การวิเคราะห์แบ่งส่วนพื้นที่ภาพถ่ายทางอากาศด้วย Generative Adversarial Networks

กิตตากร วิริยะศาสตร์ [1, 2*] วรากร เลื่องลือวุฒิ [1] วิชัย แผ้วเกษม [1]
พันธุ์เทพ แก้วมงคล [1] สัญญา มิตรเอม [2] และ พันศักดิ์ เทียนวิบูลย์ [3]

วันที่รับ 28 พฤศจิกายน 2566 วันที่แก้ไข 7 มีนาคม 2567 วันตอบรับ 3 เมษายน 2567

บทคัดย่อ

บทความนี้กล่าวถึงการแบ่งส่วนพื้นที่แบบ Semantic Segmentation จากภาพถ่ายทางอากาศ ซึ่งเป็นภาพที่ได้จากอากาศยานไร้คนขับ (Unmanned Aerial Vehicle: UAV) มาวิเคราะห์การจำแนกพื้นที่ด้วยวิธี Generative Adversarial Networks (GANs) โดยการจำแนกพื้นที่ด้วยระบบสีแบบ RGB มาทำการจำแนกพื้นที่ทั้งหมด 10 พื้นที่ เช่น สนามบิน สนามกีฬา ป่าไม้ พื้นที่ทางการเกษตร แม่น้ำ บ่อน้ำ รถ ถนน สิ่งก่อสร้าง และพื้นที่อื่น ๆ ซึ่งในการทดลองนี้ได้ใช้โมเดลใน UNET ได้แก่ MobileNetV2, ResNet50, ResNet50V2, DenseNet201 และ VGG16 มาเป็น Generator บน Generative Adversarial Networks ไว้ในการจำแนกพื้นที่ของภาพถ่ายทางอากาศ จากการทดลองพบว่า โมเดลแต่ละโมเดลมีความแม่นยำโดยประมาณ 80% และมีความเร็วในการทำงานต่อเฟรมอยู่ที่ 2 วินาที

คำสำคัญ : อากาศยานไร้คนขับ, การจำแนกพื้นที่, เครือข่ายคู่ต่อสู้ช่วยสร้าง, RGB

[1] ส่วนงานวิศวกรรมการสื่อสารข้อมูลทางอิเล็กทรอนิกส์และเครือข่ายคอมพิวเตอร์, สถาบันเทคโนโลยีป้องกันประเทศ

[2] ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์, คณะวิศวกรรมศาสตร์, มหาวิทยาลัยธรรมศาสตร์

[3] ภาควิชาวิศวกรรมไฟฟ้า, คณะวิศวกรรมศาสตร์, มหาวิทยาลัยเกษตรศาสตร์

[*] ผู้แต่ง, อีเมล: kittakorn.v@dti.or.th

Analysis of Segmentated Images by Generative Adversarial Networks

Kittakorn Viriyasatr [1, 2*] Warakorn luangluewut [1] Wichai Pawgasame [1]
Pantape Kaewmongkol [1] Sanya Mitaim [2] and Phunsak Thiennviboon [3]

Received 28 November 2023, Revised 7 March 2024, Accepted 3 April 2024

Abstract

This article presents the study on semantic segmentation for aerial image classification using Generative Adversarial Networks (GANs). The aerial images were acquired by an Unmanned Aerial Vehicle (UAV). The proposed method utilized the RGB color space to classify into the total of 10 land-cover classes, including airport, stadium, forest, agricultural area, river, pond, car, road, building, and others. The experiments were conducted using MobileNetV2, ResNet50, ResNet50V2, DenseNet201, and VGG16 as the generator in the GAN framework. The experimental results demonstrate that each model achieved the accuracy of approximately 80% and the processing speed of 2 seconds per frame.

Keywords : Unmanned aerial vehicles, Segmentation, Generative adversarial networks, RGB

[1] Data Communication Division, Defence Technology Institute

[2] Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University

[3] Department of Electrical Engineering, Faculty of Engineering, Kasetsart University

[*] Corresponding author, E-mail: kittakorn.v@dti.or.th

1. บทนำ

บทความนี้กล่าวถึงการแบ่งส่วนพื้นที่แบบ Semantic Segmentation จากภาพถ่ายทางอากาศ โดยการนำภาพที่ได้จากอากาศยานไร้คนขับ (Unmanned Aerial Vehicle: UAV) มาวิเคราะห์การจำแนกพื้นที่ (Segmentation) [1] - [3] ซึ่งที่ผ่านมาได้มีงานที่คล้ายกันนี้ [4] - [5] จำแนกพื้นที่ด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน แต่ในงานวิจัยนี้จะทำการทดลองโดยใช้ Generative Adversarial Networks (GANs) [6] - [8] ในการจำแนกพื้นที่ทั้งหมด 10 คลาส ได้แก่ สนามบิน สนามกีฬา ป่าไม้ พื้นที่ทางการเกษตร แม่น้ำ บ่อน้ำ รถ ถนน สิ่งก่อสร้าง และพื้นที่อื่น ๆ เพื่อหาความแม่นยำและความเร็วในการทำงานเพื่อให้สอดคล้องกับความเร็วของอากาศยานไร้คนขับ ซึ่งคล้ายกับการทดลอง [9] ที่ทำการตรวจจับวัตถุในภาพถ่ายทางอากาศด้วยเทคนิคโครงข่ายประสาทเทียม

