู่โรงเรียนและสนามฟุตบอล สำหรับการวิเคราะห์ว่าตัวแบบจำแนกภาพว่าเป็นอาคารประเภทนั้น ๆ จากองค์ประกอบใดของภาพ ผู้วิจัยจะแสดงโดยการสร้างภาพนิทัศน์จากข้อมูลด้วย Grad-CAM ดังแสดงในรูปที่ 12



รูปที่ 8 : ภาพรวมของการวิเคราะห์ข้อมูล

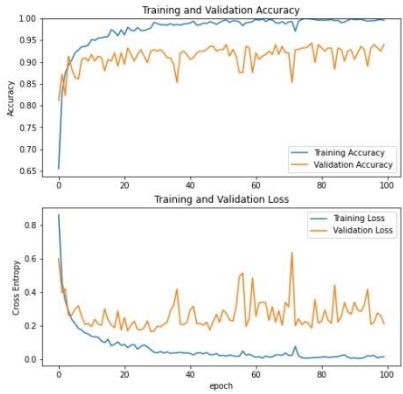
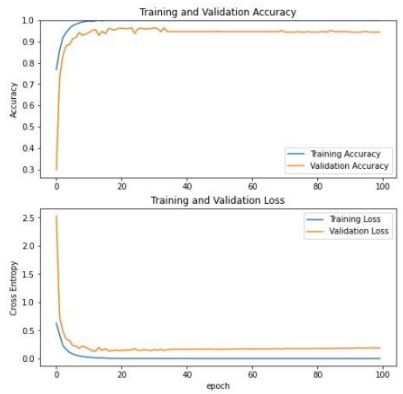
ตารางที่ 2 : ประสิทธิภาพของแต่ละตัวแบบ

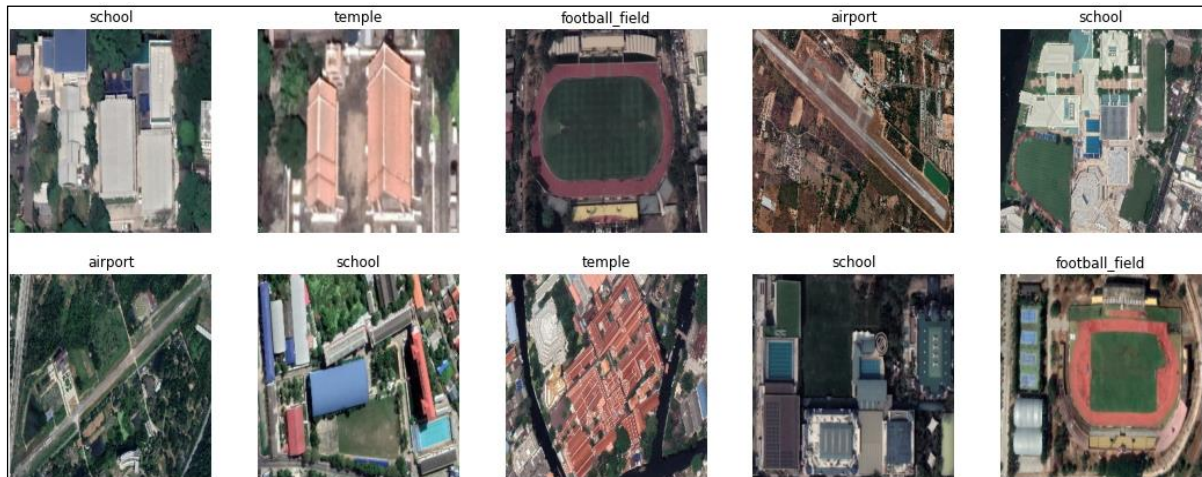
การทดลอง	Pre-Trained Model	Train (acc)	Train (loss)	Val (acc)	Val (loss)	Test (acc)	Test (loss)
1	VGG16	0.9907	0.029	0.8943	0.4294	0.8281	0.7305
	InceptionV3	0.9814	0.0503	0.8792	0.5086	0.8594	0.345
	ResNet152V2	0.9953	0.0135	0.9245	0.2751	0.8594	0.4104
	DenseNet201	0.9953	0.0153	0.9396	0.212	0.9688	0.1421
	Xception	0.9823	0.041	0.9094	0.2288	0.875	0.5842

