

การตรวจจับวัตถุในภาพถ่ายทางอากาศด้วยเทคนิคโครงข่ายประสาทเทียม

วรากร เลื่องลือวุฒิ^{1*} กิตติกร วิริยะศาสตร์¹ วิชัย แผ้วเกษม¹

พันธุ์เทพ แก้วมงคล¹ และสัญญา มิตรเอม²

วันที่รับ 18 ตุลาคม 2566 วันที่แก้ไข 6 พฤศจิกายน 2566 วันที่ตอบรับ 20 พฤศจิกายน 2566

บทคัดย่อ

บทความนี้กล่าวถึงการศึกษาการตรวจจับภาพถ่ายทางอากาศ ซึ่งเป็นภาพที่ได้จากอากาศยานไร้คนขับ (Unmanned Aerial Vehicle: UAV) มาวิเคราะห์ด้วยหลักการเรียนรู้ของโครงข่ายประสาทเทียมกับภาพถ่ายทางอากาศ โดยการศึกษาผู้เขียนทบทวนวรรณกรรมด้านการวิเคราะห์ภาพเพื่อการตรวจจับวัตถุ และศึกษาค้นคว้าว่ามีโมเดลใดบ้างที่เหมาะสมกับงานด้านการตรวจจับวัตถุ โดยเลือกโมเดลตรวจจับ YOLO, RetinaNet และ Fast R-CNN แล้วทำการศึกษาและทำการทดลองว่าการตรวจจับโดยใช้ภาพถ่ายทางอากาศด้วยอากาศยานไร้คนขับนั้น ควรใช้โมเดลตรวจจับชนิดใดที่สอดคล้องกับอุปกรณ์ที่ใช้ในการรับภาพให้เหมาะสม ซึ่งจากผลการทดลองพบว่า ผลการใช้ YOLO มีผล mAP ถึง 58.5% และความไวในการทำงาน 158.13 เฟรมต่อวินาที ซึ่งมีความแม่นยำและความรวดเร็วในการตรวจจับมากกว่าโมเดลตัวอื่น ๆ ที่ใช้ในการทดลองภาพถ่ายทางอากาศของอากาศยานไร้คนขับ

คำสำคัญ : โครงข่ายประสาทเทียม, การตรวจจับวัตถุ

¹ ส่วนงานวิศวกรรมการสื่อสารข้อมูลทางอิเล็กทรอนิกส์และเครือข่ายคอมพิวเตอร์, สถาบันเทคโนโลยีป้องกันประเทศ

² ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์, คณะวิศวกรรมศาสตร์, มหาวิทยาลัยธรรมศาสตร์

* ผู้แต่ง, อีเมล: warakorn.l@dti.or.th

Detecting Objects in Aerial Photographs using Neural Network Techniques

Warakorn luangluewut ^{1*} Kittakorn Viriyasatr ¹ Wichai Pawgasame ¹
Pantape Kaewmongkol ¹ and Sanya Mitaim ²

Received 18 October 2023, Revised 6 November 2023, Accepted 20 November 2023

Abstract

This article discusses the study of aerial image analysis obtained from Unmanned Aerial Vehicles (UAVs) using the principles of artificial neural networks for image recognition. In this study, the author reviews the literature on image analysis for object detection and explores which models are suitable for the task of object detection in aerial images. The author selects the YOLO, RetinaNet, and Fast R-CNN detection models after careful consideration. Through experimentation and study, the article reveals that when detecting objects in aerial images captured by UAVs, it is crucial to choose a detection model that aligns well with the equipment used for image capture. The experimental results show that utilizing the YOLO model yields a mean Average Precision (mAP) of up to 58.5% and a processing speed of 158.13 frames per second. These results highlight the superior accuracy and speed of object detection compared to other models tested with aerial images captured by unmanned aerial vehicles.

