บทความวิจัย

Computer Science Department
Faculty of Informatics, Mahasarakham University

ระบบจำแนกถนนชำรุด

Road Classification System

พิชยุทธ บุญตน, ยุทธนา สีขวา, ระพีพร ช่ำชอง

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

61011212012@msu.ac.th, 61011212060@msu.ac.th, rapeeporn.c@msu.ac.th

บทคัดย่อ

เทคโนโลยีการตรวจจับวัตถุที่สามารถค้นหาสิ่งต่าง ๆ ที่อยู่ในรูปภาพหรือวิดีโอ โดยสามารถสอนโมเดลการตรวจจับให้จดจำแค่บางสิ่งเพื่อใช้ในบางงานที่มีความเฉพาะเจาะจงได้ เทคโนโลยีการตรวจจับวัตถุในปัจจุบันเริ่มมีให้เห็นในอุปกรณ์ทั่ว ๆ ไปเช่น กล้องวงจรปิด มือถือ รถยนต์ เป็นต้น เมื่อเทคโนโลยีการตรวจจับวัตถุสามารถนำไปใช้ได้ในหลาย ๆ งาน ดังนั้นในงานวิจัยนี้จึงใช้โมเดลการเรียนรู้เชิงลึก (Deep learning) ได้เรียนรู้ความเสียหายของถนนโดยใช้โมเดลที่ผ่านการเรียนรู้มาแล้วที่มีพื้นฐานมาจาก CNN (Convolutional Neural Network) ที่มีชื่อว่า VGG-16 และใช้หลักการการตรวจจับวัตถุแบบ Faster R-CNN ที่มีการสร้างโมเดลแยกออกมาอีกหนึ่งขั้นตอนเพื่อทำงานในการคัดเลือกภาพก่อนจะส่งไปทำนายผลจริงชื่อว่า RPN (Region Proposal Network) โดยเรียนรู้และทดสอบจากภาพถนนจริง ๆ ผลการวัดประสิทธิภาพได้ให้ผลที่น่าพึงพอใจและนำโมเดลปรับใช้ในงานหลาย ๆ ด้าน

คำสำคัญ: โครงข่ายประสาทเทียม, การตรวจจับวัตถุ, Faster R-CNN, Region Proposal Network, Convolutional Neural Network

1.บทนำ

ในปัจจุบันมีการใช้รถใช้ถนนเพิ่มขึ้นอย่างต่อเนื่องและมีการใช้รถยนต์หลากหลายประเภทบนท้องถนน เพิ่มความเสี่ยงให้ผู้ขับขี่ที่ใช้รถใช้ถนนอาจก่อให้เกิดอุบัติเหตุและการสูญเสียชีวิตหรือทรัพย์สินเสียหาย ซึ่งอันเนื่องมาจากการขับขี่ของผู้ขับรถด้วยความเร็วและความประมาท การดื่มสุรา ความไม่ใส่ใจหรือแม้กระทั้งการขับรถที่ผิดประเภทและการบรรทุกเกินขนาด ซึ่งเกินกว่ามาตรฐานของถนนที่จะรับน้ำหนักได้ จึงทำให้ถนนเกิดความเสียหายหรือชำรุด ก่อให้เกิดการจราจรที่ช้าลงและมีอัตราการเพิ่มอุบัติเหตุบนท้องถนนมากขึ้น

เนื่องจากในปัจจุบันเทคโนโลยีมีความก้าวหน้าอย่างก้าวกระโดด ทำให้คอมพิวเตอร์มีประสิทธิภาพมากที่จะประมวลผลข้อมูลจำนวนมหาศาลในเวลาไม่นาน จึงทำให้เกิดทฤษฎีที่ชื่อว่า Deep learning (การเรียนรู้เชิงลึก) ซึ่งทำให้คอมพิวเตอร์สามารถเรียนรู้และพัฒนาตัวเองได้คล้ายกับสมองมนุษย์ โดยเรียนรู้จากข้อมูลเป็นจำนวนมาก จึงสามารถคัดแยกข้อมูลได้อย่าง

Computer Science Department
Faculty of Informatics, Mahasarakham University

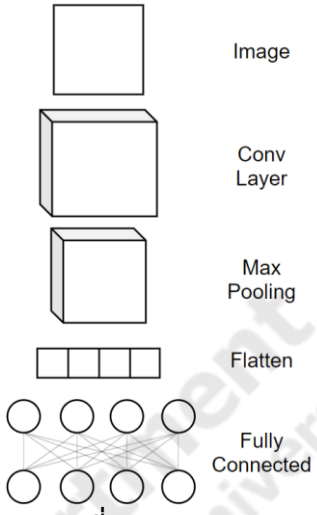
รวดเร็วและแม่นยำเช่น ภาพคน สัตว์ สิ่งของ เป็นต้น