ตารางที่ 2 : ประสิทธิภาพของแต่ละตัวแบบ (ต่อ)

การทดลอง	Pre-Trained Model	Train (acc)	Train (loss)	Val (acc)	Val (loss)	Test (acc)	Test (loss)
2	VGG16	1	1.09E-04	0.9405	0.3554	0.8281	0.8924
	InceptionV3	1	8.55E-06	0.9405	0.3057	0.8281	0.8924
	ResNet152V2	1	4.53E-06	0.9345	0.3546	0.875	0.4277
	DenseNet201	1	5.44E-06	0.9444	0.1874	0.9688	0.1159
	Xception	1	7.79E-06	0.9067	0.3652	0.8594	0.7374

ตารางที่ 3 : ประสิทธิภาพตัวแบบที่ได้จากการถ่ายโอนความรู้ (Transfer Learning) มาจากตัวแบบ DenseNet201

	การทดลองที่ 1	การทดลองที่ 2
กราฟ Accuracy/Loss		
ขนาดตัวแบบ	89.5 MB	89.5 MB
เวลาที่ใช้ในการฝึกต่อรอบ	1 ชั่วโมง 15 นาที	20 นาที



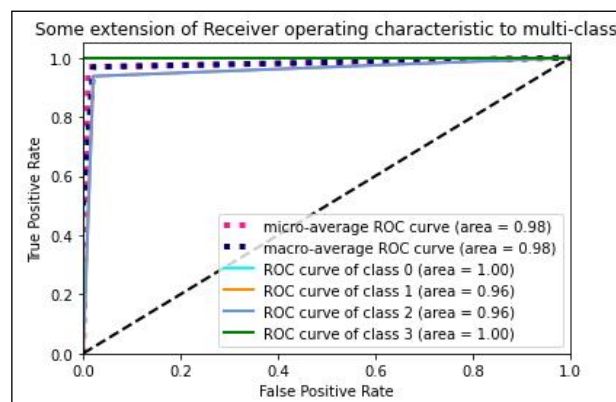
รูปที่ 9 : ตัวอย่างผลการจำแนกอาคารประเภทต่าง ๆ จากตัวแบบที่ได้

แหล่งที่มา : Google Earth [10]

สนามบิน				สนามฟุตบอล			
		Predicted Class				Predicted Class	
Actual		Negative	Positive	Actual		Negative	Positive
	Negative	48	0		Negative	47	1
	Positive	0	16		Positive	1	15

โรงเรียน				วัด			
		Predicted Class				Predicted Class	
Actual		Negative	Positive	Actual		Negative	Positive
	Negative	47	1		Negative	48	0
	Positive	1	15		Positive	0	16