Keywords : Neural networks, Object detection

¹ Data Communication Division, Defence Technology Institute

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

* Corresponding author: warakorn.l@dti.or.th

1. บทนำ

บทความนี้กล่าวถึงการศึกษาภาพถ่ายทางอากาศ โดยการนำภาพที่ได้จากอากาศยานไร้คนขับ (Unmanned Aerial Vehicle: UAV) มาวิเคราะห์ด้วยหลักการเรียนรู้ของโครงข่ายประสาทเทียม (Neural Network) โดยที่ผ่านมามีงานวิจัยที่ได้ทำการทดสอบทดลองด้วย Neural Network กับภาพถ่ายทางอากาศ [1] - [5] ซึ่งมีการนำไปประยุกต์ใช้กับงานที่หลากหลาย แต่ในบทความนี้ผู้เขียนและคณะได้ทำการศึกษางานวิเคราะห์ภาพถ่ายทางอากาศในรูปแบบการตรวจจับวัตถุ (Object Detection) โดยทำการศึกษาโมเดลตรวจจับวัตถุในแบบต่าง ๆ ว่ามีข้อดีและข้อเสียแตกต่างกันอย่างไร โดยนำโมเดล 3 โมเดล ได้แก่ YOLO [6], RetinaNet [7] และ Fast R-CNN [8] มาทำการศึกษาว่าโมเดลแบบใดสอดคล้องกับการทำงานของกล้องที่ใช้ในอากาศยานไร้คนขับ (UAV) เพื่อวัตถุประสงค์ในการตรวจจับวัตถุต่าง ๆ และมีความเหมาะสมในการนำไปประยุกต์ใช้ในการกิจด้านความมั่นคงในประเทศ เช่น ด้านป่าไม้ ด้านการกู้ภัย เป็นต้น และอีกหลาย ๆ งานที่ใช้ภาพถ่ายทางอากาศ

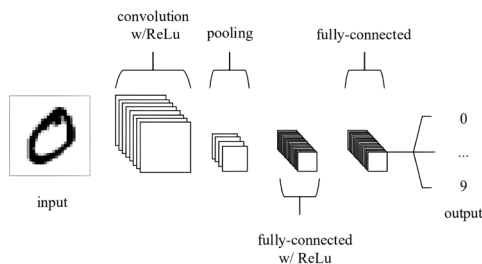
2. ทฤษฎีที่เกี่ยวข้อง

ที่ผ่านมามีงานวิจัยที่ได้ทำการทดลองโครงข่ายประสาทเทียมกับภาพถ่ายทางอากาศไว้หลากหลาย โดยงานวิจัยนี้ผู้เขียนได้ทำการศึกษางานวิเคราะห์ภาพในรูปแบบการตรวจจับวัตถุ ซึ่งจากการศึกษาค้นคว้าพบว่า มีงานวิจัยรูปแบบใกล้เคียงในลักษณะนี้แล้ว ซึ่งทำการศึกษาเพื่อตรวจจับสิ่งก่อสร้างกับยานพาหนะ โดยศึกษาว่ามีโมเดลใดบ้างที่เหมาะสมกับงานของการตรวจจับวัตถุ

2.1 Convolutional Neural Networks (CNNs) Neural Network (โครงข่ายประสาทเทียม) [9]

โครงข่ายประสาทเทียม คือ โมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วยเซลล์ประสาท (Neurons) และจุดประสานประสาท (Synapses) ตามโมเดลนี้ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทจนเป็นเครือข่ายที่ทำงานร่วมกัน ส่วน Deep Learning คือ วิธีการเรียนรู้แบบอัตโนมัติ ด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Neurons) โดยนำระบบโครงข่ายประสาทเทียม (Neural Network) มาซ้อนกันหลายชั้น (Layer) และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูลดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบ (Pattern) หรือจัดหมวดหมู่ข้อมูล (Classify the Data) โดยทั่วไประบบโครงข่ายประสาทเทียมจะเรียนรู้ได้เพียงไม่กี่ชั้น เนื่องจากยังไม่มีข้อมูลสอน (Training Data) หรือความสามารถด้านคอมพิวเตอร์ยังไม่สูงพออย่างใดก็ดี ในช่วงหลายปีที่ผ่านมาเทคโนโลยีได้มีการพัฒนามากขึ้น ทำให้มีการเพิ่มจำนวนชั้นของโครงข่ายได้ง่ายและมากยิ่งขึ้น ยังมีการซ้อนกันหลายชั้นโครงข่ายก็ยิ่งมีความซับซ้อนและลึกขึ้น จึงเป็นที่มาของคำว่า Deep Learning นอกจากนี้ ตามรูปแบบของ Machine Learning โดยทั่วไปเมื่อมีข้อมูลดิบเข้ามาจะไม่มี การสกัดคุณลักษณะของข้อมูลนำเข้า (Feature Extraction) โดยอัตโนมัติ อย่างเช่น Deep Learning แต่จะต้องอาศัยความรู้เฉพาะทาง (Domain Knowledge) สำหรับสกัดคุณลักษณะ (Feature) ด้วยตนเอง โดยตัวอย่าง CNN จะเห็นได้จากรูปที่ 1 แสดงการใช้โมเดล CNN ในการอ่าน