สำหรับการวิเคราะห์จำแนกพื้นที่นั้น จำเป็นต้องมีผู้เชี่ยวชาญด้านภูมิศาสตร์เป็นผู้วาดแบบร่างเพื่อตีกรอบจำแนกพื้นที่ ซึ่งวิธีการนี้ใช้ระยะเวลานาน สามารถแก้ไขให้เร็วขึ้นได้โดยใช้ผู้เชี่ยวชาญหลายท่าน ซึ่งอาจช่วยลดระยะเวลาได้ แต่ก็มีค่าใช้จ่ายที่สูงขึ้น นอกจากนี้ยังมีปัญหาด้านพื้นที่ ถ้าพื้นที่ที่เป็นจุดสนใจนั้นยากต่อการเข้าถึง เช่น พื้นที่ป่าที่เป็นเนินเขารกทึบ พื้นที่ที่ต้องใช้เจ้าหน้าที่ในการสำรวจจำนวนหลายท่าน พื้นที่อันตรายที่อยู่ในเขตการทดสอบของจรวด ซึ่งอาจทำให้เกิดอันตรายต่อตัวเจ้าหน้าที่ได้ เป็นต้น

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อมุ่งแก้ไขปัญหาในการจำแนกพื้นที่ดังกล่าวด้วย Generative Adversarial Networks (GANs) โดยใช้ภาพถ่ายทางอากาศเพื่อให้สามารถนำไปประยุกต์ใช้งานได้ต่อไป อย่างไรก็ตาม ข้อจำกัดในงานวิจัยมีปัจจัยดังนี้

1) ด้านข้อมูลภาพถ่ายทางอากาศ คลาดเคลื่อนหรือข้อมูลไม่สมดุลที่ส่งผลให้มีข้อจำกัดในการใช้งานโมเดล ตัวอย่างเช่น เมื่อทำการตีกรอบพื้นที่จำนวนภาพที่มีพื้นที่ของป่ามากกว่าแม่น้ำมาก ๆ สมมุติว่า มี 1,000 ภาพ ที่มีป่า แต่มีภาพที่มีแม่น้ำเพียง 100 ภาพ อาจทำให้ประสิทธิภาพของโมเดลเอนเอียงไปในพื้นที่ของป่ามากกว่าแม่น้ำ เป็นต้น

2) ข้อจำกัดด้านการตีความพื้นที่ ขนาดของภาพถ่ายทางอากาศเป็นสิ่งสำคัญ ภาพขนาดเล็กอาจตีความพื้นที่ได้ยากกว่าภาพที่มีขนาดใหญ่ในการจำแนกว่าเป็นคลาสใด เนื่องจากมุมกล้องและความแตกต่างของภาพในการถ่าย ทำให้พื้นที่ย่อขยายแตกต่างกัน ส่วนภาพที่ขนาดใหญ่เกินไป ในที่นี้หมายถึงมีขนาดพิกเซลจำนวนมาก เป็นอุปสรรคในการทดสอบโมเดล เนื่องจากประสิทธิภาพในการประมวลผลของคอมพิวเตอร์อาจไม่เพียงพอและระยะเวลาการประมวลผลข้อมูล อาจจำเป็นต้องแบ่งภาพหรือลดขนาดพิกเซลตามความเหมาะสมก่อนทำการทดสอบโมเดล

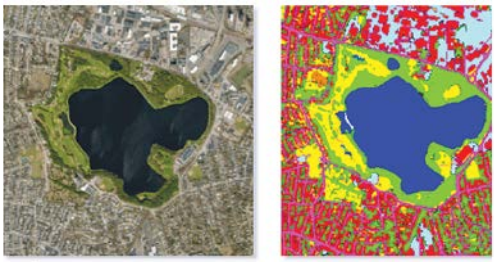
ดังนั้น งานวิจัยนี้จึงเน้นการแก้ไขปัญหาและข้อจำกัดดังกล่าวข้างต้น เพื่อเพิ่มประสิทธิภาพในการใช้งานโมเดลในการวิเคราะห์ภาพถ่ายทางอากาศได้มากยิ่งขึ้น และสะดวกต่อการจำแนกพื้นที่ โดยยึดหลักการเพิ่มความแม่นยำของการจำแนกพื้นที่และความน่าเชื่อถือในการจำแนกพื้นที่ในภาพถ่ายทางอากาศ รวมถึงความเร็วในการประมวลผลข้อมูลภาพถ่ายทางอากาศ