จากปัญหาที่กล่าวมาข้างต้น ทำให้ผู้พัฒนาเล็งเห็นความสำคัญของระบบนี้ จึงได้นำเสนอแอปพลิเคชันที่จะมาช่วยในเรื่องของการแจ้งรายงาน ความชำรุดเสียหายของถนน ทำให้ผู้ใช้ทั่วไปสามารถรายงานสภาพของถนนโดยการถ่ายภาพนิ่ง และส่งผ่านแอปพลิเคชันไปยังหน่วยงานที่เกี่ยวข้องกับการซ่อมบำรุงถนน เพื่อมาซ่อมแซมถนนได้อย่างรวดเร็วในตำแหน่งที่ถูกรายงาน สำหรับการใช้ในหน่วยงานเอง จะติดกล้องรถยนต์ถ่ายภาพถนนแบบวีดีโอ ทำงานควบคู่กับคอมพิวเตอร์เพื่อความเร็วในการประมวลผล โดยให้เจ้าหน้าที่ขับรถไปตามถนนก็จะสามารถตรวจสอบถนนชำรุดได้เช่นกัน ระบบทั้งหมดนี้จะสามารถช่วยคัดแยกประเภทของถนนชำรุดและตำแหน่งของถนนเพื่อมาซ่อมแซมได้อย่างรวดเร็ว

2. ทฤษฎีและงานที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

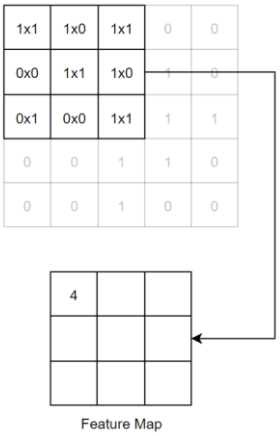
Convolutional Neural Network (CNN) [3] คือโครงข่ายประสาทแบบคอนโวลูชันที่มีประสิทธิภาพในการทำงานเกี่ยวกับรูปภาพมีขั้นตอนสกัดเอาส่วนสำคัญ ๆ ต่าง ๆ ในรูปภาพก่อนจะส่งเข้าไปทำงานในขั้นตอนของโครงข่ายประสาทเทียม (Artificial Neural Network) โดยขั้นตอนของ Convolutional Neural Network มีดังต่อไปนี้



ภาพประกอบที่ 1 การทำงานของ CNN

(1) Convolutional

เป็น Layer หลักของ CNN ทำหน้าที่รับ Input เข้ามา จากนั้นจะใช้การดำเนินการทางคณิตศาสตร์เพื่อหาคุณสมบัติที่สำคัญจากรูปภาพการคำนวณจะเริ่มจากการกำหนดค่าในตัวกรอง (Filter) หรือ เคอร์เนล (Kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออกมาหรือที่เรียกว่า Feature Map



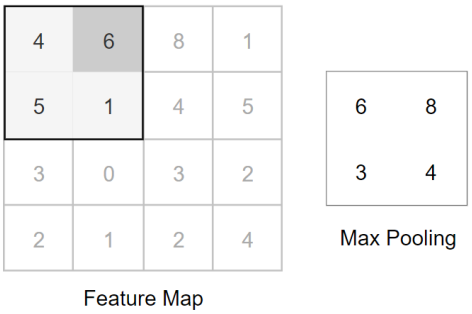
ภาพประกอบที่ 2 Feature Map

(2) Max Pooling

Max Pooling เป็นชั้นที่เชื่อมจาก Convolutional Layer โดยมีเป้าหมายคือทำให้ขนาดของ Feature Map ลดลงด้วยการหาค่าเฉลี่ย (Average Pooling) หรือหาค่าที่สูง

Computer Science Department
Faculty of Informatics, Mahasarakham University

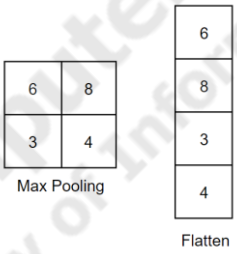
ที่สุด (Max Pooling) และจะเลื่อนตัวกรองไปตาม Stride ที่กำหนดไว้ โดยขนาดตัวกรองของการหาค่าที่สูงที่สุด (Max Pooling) นิยมเรียกกันว่า Pool Size