ตัวเลขจากรูปภาพว่าในรูปนั้นเป็นตัวเลขอะไร แล้ว
โมเดลจะทำการทำนายผลออกมาว่าเป็นเลขอะไร

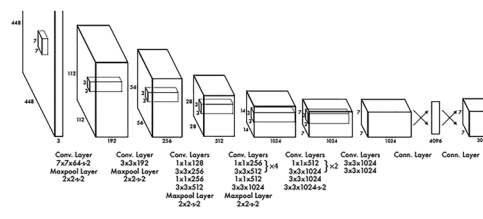


รูปที่ 1 โครงสร้างของ CNNs [9]

2.2 YOLO (You Only Look Once)

ความเร็วในการตรวจจับวัตถุมีความสำคัญอย่างยิ่งในแอปพลิเคชันจำนวนมากที่ต้องการการตัดสินใจแบบเรียลไทม์ ยกตัวอย่างเช่นในขณะที่ยานอากาศยานไร้คนขับ (UAV) อยู่ในอากาศ แอปพลิเคชันการเฝ้าระวังต้องสามารถตรวจจับภาพทางอากาศด้วยความเร็วสูง แบบจำลองการตรวจจับวัตถุต้องสามารถตรวจจับได้เร็วขึ้นตามไปด้วย ซึ่งจะทำให้การตัดสินใจมีความน่าเชื่อถือมากยิ่งขึ้นเท่านั้น Redmon [6] เสนอการตรวจจับวัตถุครั้งเดียวที่ล้ำสมัยด้วยการตรวจจับที่เรียกว่า You Only Look Once (YOLO) ซึ่งมีจุดมุ่งหมายเพื่อแก้ไขปัญหาความเร็วของการตรวจจับ YOLO แบ่งภาพอินพุตลงในกริด แต่ละภาพมีขนาดเท่ากัน CNN เดียวใช้ในการทำนายความน่าจะเป็น คลาส (Class) และขอบเขตกล่องของวัตถุในแต่ละตาราง YOLO ใช้กล่องยึดที่กำหนดไว้ล่วงหน้าในขนาดและอัตราส่วนกว้างยาวที่แตกต่างกันประมาณกรอบขอบเขตของวัตถุในภาพ มีการกำหนดกล่องยึดสำหรับแต่ละกริด ตั้งแต่การเปิดตัว YOLO [6] ได้มีการต่อยอดพื้นฐานออกไปจากสถาปัตยกรรม YOLO มากมาย นั่นคือมีการพัฒนาโมเดลขึ้นมากมาย และแต่ละโมเดลเป็น

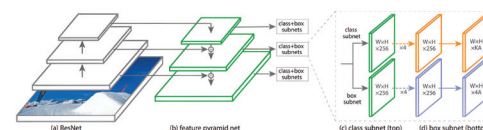
ที่นิยมและถูกนำไปใช้ในแอปพลิเคชันที่หลากหลาย และแตกต่างกันออกไปตามวัตถุประสงค์ โดยโมเดล YOLO จะมีโครงสร้างดังรูปที่ 2



รูปที่ 2 โครงสร้างโมเดล YOLO [6]

2.3 RetinaNet

RetinaNet เป็นโมเดลการตรวจจับวัตถุ (Object Detection) ที่มีรูปแบบเป็นแบบพีระมิด มีการอ่านภาพจากยอดพีระมิดไปยังฐานพีระมิดโดยเริ่มจากการซูมตำแหน่งในภาพแล้วปรับขนาดกรอบตรวจจับภาพไปเรื่อย ๆ จนเจอวัตถุที่ต้องการตรวจจับ ตัวอย่างโมเดล [7] ดังแสดงในรูปที่ 3

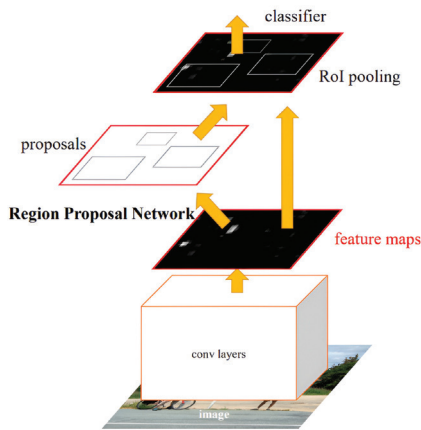


รูปที่ 3 RetinaNet model architecture [7]