2. ทฤษฎีที่เกี่ยวข้อง

2.1 Image Segmentation

Image Segmentation คือ การจำแนกว่าพิกเซลแต่ละจุดคือพื้นที่ของคลาสใดที่เราสนใจจะได้ผลออกมาเป็นพื้นที่สีต่าง ๆ ซึ่งแต่ละสีหมายความถึง

ลักษณะที่แตกต่างกัน เช่น บ้าน ถนน ต้นไม้ ฯลฯ ตัวอย่างการทำงานของ Image Segmentation ดังรูปที่ 1



รูปที่ 1 *ตัวอย่างการทำงานของ Image Segmentation [10]*

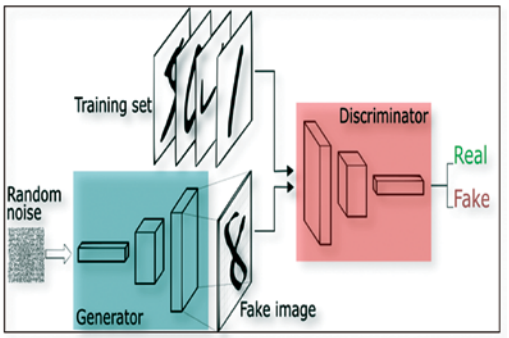
วิธีการ Image Segmentation เหมาะกับการใช้งานเชิงพื้นที่ โดยเฉพาะกับภาพถ่ายทางอากาศ เนื่องจากต้องการตีความภาพเชิงพื้นที่ให้ละเอียดสูงสุดในระดับพิกเซล ซึ่งเพียงพอและเหมาะสมสำหรับการใช้งานดังกล่าว

2.2 Generative Adversarial Networks (GANs) [11] เหมาะสำหรับงาน Segmentation เนื่องจาก GANs สามารถสร้างภาพจำลองขึ้นมาเพื่อใช้ในการจำแนกพื้นที่ ซึ่งหน้าที่ของ GANs คือ การสร้างข้อมูล ซึ่งการที่ GANs จะสร้างข้อมูลออกมาได้นั้นประกอบไปด้วย 2 ส่วน คือ

2.2.1 Generator ทำหน้าที่สร้างข้อมูลให้ตรงกับข้อมูลรูปภาพจริงที่เราต้องการให้มากที่สุด

2.2.2 Discriminator จะเป็นตัว Classifier ทำการเรียนรู้ข้อมูลที่รับมาว่าอะไรเป็นสิ่งที่มาจากข้อมูลจริง และอะไรมาจาก Generator ที่สร้างขึ้นมาเลียนแบบข้อมูลภาพจริงให้เหมือนภาพจริงมากที่สุดดังรูปที่ 2

Generator เป็นองค์ประกอบหลักของ GANs ที่ใช้ในการสร้างภาพ โดยเริ่มจากการใช้ Noise เป็นอินพุต (Input) สำหรับ Generator และ



รูปที่ 2 *จำลองการทำงานของ GANs [12]*

เอาต์พุต (Output) เป็นภาพจำลอง (Fake Image) ที่พยายามให้ใกล้เคียงกับภาพที่ต้องการมากที่สุด โดยสมการของ Generator แสดงดังสมการที่ (1)

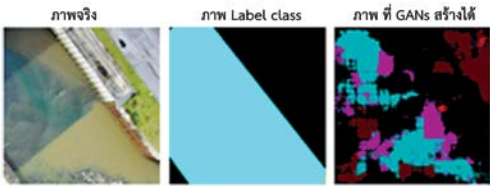
$$\min_G V(G) = \mathbb{E}_{z\sim p_z(z)}[log\,(1 - D(G(z)))] \quad (1)$$

Discriminator [12] เป็นตัว Classifier ที่เรียนรู้จากข้อมูลที่รับมา เพื่อจำแนกว่าอะไรเป็นข้อมูลจริงและอะไรมาจาก Generator จากนั้นจะส่ง Feedback กลับไปให้ Generator เพื่อให้เรียนรู้และปรับปรุงคุณภาพของภาพจำลอง เพื่อให้ใกล้เคียงกับรูปภาพต้นฉบับมากที่สุด คล้ายกับเป็นการแข่งขันระหว่างสองฝ่าย โดยสมการของ Discriminator แสดงในสมการที่ (2)

$$\max_{\theta_d}[\mathbb{E}_{x\sim\rho_{data}} log D_{\theta_d}(x) + \mathbb{E}_{z\sim\rho_{(z)}} \log(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right))] \quad (2)$$