ภาพประกอบที่ 3 Max Pooling

(3) Flatten

ในขั้นตอนต่อไปจะเป็นการนำผลลัพธ์ทั้งหมดเข้าไปทำงานในขั้นตอน Full connection หน้าที่ของ Flatten เลยต้องแปลงชุดข้อมูลให้เป็นแนวตั้งเพื่อส่งข้อมูลไปยังชั้นของโครงข่ายประสาทเทียม (Artificial Neural Network) ที่เป็นการรับ Input แบบแนวตั้ง



ภาพประกอบที่ 4 Flatten

(4) Full connection

เป็นชั้นของการตัดสินใจ โดยจะมีฟังก์ชันการทำงานที่เรียกว่า Loss Function ที่เป็นตัวให้คะแนนการตัดสินใจหากค่าความน่าจะเป็นเท่ากับ 1 ที่หมวดหมู่ไหน โมเดลนั้นก็จะทำนายว่า Input นั้นเป็นหมวดหมู่ที่ทำนายออกมา

2.1.2 VGG-16

เป็นหนึ่งในโมเดลที่มีพื้นฐานมาจากโครงข่ายประสาทแบบคอนโวลูชันที่มีการแข่งขันด้วยชุดข้อมูล ImageNet และติดอันดับโมเดลห้าอันดับที่ดีที่สุดโดยการทำงานมีโครงสร้างดังนี้

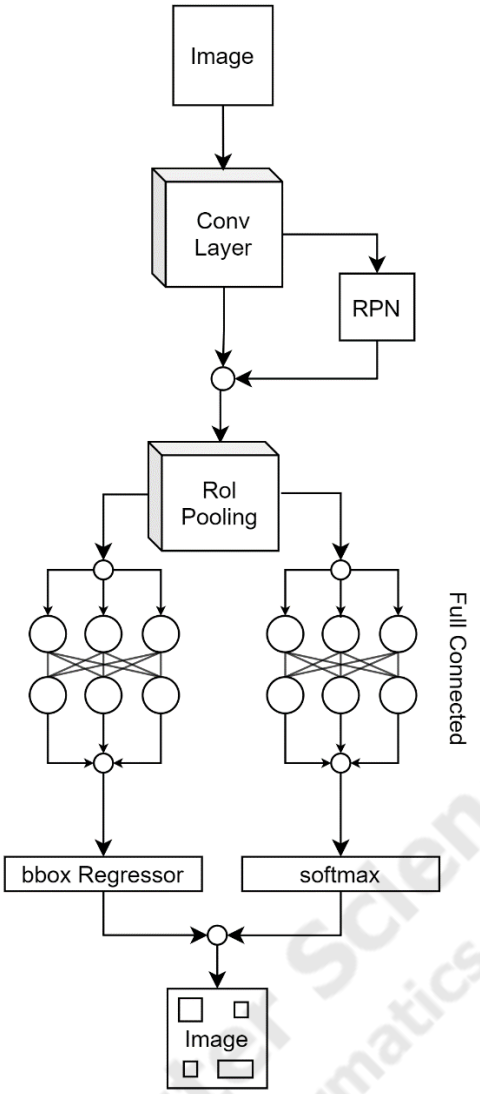
ตารางที่ 1 โครงสร้างของ VGG-16

Input	Layer
1	Convolution 2 ครั้ง
	Max Pooling
3	Convolution 2 ครั้ง
	Max Pooling
5	Convolution 2 ครั้ง
	Max Pooling
7	Convolution 3 ครั้ง
	Max Pooling
10	Convolution 3 ครั้ง
	Max Pooling
13	Full connection
14	Full connection
15	Full connection
Output	Full connection

2.1.2 Faster R-CNN

Faster Regional-Convolutional Neural Networks พัฒนาต่อจาก CNN, R-CNN, Fast R-CNN [6] โดยมีขั้นตอนดังนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 5 Faster R-CNN

หลักการของโมเดล Faster R-CNN จะส่งภาพนำเข้า (Input Image) ไป [2] ประมวลผลด้วย Convolutional Layer ของ VGG-16 ก่อนแล้วจะได้ผลลัพธ์เป็นภาพที่มีเพียงคุณลักษณะเด่น (Feature Map) เท่านั้นก่อนส่งเข้าไปคัดเหลือคำตอบ

(1) Conv Layer

เป็นการ Convolutional ที่ทำการสกัดภาพให้เหลือเพียงแค่คุณลักษณะเด่นแบบเดียวกันกับโครงข่ายประสาทแบบคอนโวลูชัน