2.4 Fast R-CNN

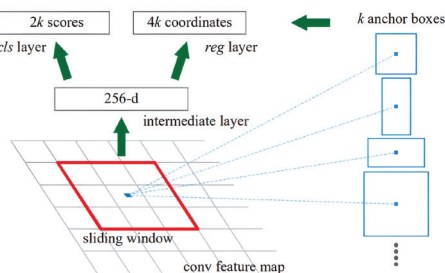
Fast R-CNN เป็นโมเดลตรวจจับวัตถุที่มีรูปแบบเป็นการเลือกกล่อง (Box) ที่จะใช้ในการตรวจจับวัตถุที่ต้องการ โดยปรับขนาดให้เหมาะสมกับข้อมูลที่ต้องการใช้ตรวจสอบ หลังจากนั้นกล่องจะถูกเลื่อนกันไปเรื่อย ๆ ตามขนาดของภาพทั้งหมด หากเจอวัตถุตรงกับคลาสที่ต้องการตรวจจับโมเดลจะทำการสร้างกล่องเพื่อตรวจจับตำแหน่งวัตถุในภาพนั้น [8]

ดังแสดงในรูปที่ 4



รูปที่ 4 แสดงการทำงานของ Fast R-CNN [8]

หลังจากทำการเลื่อนกล่องในภาพแล้ว โมเดล จะทำการ convolutional โดยวิธีการนี้เรียกว่า Region Proposal Network (RPN) โดยที่ RPN จะปรับขนาดกล่องให้เหมาะสมกับข้อมูลที่ตรวจจับ ในภาพ แล้วจะให้ช่วงคะแนนที่ดีที่สุดว่าควรสร้าง กล่องแบบใดมาตรวจจจับภาพ โดยตัวอย่างการทำงาน จะเป็นไปดังแสดงในรูปที่ 5



รูปที่ 5 แสดงการทำงานของ Region Proposal Network [8]

2.5 การตรวจจับวัตถุโดยอาศัยกล่องและตำแหน่ง ของระบบกำหนดตำแหน่งบนอากาศยานไร้คนขับ สำหรับวัตถุประสงค์การค้นหาและช่วยชีวิต [10]

ในงานนี้คือการวิจัยในเรื่องของกล่องที่ใช้ติด ในอากาศยานไร้คนขับ (UAV) ที่ใช้ในการกู้ภัยว่ามี

ประสิทธิภาพในการทำงานเท่าไร มีความเร็วในการ ทำงานมากน้อยเพียงใด และพบว่าความละเอียดของ ภาพมีประสิทธิภาพถึง 1920×1080 จุดภาพ และ มีความไวในการทำงาน (Frame Rate) ถึง 25 เฟรม ต่อวินาที (Frame per Second: FPS) โดยภาพถ่าย ทางอากาศจากกล้องของอากาศยานไร้คนขับแสดง ดังรูปที่ 6



รูปที่ 6 ตัวอย่างภาพถ่ายทางอากาศจากกล้องของอากาศยาน ไร้คนขับ (UAV) [10]

3. วิธีการดำเนินการ

ข้อมูลการทดลองจะเป็นภาพถ่ายทางอากาศ มุมสูงโดยใช้อากาศยานไร้คนขับ (UAV) เป็นตัว ถ่ายภาพ ซึ่งภาพที่ใช้ในการทดลองนี้เป็นภาพที่มี ขนาดภาพคือ 5472×3648 ซึ่งในการทดลองนี้ จะทำการทดลองทั้งหมด 3 โมเดลหลัก คือ YOLOv7, RetinaNet และ Fast R-CNN โดยในการทดลองนี้ จะใช้คลาสการทดลองทั้งหมด 2 คลาส คือ สิ่งก่อสร้าง กับพาหนะ โดยตัวอย่างข้อมูลภาพที่ใช้ทดลองจะ เป็นดังแสดงในรูปที่ 7 ซึ่งจำนวนภาพที่ใช้มีจำนวน 730 ภาพ ประกอบด้วย พาหนะ 3,937 คัน และ สิ่งก่อสร้าง 14,263 หลัง โดยใช้ข้อมูลในการ Train 80% Validation 10% และ Test อีก 10% ของ ภาพทั้งหมด



รูปที่ 7 ตัวอย่างภาพถ่ายทางอากาศจากกล้องของอากาศยานไร้คนขับ (UAV)