Generator และ Discriminator [12] ใช้โมเดล CNN (Convolutional Neural Network) โดยงานวิจัยนี้ใช้ U-Net ซึ่งเป็นโครงข่ายประสาทเทียมแบบคอนโวลูชันชนิดหนึ่ง โดย U-Net ประกอบด้วยโมเดลหลายรูปแบบ การวิเคราะห์ว่าโมเดลใดเหมาะสมที่สุดสำหรับอากาศยานไร้คนขับ

พิจารณาจากความแม่นยำ (Accuracy) และความเร็วในการทำงานของโมเดล ผลการทดลองแสดงในรูปที่ 3 ซึ่งภาพที่ GANs อ่านได้จะแสดงผลตามสีที่กำหนดไว้ใน Label Map ว่าพื้นที่ใดควรมีสีอะไร



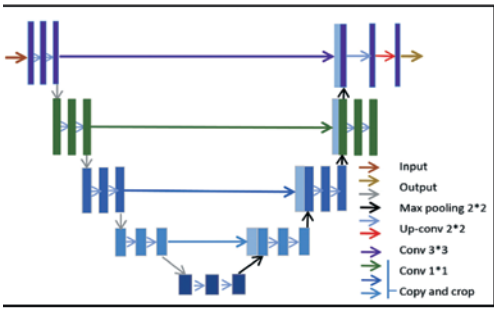
รูปที่ 3 *ตัวอย่างผลการอ่านภาพด้วย GANs*

ตามที่กำหนดไว้แต่แรก

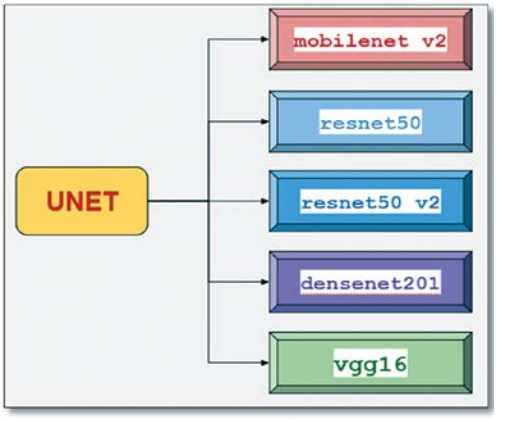
2.3 U-Net [13] U-Net เป็นโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ที่เรียนรู้แบบ Supervised Learning หรือการเรียนรู้แบบมีผู้สอน ยกตัวอย่างเช่น การทำ Classification แยกรูปภาพว่ามีรถอยู่ในภาพหรือไม่มี ในกรณีนี้รูปภาพจะถูกส่งผ่าน Layers ต่าง ๆ ของ Classification Network ที่ทำหน้าที่ เช่น Convolution, Max Pooling, Dropout จนกระทั่งได้ผลลัพธ์ออกมาดังรูปที่ 4 โดย งานวิจัยนี้เลือกใช้ U-Net ด้วยเหตุผลที่ว่าโครงสร้างนี้เป็นโครงสร้างที่มีการสกัดคุณลักษณะของภาพได้อย่างมีประสิทธิภาพจากงานวิจัยที่ผ่านมา

ส่วนแรก ของ U-Net (ส่วนที่ 1 ในรูปที่ 4) ประกอบด้วย Convolution Block (3×3 conv) และการลดทอน down-sampling ด้วย Max Pooling (2×2 max pool) เพื่อดึงคุณลักษณะ (Features) ออกมาจากภาพ จะเห็นว่าการดึงคุณลักษณะนี้เกิดขึ้นหลายระดับตั้งแต่ High-resolution Features ไปจนถึง Low-resolution Features

ส่วนที่สอง ของ U-Net (ส่วนที่ 2 ในรูปที่ 4) ประกอบไปด้วยการทำ Up-sampling (up-conv 2×2) และการทำ Convolution คือ การนำ Features ที่ได้จากส่วนที่ 1 มาใช้ในการสร้างภาพที่ถูกตัดเรียบร้อยแล้ว (Segmented Output) ลูกศรสีเทาในรูปแสดงให้เห็นว่า ส่วนที่ 2 จะนำ Features ในแต่ละระดับจากส่วนที่หนึ่งมารวมในการคำนวณด้วย สุดท้ายขั้นตอนจะประกอบด้วย conv 1×1 เพื่อแปลงขนาดของ Feature จาก 62 (ซึ่งเท่ากับความลึกของ Layer ก่อนสุดท้าย) ให้เหลือเท่ากับจำนวนคลาส (Class) ที่ผู้วิจัยได้จำแนกเอาไว้ในรูปที่ 4 ซึ่งใน U-Net จะประกอบด้วยโมเดลหลายโมเดลดังรูปที่ 5 โมเดลที่มีทั้งหมดใน U-Net โดยที่โมเดล U-Net นั้นจะนำมาใช้เป็นส่วน Generator ของ Generative Adversarial Networks สำหรับการทำ Image Segmentation