(2) RPN

Region proposal network (RPN) รับภาพที่เหลือเพียงคุณลักษณะเด่น (Feature Map) เพื่อทำหน้าที่ในการเลือกส่วนที่คิดว่าน่าจะเป็นวัตถุที่สนใจออกมาด้วยความน่าจะเป็น

(3) ROI Pooling

ทำหน้าที่รวมผลลัพธ์ระหว่าง Conv Layer และ RPN เข้าด้วยกัน จากนั้นจะสกัดส่วนที่สำคัญที่สุดในผลลัพธ์ทุกค่าที่คัดเลือกมาคล้ายกับขั้นตอน Max Pooling ของโครงข่ายประสาทแบบคอนโวลูชัน

(4) Full connection

เป็นชั้นของการตัดสินใจและแยกออกเป็นสองส่วน

- B-box Regressor ทำหน้าที่ในการเรียนรู้และปรับค่าที่ผิดพลาดด้วยผลเฉลย (Ground Truth)

- Softmax ทำหน้าที่เลือกคำตอบที่มีอยู่ใน Output และทำนายผลนั้นออกมา

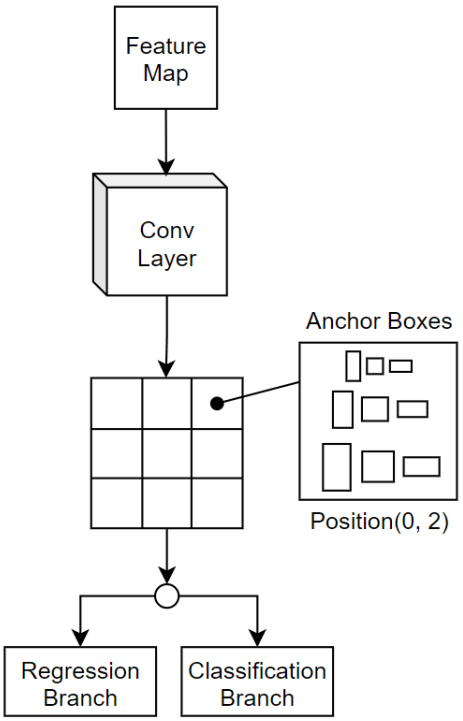
(5) Output

ผลลัพธ์ที่ได้จะมีทั้งค่า กรอบบอกตำแหน่งของภาพและค่าความมั่นใจ กรอบที่ทำนายออกมาจะเป็นกรอบที่จะวางทับกับรูปภาพเพื่อบ่งบอกว่ามีสิ่งที่สนใจอยู่ในตำแหน่งนั้น ๆ

2.1.3 Region proposal network

การทำงานของ RPN ถือว่าเป็นโมเดลขนาดเล็ก ๆ ที่ฝังอยู่ในขั้นตอนของ Faster R-CNN มีขั้นตอนการทำงานของ Convolutional อยู่ข้างในและมีการเรียนรู้และปรับค่าด้วยตัวเองขั้นตอนของ RPN แสดงดังรูปต่อไปนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 6 Region proposal network

(1) Anchor Boxes

มีกล่องที่อยู่ข้างในอีกเก้ากล่อง ทำหน้าที่คัดเลือกวัตถุที่อาจจะมีสิ่งที่ซ้อนอยู่ในภาพ กรณีที่การคัดเลือกตำแหน่งที่คาดว่าเป็นวัตถุ Anchor Boxes จะถูกคำนวณทั้งเก้ากล่องในครั้งเดียวเป็นเพราะว่าบางครั้งวัตถุก็อาจจะออกนอกกรอบเลยต้องสร้าง Anchor Boxes ที่มีขนาดที่แตกต่างกัน

(2) Regression Branch

เรียนรู้การชดเชย (หรือความแตกต่าง) สำหรับค่า X, Y, W, H ที่โมเดลทำนาย (Predicted Box) ออกมาแล้วนำค่ามาเทียบกับผลเฉลย (Ground Truth) เพื่อปรับค่า Anchor Boxes ให้มีการวางตำแหน่งให้ดีขึ้นกว่าเดิม

(3) Classification Branch

คำนวณคะแนน IoU ของผลเฉลย (Ground Truth) ด้วยกล่อง Anchor Boxes และจัดประเภทกล่อง Anchor Boxes ที่เป็นวัตถุหรือเป็นพื้นหลังด้วยความน่าจะเป็น

2.1.2 การประเมินประสิทธิภาพ

การวัดประสิทธิภาพการทำนายของโปรแกรม เปรียบเทียบกับผลลัพธ์จริง ๆ ที่ถูกต้องโดยมีความหมายแต่ละตัวดังนี้