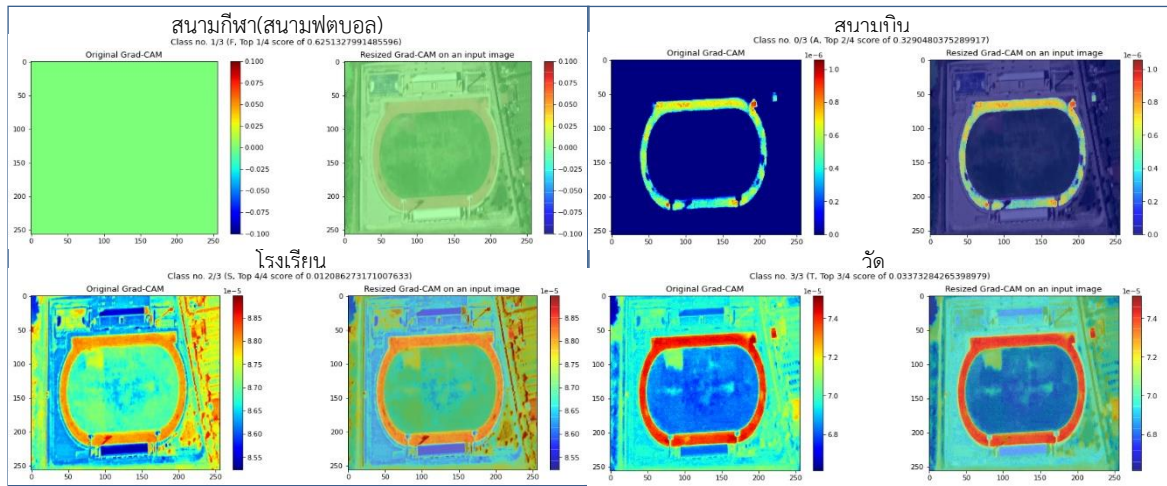
รูปที่ 10 : Confusion Matrix



รูปที่ 11 : ROC Curve

ตารางที่ 4 : Classification Report

	precision	recall	f1-score	support
สนามบิน	1.00	1.00	1.00	16
สนามฟุตบอล	0.94	0.94	0.94	16
โรงเรียน	0.94	0.94	0.94	16
วัด	1.00	1.00	1.00	16
accuracy			0.97	64
macro avg	0.97	0.97	0.97	64
weighted avg	0.97	0.97	0.97	64



รูปที่ 12 : การสร้างภาพนิทัศน์จากข้อมูล (Data Visualization) ด้วย Grad-CAM (Gradient-weighted CAM)

แหล่งที่มา : Google Earth, Image © 2021 Maxar Technologies [10]

จากรูปที่ 3 ข้อ (4) ได้ดำเนินการตามหัวข้อที่ 3.3) เพื่อตรวจสอบประสิทธิภาพของตัวแบบจาก Grad-CAM ซึ่งผลลัพธ์ที่ได้แสดงในรูปที่ 12 เป็นการแสดงผลการวิเคราะห์ตัวแบบจำแนกภาพสนามกีฬาเฉลิมพระเกียรติ 84 พรรษา บางบอน ว่าเป็นอาคารประเภทใด และจากองค์ประกอบใดของภาพ ซึ่งสรุปได้ว่าตัวแบบจำแนกภาพนี้ออกมาเป็นหมวดหมู่สนามกีฬา (สนามฟุตบอล) โดยดูจากองค์ประกอบของภาพ (รูปที่ 12 ภาพซ้ายบน) ดังนั้นในหมวดหมู่สนามกีฬา (สนามฟุตบอล) ตัวแบบดูจากสนามฟุตบอลและสิ่งที่เป็นสีเขียว หมวดยุทธศาสตร์โรงเรียนตัวแบบดูจากสนามฟุตบอลและอาคารต่าง ๆ (รูปที่ 12 ภาพซ้ายล่าง) ด้วยเหตุที่ตัวแบบจำแนกภาพว่าเป็นหมวดหมู่สนามกีฬา (สนามฟุตบอล) หรือโรงเรียนนั้น ตัวแบบดูจากสนามฟุตบอลจึงทำให้จำแนกคลาดเคลื่อนอยู่บ้างในหมวดหมู่นี้ (รูปที่ 12 ภาพซ้าย) ซึ่งคนเองก็อาจเกิดความสับสนเช่นกันหากภาพพิจารณาเฉพาะส่วนสนามฟุตบอล สำหรับหมวดหมู่สนามบิน ตัวแบบดูจากลานบิน (Runway) และต้นไม้จำนวนมาก (รูปที่ 12 ภาพขวาบน) หมวดหมู่วัด ตัวแบบดูจากหลังคาสี่เหลี่ยมเป็นกลุ่ม ๆ (รูปที่ 12 ภาพขวาล่าง) จึงทำให้จำแนกได้ดีในหมวดหมู่นี้ (รูปที่ 12 ภาพขวา)

5.5) การประยุกต์ใช้

5.5.1) *ปรับปรุงแผนที่ทางภูมิศาสตร์* ตัวแบบที่สร้างสามารถนำไปใช้ในเรื่องปรับปรุงแผนที่ทางภูมิศาสตร์ว่าพื้นที่บนแผนที่นี้เป็นสิ่งปลูกสร้างอะไร หรือตรวจสอบพื้นที่นั้น ๆ ว่าติดแท็กใน