รูปที่ 4 *แสดงโครงสร้างของ U-Net*



รูปที่ 5 *โมเดลที่มีทั้งหมดใน U-Net*

3. วิธีการดำเนินการ

3.1 การรวบรวมข้อมูล โดยชนิดข้อมูลเป็นไฟล์ภาพขนาด 5472×3648 pixels (8.8 MB)

เนื่องจากภาพต้นฉบับมีขนาดใหญ่เกินกว่าที่ U-Net จะนำเข้าข้อมูลภาพได้ จำเป็นต้องแบ่งภาพเป็นส่วน ๆ แต่ละส่วนทำหน้าที่เป็นอินพุตเพียงชุดเดียวของ U-Net ดังนั้น จึงต้องแบ่งภาพใหญ่เป็นหลายส่วน เพื่อสร้างอินพุตหลายชุด กระบวนการนี้จะอธิบายในขั้นตอนการเตรียมภาพ (Image Pre-processing) โดยรูปที่ 6 แสดงตัวอย่างภาพต้นฉบับขนาด 5472×3648 pixels มีภาพทั้งหมด 1,300 ภาพ แต่สามารถใช้ได้ 1,286 ภาพ เนื่องจากบางภาพไม่ตรงกับคลาสที่กำหนดไว้

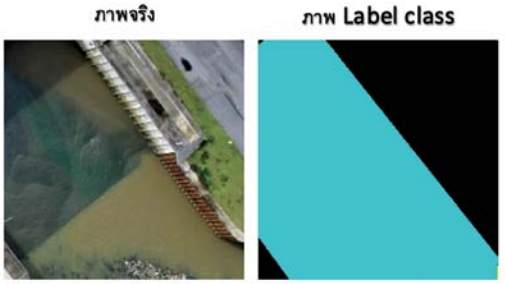
รูปที่ 6 *ตัวอย่างไฟล์รูปต้นฉบับ*

3.2 การ Label ข้อมูลหรือการตีกรอบจุดสังเกต

ขั้นตอนการติดป้ายกำกับ (Label) เริ่มต้นด้วยการกำหนดคลาสทั้งหมด ซึ่งผู้วิจัยสนใจทั้งหมด 10 คลาส ได้แก่ ป่า พื้นที่ทางการเกษตร รถ ถนน บ่อน้ำ แม่น้ำ สนามบิน สนามกีฬา สิ่งก่อสร้าง และพื้นที่อื่น ๆ ตามที่ผู้วิจัยต้องการวิเคราะห์จากภาพ จากนั้นผู้วิจัยจะเลือกการติดป้ายกำกับหรือการระบุตำแหน่งที่สนใจ โดยใช้โปรแกรม CVAT (Computer Vision Annotation Tool) สำหรับการติดป้ายกำกับหรือการระบุตำแหน่งที่สนใจจะใช้รูปแบบ Polygons เพื่อกำหนดให้เป็นรูปทรงที่ไม่ได้เป็นสี่เหลี่ยม ดังแสดงในรูปที่ 7 สำหรับการติดป้ายกำกับข้อมูลแบบ Polygons

รูปที่ 7 *การติดป้ายกำกับข้อมูลแบบ Polygons*

หลังจากนั้นไฟล์ที่ผู้วิจัยได้จะเป็นไฟล์ RGB ในแต่ละรูปที่ทำการติดป้ายกำกับเอาไว้ ซึ่งผลที่ได้แสดงดังรูปที่ 8



รูปที่ 8 *ผลที่ได้จากการติดป้ายกำกับ*

ผู้วิจัยจะทำการติดป้ายกำกับให้กับแต่ละคลาสที่สนใจ ด้วยการกำหนดค่าสี RGB เพื่อแยกความแตกต่างของลักษณะพื้นที่ โดยแต่ละพื้นที่ที่มีลักษณะแตกต่างกันจะได้รับสีที่แตกต่างกันตามรูปที่ 9

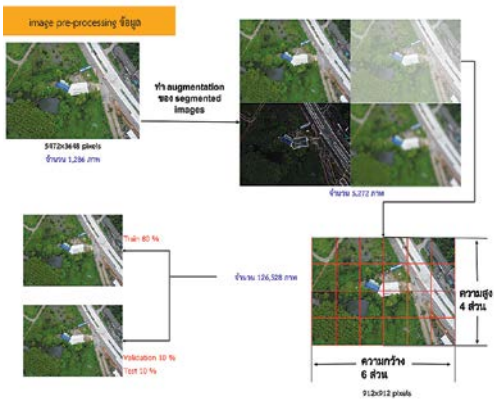
3.3 Image Pre-processing ข้อมูล

หลังจากตีความพื้นที่ที่สนใจแล้ว ภาพถ่ายทางอากาศทั้งหมด 1,286 ภาพ จะถูกขยายข้อมูล (Augmentation) ด้วยการทำภาพเบลอ การปรับความสว่างของภาพ และการเปลี่ยนสีของภาพถ่ายทางอากาศ เพื่อเพิ่มความหลากหลายของข้อมูล