ซึ่งผลที่ได้จากการทดลองจะเป็นการที่ Detectors แต่ละตัวสามารถตรวจจับวัตถุแต่ละคลาสโดยการนำโมเดลมาตรวจจับวัตถุในภาพได้โดยใช้ Bounding Boxes ดังแสดงในรูปที่ 8 และสามารถบอกถึงความเร็ว (Speed) ในการทำงาน ความแม่นยำได้ว่าโมเดลแต่ละตัวมีประสิทธิภาพดีเพียงใด และขนาดความจุของโมเดลแต่ละตัวว่าใช้ความจุมากแค่ไหน



รูปที่ 8 ตัวอย่างการใช้ Bounding Boxes มาตรวจจับวัตถุ

4. ผลการทดลอง

model	mAP	FPS	Size (MB)
YOLOv7	58.5%	158.13	36.5
Fast R-CNN	21.2%	73	28.3
RetinaNet	1.2%	152.56	50.7

ตารางที่ 1 ผลการทดลองของภาพถ่ายทางอากาศโดยใช้ Object Detection

จากการทดลองของโมเดลแต่ละตัวพบว่า Frame Per Second (FPS) ทุกตัวมีค่าเกิน 25 เฟรมต่อวินาทีทั้งหมด แต่ประสิทธิภาพของโมเดลยังดีไม่เท่ากันอยู่ และบางโมเดลจะมีค่า mAP ที่ต่ำมากปัจจัยหนึ่งเนื่องจากข้อมูลที่ใช้มีจำนวนต่ำเกินไป แต่ในการทดลองนี้เพียงแค่อ้างอิงเปรียบเทียบข้อดีข้อเสียในงานภาพถ่ายทางอากาศเท่านั้น จึงใช้ผลการทดลองนี้มาเป็นผลการทดสอบ ซึ่งผลการทดลองพบว่า YOLOv7 มีค่าความแม่นยำที่สูงมากเมื่อเทียบกับโมเดลตัวอื่น ๆ แต่ Fast R-CNN กับ RetinaNet มีข้อดีข้อเสียที่แตกต่างกัน เนื่องจาก RetinaNet มี FPS ที่สูงกว่า Fast R-CNN มาก แต่ความแม่นยำของ RetinaNet ต่ำกว่ามาก ในขณะที่ Fast R-CNN มีขนาดความจุต่ำกว่าโมเดลตัวอื่น ๆ มาก โดยใช้ความจุเพียง 28.3 เมกะไบต์

5. การอภิปรายผล

ตารางที่ 2 ข้อดี-ข้อเสียของแต่ละโมเดล

model	ข้อดี	ข้อเสีย
YOLOv7	มีความไวมากที่สุดในการทำงานเมื่อเทียบกับตัวตรวจจับอื่น ๆ	ใช้ปริมาณในความจุโมเดลยังมากอยู่เมื่อเทียบกับโมเดลบางตัว
RetinaNet	มีความไวในการทำงานเกือบจะไวที่สุด	โมเดลมีขนาดความจุที่มากและความแม่นยำต่ำมาก
Fast R-CNN	ใช้ความจุในการเก็บโมเดลต่ำมากเมื่อเทียบกับตัวอื่น	ความเร็วและความแม่นยำยังไม่ดีที่สุดเมื่อเทียบกับ YOLO

เมื่อนำผลการทดลองมาสรุปรวมกันในตารางที่ 2 พบว่า YOLOv7 มีความไวสูงที่สุดเมื่อเทียบกับโมเดลชนิดอื่น ๆ และความไวในการทำงานสอดคล้องกับการทำงานบนกล้องของอากาศยานไร้คนขับ (UAV) ที่มีความไว 25 FPS ซึ่งเหมาะสมต่อการใช้งานของ YOLOv7 ที่มีความไวถึง 158.13 FPS แต่โมเดลตัวอื่น ๆ ถึงแม้จะมีความไวที่สามารถใช้กับอากาศยานไร้คนขับ (UAV) ได้ แต่ความแม่นยำยังต่ำอยู่มากเมื่อเทียบกับ YOLOv7

6. สรุป

- YOLOv7 มีข้อดีด้านความเร็วในการประมวลผลสูงมาก สามารถทำงานได้แบบเรียลไทม์ ใช้งานง่าย สะดวกในการนำไปประยุกต์ใช้งาน รองรับ การตรวจจับวัตถุได้หลากหลายประเภท แต่ความแม่นยำในการตรวจจับวัตถุอาจไม่สูงเมื่อเทียบกับโมเดลตัวอื่น ๆ ที่วัดจาก mAP