แผนที่ถูกต้องหรือไม่ เพื่อประหยัดเวลาและบุคลากรในการวิเคราะห์

5.5.2) *การตรวจจับการใช้ที่ดิน* ตัวแบบที่สร้างสามารถนำไปใช้ในเรื่องการตรวจจับการใช้ที่ดินว่าสิ่งปลูกสร้างนั้นขออนุญาตสร้างอย่างถูกต้องหรือลักลอบทำแบบผิดกฎหมาย เพื่อที่จะสามารถตรวจสอบได้อย่างทันทั่วถึง และยังสามารถทำการสำรวจพื้นที่เดียวกันในเวลาที่แตกต่างกัน ดังแสดงในรูปที่ 13 ซึ่งผลการจำแนกจากตัวแบบที่สร้างได้ จำแนกทั้ง 2 ภาพว่าเป็นหมวดหมู่สนามบิน จึงสามารถสรุปได้ว่าตัวแบบสามารถคาดการณ์สิ่งปลูกสร้างได้พอประมาณ

6) สรุปและอภิปรายผลการวิจัย

จากการดำเนินการตามบทวิจัยข้างต้น จะได้ระบบการจำแนกอาคารประเภทต่าง ๆ จากภาพถ่ายดาวเทียมในหมวดหมู่สนามบิน สนามกีฬา (สนามฟุตบอล) โรงเรียนและวัดซึ่งอยู่ในประเทศไทยที่มีประสิทธิภาพและประหยัดเวลาและบุคลากรมากขึ้น ตลอดจนสามารถนำมาประยุกต์ใช้ในเรื่องการตรวจจับการใช้ที่ดิน ปรับปรุงแผนที่ทางภูมิศาสตร์ เป็นต้น แต่ชุดข้อมูลภาพถ่ายทางดาวเทียมในแต่ละประเภท ผู้วิจัยจะเลือกเฉพาะภาพที่มีคุณลักษณะเด่น ๆ ตามที่ผู้วิจัยกำหนด ดังนั้นหากเป็นภาพที่ไม่มีคุณลักษณะเด่นตามที่ผู้วิจัยกำหนดอาจจำแนกประเภทคลาดเคลื่อนได้ นอกจากนี้ผู้วิจัยยังได้แสดงการสร้างภาพนิทัศน์จากข้อมูล ว่าตัวแบบที่ดีที่สุดจำแนกภาพว่าเป็นอาคารประเภทนั้น ๆ จากองค์ประกอบใดของภาพด้วย Grad-CAM ซึ่งทำให้ทราบว่าตัวแบบยังสามารถพัฒนาเพื่อเพิ่มประสิทธิภาพได้ แต่

อาจจะต้องเตรียมชุดข้อมูลในแต่ละหมวดหมู่ให้มีความหลากหลายมากยิ่งขึ้น อาจมีการเตรียมข้อมูลรูปแบบอื่น ๆ ร่วม

ด้วย เช่น การแยกบริเวณรูปภาพ (Image Segmentation) หรือหาวิธีการเพิ่มประสิทธิภาพตัวแบบวิธีใหม่ ๆ ร่วมด้วย



รูปที่ 13 : ภาพถ่ายดาวเทียมท่าอากาศยานนานาชาติเบตงสำรวจเมื่อ 5/8/2016 (ภาพซ้าย)
และภาพถ่ายดาวเทียมท่าอากาศยานนานาชาติเบตงสำรวจเมื่อ 17/4/2019 (ภาพขวา)

แหล่งที่มา : Google Earth, Image © 2021 Maxar Technologies [10], Google Earth, Image © 2021 CNES / Airbus [10]

REFERENCES

- [1] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
- [2] *TensorFlow 2.6.0*. (2021). Google.
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” 2011. [Online]. Available: arXiv:1106.1813.
- [4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [5] T. Siriborvornratanakul. (2021). CNN Wrap-up [PowerPoint slides]. Available: <https://as.nida.ac.th/~thitirat/homepage/index.html>
- [6] M. Pritt and G. Chern, “Satellite image classification with deep learning,” in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop (AIPR)*, Washington, DC, USA, Oct. 10–12, 2017, pp. 1–7.
- [7] P. Helber, B. Bischke, A. Dengel, and D. Borth, “EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [8] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014. [Online]. Available: arXiv:1412.6980.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034, doi: 10.1109/ICCV.2015.123.
- [10] *Google Earth*. (2015). Google. Accessed: Feb. 1, 2021. [Online]. Available: <https://earth.google.com/web/>
- [11] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014. [Online]. Available: arXiv:1409.1556.
- [12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” 2015. [Online]. Available: arXiv:1512.00567.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” 2016. [Online]. Available: arXiv:1603.05027.
- [14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [15] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” 2016. [Online]. Available: arXiv:1610.02357.