รวมเป็นจำนวนทั้งหมด 5,272 ภาพ จากนั้นทำการแบ่งภาพเป็นพาร์ทิชัน (Partition) โดยแบ่งภาพตามมิติความกว้างออกเป็น 6 ส่วน และมิติความสูงออกเป็น 4 ส่วน รวมทั้งหมด 126,528 ภาพ ขั้นตอนนี้แสดงในรูปที่ 10

class_image	RGB
actions background	0,0,0
air_field	128,0,0
sports_field	0,128,0
forest	128,128,0
farm	0,0,128
water_surface	128,0,128
waterway	0,128,128
car	128,128,128
road	64,0,0
building	192,0,0

รูปที่ 9 *จำแนกคลาสแต่ละคลาสด้วย RGB*



รูปที่ 10 *Image Pre-processing ข้อมูล*

3.4 Train และ Test ข้อมูล

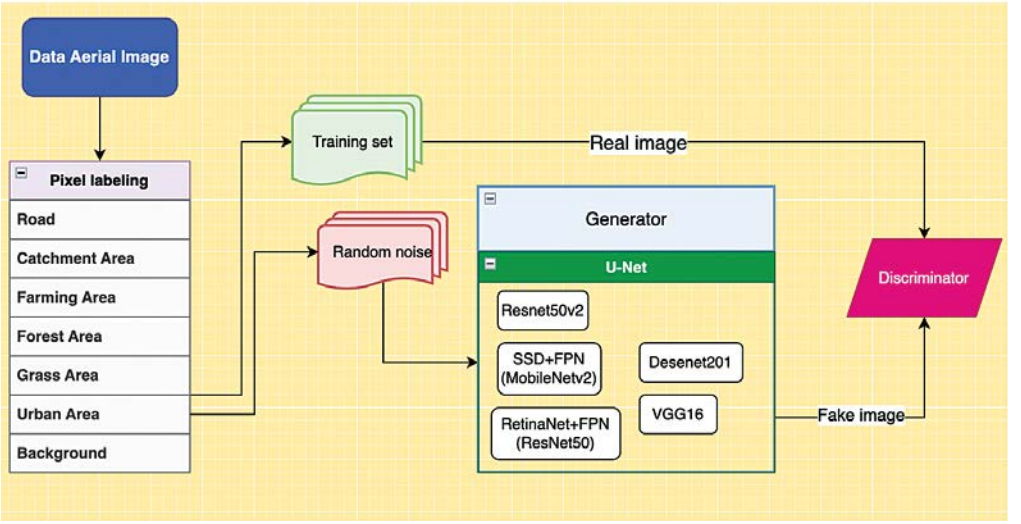
โมเดลที่เลือกสำหรับ Semantic Segmentation (U-Net) คือ MobileNetV2, ResNet50, ResNet50V2, DenseNet201 และ VGG16 ซึ่งเป็นส่วนของ U-Net สามารถแทนที่ด้วย Convolutional Neural Networks เพื่อสร้างคุณลักษณะของภาพ โมเดลเหล่านี้ถูกนำมาใช้เพื่อเปรียบเทียบประสิทธิภาพของ U-Net โดยพิจารณาจากความแม่นยำที่ได้จาก Generator ของ Generative Adversarial Networks ในการ Image Segmentation เนื่องจากภาพต้นฉบับมีขนาดใหญ่เกินกว่าที่ U-Net จะจัดการได้ ผู้วิจัยแก้ไขปัญหานี้โดยแบ่งภาพเป็นส่วน ๆ แต่ละส่วนจะทำหน้าที่เป็นอินพุตเพียงชุดเดียวของ U-Net

ดังนั้น จึงต้องแบ่งภาพใหญ่ออกเป็นหลายส่วนเพื่อสร้างหลายอินพุต ผู้วิจัยพบว่า การแบ่งพาร์ทิชันที่มีประสิทธิภาพ คือ การแบ่งมิติความกว้างของภาพออกเป็น 6 ส่วน และมิติความสูงเป็น 4 ส่วน ตามแผนการฝึกอบรม (Training) ที่แสดงในรูปที่ 11

ในขั้นตอนนี้ข้อมูลจะถูกแบ่งเป็น Train 80%, Validation 10% และ Test 10% สำหรับการฝึกโมเดลที่แตกต่างกัน โดยพารามิเตอร์สำหรับการฝึกจะเหมือนกันสำหรับโมเดลทั้งหมด และผู้วิจัยจะเปลี่ยนเฉพาะโมเดลพื้นฐานเท่านั้นในการทำ Image Segmentation จากจำนวนทั้งหมด 126,528 ภาพ