- RetinaNet มีข้อดีด้านความเร็วในการประมวลผลสูง ใกล้เคียงกับ YOLOv7 ความแม่นยำในการตรวจจับวัตถุสูง รองรับการทำงานตรวจจับวัตถุได้หลากหลายประเภท แต่การใช้งานในการประมวลผลอาจซับซ้อนกว่า YOLOv7 ทำให้ขนาดโมเดลมีความจุมาก

- Fast R-CNN มีข้อดี คือ ความแม่นยำในการตรวจจับวัตถุสูงและรองรับการทำงานตรวจจับวัตถุได้หลากหลายประเภท แต่มีความเร็วในการประมวลผลต่ำ อาจไม่เหมาะกับการทำงานแบบเรียลไทม์ การใช้งานซับซ้อนน้อยกว่า YOLOv7 และ RetinaNet จากการศึกษาพบว่า โมเดล YOLOv7 เป็นโมเดลที่มีความเร็วเมื่อเทียบกับโมเดลชนิดอื่น ๆ ถึงแม้จะมีความซับซ้อนมากกว่า Fast R-CNN ก็ตาม ซึ่งเป็น

โมเดลที่น่าสนใจสำหรับงานวิจัยและการประยุกต์ใช้กับภาพถ่ายทางอากาศที่ได้มาจากอากาศยานไร้คนขับ (UAV) ขณะเรียลไทม์ให้สอดคล้องกับความเร็วในการทำงานกับกล้องตรวจจับ และในงานทดลองขั้นถัด ๆ ไปในการตรวจจับวัตถุในภาพถ่ายทางอากาศในขณะเคลื่อนที่ ซึ่งจะเป็นประโยชน์ทั้งในการประยุกต์ใช้งานด้านความมั่นคง ตัวอย่างเช่น การนับจำนวนกำลังพล จำนวนคน จำนวนรถ จำนวนเครื่องบิน เป็นต้น อีกทั้งในการกิจการกู้ภัยในพื้นที่ต่าง ๆ ที่ประสบปัญหาในการเข้าถึงพื้นที่นั้น ๆ ได้อย่างยากลำบาก โมเดลนี้มีความเหมาะสมในการนับจำนวนวัตถุในภาพถ่ายทางอากาศ รวมถึงการนำเอาข้อมูลต่าง ๆ ไปใช้ในการตรวจจับวัตถุในภาพถ่ายทางอากาศทั้งที่เป็นภาพถ่ายแบบนิ่งและภาพเคลื่อนไหว

7. เอกสารอ้างอิง

- [1] Q. M. Chung, T. D. Le, T. V. Dang, N. D. Vo, T. V. Nguyen, and K. Nguyen, "Data Augmentation Analysis in Vehicle Detection from Aerial Videos," in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, Ho Chi Minh City, Vietnam, 2020, pp. 1-3, doi: 10.1109/RIVF48685.2020.9140740.
- [2] S. Ali, A. Siddique, H. F. Ates, and B. K. Güntürk, "Improved YOLOv4 for Aerial Object Detection," in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021, pp. 1 - 4, doi: 10.1109/SIU53274.2021.9478027.

- [3] Y. Zuo, J. Yang, Z. Zhu, R. Li, Y. Zhou, and Y. Zheng, "Real-Time Semantic Segmentation of Aerial Videos Based on Bilateral Segmentation Network," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium, 2021, pp. 2763-2766, doi: 10.1109/IGARSS47720.2021.9554952.
- [4] I. Yurchuk, V. Kovdrya, and L. Bilyanska, "Segmentation of Digital Images of Aerial Photography," in *2019 IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, Kiev, Ukraine, 2019, pp. 258-261, doi: 10.1109/APUAVD47061.2019.8943841.
- [5] G. - S. Xia et al., "DOTA: A Large-scale Dataset for Object Detection in Aerial Images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3974 - 3983.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016 pp. 779-788. doi: 10.1109/CVPR.2016.91.
- [7] T. -Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, Eds. 2015, pp. 1 - 9.
- [9] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *ArXiv e-prints*, pp. 1 - 11, 2015, doi: 10.48550/arXiv.1511.08458.
- [10] J. Sun, B. Li, Y. Jiang, and C.-y. Wen, "A Camera-Based Target Detection and Positioning UAV System for Search and Rescue (SAR) Purposes," *Sensors*, vol. 16, no. 11, pp. 1778, 2016, doi: 10.3390/s16111778.