4. ผลการทดลอง

จากการทดลอง พบว่า โมเดลแต่ละโมเดลให้ผลความแม่นยำที่ไม่แตกต่างกันมากนัก ยกเว้น RestNet50 ซึ่งมีความแม่นยำน้อยที่สุดที่ 66.23% ส่วน MobileNetV2 มีความแม่นยำสูงสุดถึง 82.33% เนื่องจาก MobileNetV2 ออกแบบตามหลักการ Depthwise Separable Convolution เพื่อลดความ



รูปที่ 11 *ผังการ Training ข้อมูล*

ซับซ้อนของโมเดลและประหยัดทรัพยากร รวมถึงการใช้ Batch Normalization และ ReLU activation ระหว่าง Layers เพื่อเพิ่มความเร็วในการฝึกโมเดล นอกจากนี้ยังใช้โครงสร้างแบบ Bottleneck เพื่อลดการใช้ทรัพยากรและเพิ่มความลึกของโมเดล ทำให้ MobileNetV2 มีประสิทธิภาพสูงกว่าโมเดลอื่น ๆ ที่ทดสอบในชุดข้อมูลเดียวกัน

ตารางที่ 1 ผลการทดลองความแม่นยำของแต่ละโมเดล

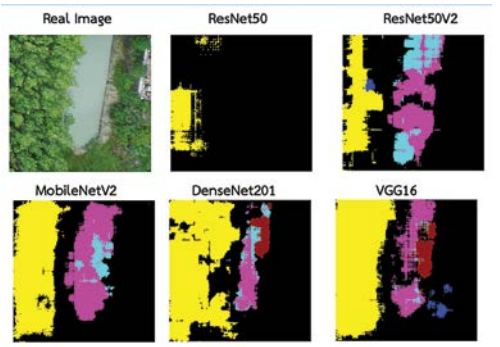
Model	mAP (%
MobileNetV2	82.33
ResNet50	66.23
ResNet50V2	79.32
DenseNet201	76.03
VGG16	80.34

ในการทดสอบเวลาที่ใช้ทำ Image Segmentation ในแต่ละเฟรม ผู้วิจัยทำการทดลองทั้งหมด 4 ครั้ง และคำนวณค่าเฉลี่ยของเวลาที่ใช้ในการทำ Image Segmentation ในแต่ละเฟรม ผลการทดลองดังแสดงในตารางที่ 2 พบว่า ResNet50V2 ใช้เวลาในการทำ Image Segmentation เร็วที่สุด

ตารางที่ 2 ผลการทดลองความเร็วของแต่ละโมเดล

Model	เวลาที่ใช้ตามจำนวนครั้ง (วินาที/sec)			
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	ครั้งที่ 4
MobileNetV2	2.46	1.50	1.50	1.81
ResNet50	4.21	4.00	3.79	4.00
ResNet50V2	1.63	1.61	1.59	1.59
DenseNet201	2.46	2.21	2.31	2.32
VGG16	2.41	2.31	2.16	2.29

จากการทำ Image Segmentation จะได้ผลการจำแนกด้วยค่าสี RGB ตามที่แสดงในรูปที่ 12 ซึ่งแสดงผลลัพธ์ของ Image Segmentation จากโมเดลต่าง ๆ ที่ได้ทดลอง



รูปที่ 12 *ผลที่ได้จากโมเดลแต่ละโมเดล*

5. สรุป

จากการทดลอง พบว่า โมเดลแต่ละโมเดลมีผลการจำแนกพื้นที่ที่ความแม่นยำใกล้เคียงกัน โดยจะมีความแม่นยำอยู่ที่ประมาณ 80% อย่างไรก็ตาม เวลาที่ใช้ในการทำงานมีความแตกต่างกันอย่างมีนัยสำคัญ โมเดลที่มีความแม่นยำสูงสุด คือ MobileNetV2 แต่กลับใช้เวลาทำงานเฉลี่ยช้ากว่า ResNetV2 ทั้งนี้ MobileNetV2 แสดงผลการทำงานที่เร็วที่สุดในบางเฟรม ทำให้เป็นโมเดลที่เหมาะสมที่สุดสำหรับการจำแนกพื้นที่จากภาพถ่ายทางอากาศเมื่อเทียบกับโมเดลอื่น ๆ ที่ใช้ในการทดลอง

อย่างไรก็ตาม การทำ Segmentation ยังไม่เหมาะสำหรับการใช้ในอากาศยานไร้คนขับ เนื่องจากความเร็วของกล้องที่ใช้อ้างอิงในงานทดลองที่ผ่านมา [9] แสดงให้เห็นว่าจำเป็นต้องมีการจำแนกพื้นที่ที่เร็วขึ้นมาก เนื่องจากกล้องอากาศยานไร้คนขับมีความเร็วถึง 25 FPS ทำให้การทำ Image Segmentation เหมาะกับภาพถ่ายทางดาวเทียมที่เป็นภาพนิ่งมากกว่า

ผลการวิจัยนี้สามารถนำไปประยุกต์ใช้ในการจำแนกพื้นที่จากภาพถ่ายทางอากาศ หรือในงาน Remote Sensing หรือการทำ Change Detection [14] - [15] ได้หลากหลาย ซึ่งข้อจำกัดที่พบในการวิจัย คือ การประมวลผลภาพที่ใช้ระยะเวลานานในการฝึกและทดสอบข้อมูล แนะนำให้ใช้ GPU (Graphics Processing Unit) ที่มีความเร็วสูงกว่า เพื่อลดระยะเวลาในการปฏิบัติงาน และควรเพิ่มจำนวนภาพต้นฉบับมากกว่า 1,300 ภาพ และใช้วิธีการ Augmentation ที่หลากหลายมากขึ้นในการวิจัยในอนาคต ผู้วิจัยหวังว่างานวิจัยนี้จะเป็นประโยชน์ต่อผู้ที่สนใจและนำไปพัฒนาต่อยอดในอนาคต

6. เอกสารอ้างอิง

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *MICCAI 2015 - 18th Int. Conf. Med. Imag. Comput. Comput. Assisted Intervention*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi , Eds., Munich, Germany, 2015, pp. 234 – 241.

[2] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, and S. L. Waslander, “Temporary Removal: Aerial Imagery for Roof Segmentation: A Large-scale Dataset Towards Automatic Mapping of Buildings,” *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 42 - 55, 2019.

[3] L. - C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834 - 848, 2018.

[4] L. Mezeix and M. G. Casanova, "Dataset Creation Methodology for CNN Land Use/ Cover Classification: Thailand's Rural Area Study Case," *Def. Technol. Acad. J.*, vol. 5, no. 11, pp. 74 - 95, Feb. 2023.

[5] L. Mezeix, C. Arnal, S. Bassanetti, H. Corbin, and V. Mungkung, "Land Cover Analysis for Agricultural Area in Thailand Using CNN Method," *Def. Technol. Acad. J.*, vol. 5, no. 11, pp. 62 - 73, Feb. 2023.

[6] I. Goodfellow *et al.*, "Generative Adversarial Networks," *Comm. ACM*, vol. 63, no. 11, pp. 139 – 144, 2020.

[7] T. Karras, S. Laine, and T. Aila, "A Style-based Generator Architecture for Generative Adversarial Networks," in *2019 IEEE/ CVF Conf. Comput. Vision Pattern Recognit. (CVPR),* Long Beach, CA, USA, 2019, pp. 4396 - 4405, doi: 10.1109/ CVPR.2019.00453.

[8] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *2020 IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2020, pp. 8107 – 8116, doi: 10.1109/ CVPR42600.2020.00813.

[9] W. Luangluewut, K. Viriyasatr, W. Pawgasame, P. Kaewmongkol, and S. Mitaim, "Detecting Objects in Aerial Photographs Using Neural Network Techniques," *Def. Technol. Acad. J.*, vol. 5, no. 12, pp. 4 - 11, Nov. 2023.

[10] C. Sundelius, "Deep Fusion of Imaging Modalities for Semantic Segmentation of Satellite Imagery," M.S. thesis, Dept. Elect. Eng., Linköping Univ., Linköping, Sweden, 2017.

[11] I. Goodfellow *et al.*, "Generative Adversarial Nets," 2014, arXiv:1406.2661v1.

[12] R. Gandhi. "Generative Adversarial Networks- Explained." TOWARDSDATASCIENCE. com. https://towardsdatascience.com/ generative-adversarial-networks -explained-34472718707a (accessed Nov. 2, 2023).

[13] O. Ronneberger, P. Fischer, and T. Brox, "U - Net: Convolutional Networks for Biomedical Image Segmentation," 2015, arXiv:1505.04597.

[14] R. Shao, C. Du, H. Chen, and J. Li, "SUNet: Change Detection for Heterogeneous Remote Sensing Images from Satellite and UAV Using a Dual-Channel Fully Convolution Network," *Remote Sens.*, vol. 13, no. 8, p. 3750, 2021, doi:10.3390/rs13183750.

[15] X. Li, L. Yan, Y. Zhang, and N. Mo, "SDMNet: A Deep - Supervised Dual Discriminative Metric Network for Change Detection in High-Resolution Remote Sensing Images," *IEEE Geosc. Remote Sens. Lett.*, vol. 19, pp. 1 - 5, 2022, Art no. 5513905, doi: 10.1109/LGRS.2022.3216627.