(1) True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริง และคนบอกว่ามันจริง (IoU มากกว่าหรือเท่ากับเกณฑ์ที่กำหนด)

(2) True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง และคนบอกว่ามันไม่จริง (ไม่สามารถใช้ในงานตรวจจับวัตถุได้ เพราะกรอบที่ตรวจจับไม่ได้ถือเป็นเรื่องปกติในงานตรวจจับวัตถุ)

(3) False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริง แต่คนบอกว่าไม่จริง (IoU น้อยกว่าหรือเท่ากับเกณฑ์ที่กำหนด แต่ IoU ต้องมากกว่าศูนย์)

(4) False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง แต่คนบอกว่าจริง (การทำนายออกนอกค่าความจริง Ground Truth จนไม่สามารถคำนวณหา IoU ได้)

(5) Precision and Recall

การหาค่าระลึก (Recall) คือ ค่าที่บอกว่าโปรแกรมทำนายได้จริง เป็นอัตราส่วนเท่าใดของค่าจริงทั้งหมด

$$Recall = \frac{TP}{(TP + FN)} = \frac{TP}{All\ Ground\ Truth}$$

Computer Science Department Faculty of Informatics, Mahasarakham University

ค่าความแม่นยำ (Precision) คือ ค่าที่โปรแกรมทำนายว่าจริงถูกต้อง

$$Precision = \frac{TP}{(TP + FP)} = \frac{TP}{All\ Predict}$$

(6) Average Precision

ความแม่นยำเฉลี่ย (Average Precision) ใช้วัดความแม่นยำของโมเดลตรวจจับวัตถุโดยใช้ค่า Precision และ Recall และเรียงความมั่นใจที่การทำนายสูงสุดตามลำดับ เมื่อกำหนดให้ IoU มากกว่าหรือเท่ากับ 0.5 จากนั้นปรับค่า Precision ด้วยฟังก์ชัน Interpolated Precision เพื่อหาพื้นที่ใต้กราฟ (AUC) แล้วจะแบ่งกราฟออกเป็นส่วน ๆ เพื่อคำนวณความแม่นยำเฉลี่ย (Average Precision) ในขั้นตอนสุดท้ายดังนี้

$$P_{interp}(r) = \max_{\tilde{r}:\tilde{r} \geq r} p(\tilde{r})$$

โดยที่ p คือ Precision

$\tilde{r}$ คือ Recall

การแบ่งกราฟออกเป็นส่วน ๆ เพื่อหาพื้นที่ใต้กราฟ (AUC) ทำได้ดังนี้

$$AP = \sum (r_{n+1} - r_n)\, p_{interp}(r_{n+1})$$

โดยที่ผลลัพธ์ของ AP คือความแม่นยำเฉลี่ย (Average Precision) ของการทำนายที่ถูก (TP) และผิด (FP) หรือไม่ถูกทำนาย (FN)

2.2 งานวิจัยที่เกี่ยวข้อง

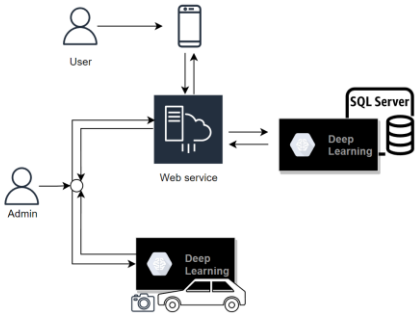
จากงานวิจัยของ Zhun Fan, Senior Member, IEEE, Yuming Wu, Jiewei Lu, and Wenji Li เรื่อง “Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network” [3] งานวิจัยนี้ต้องการตรวจจับรอยแตกบนพื้นผิวทางอัตโนมัติ ได้รับการวิจัยมาหลายสิบปี เนื่องจากพื้นผิวทางที่ซับซ้อนในโลกความจริง ในบทความนี้ใช้วิธีการตรวจสอบที่อยู่บนพื้นฐานการเรียนรู้ลึก โดยเฉพาะ Convolutional Neural Network (CNN) ใช้เรียนรู้โครงสร้างของรอยแตกจากภาพดิบโดยไม่ต้องประมวลผลล่วงหน้า หรือ non-pre-processing

จากงานวิจัยของ Jonathan Long, Evan Shelhamer UC Berkeley, Trevor Darrell เรื่อง “Fully Convolutional Networks for Semantic Segmentation” [4] งานวิจัยนี้ต้องการปรับปรุงแก้ไขเครือข่ายการจำแนกประเภท (AlexNet , VGG net และ GoogLeNet) ลงในเครือข่ายแบบ Fully-convolutional และทำการปรับปรุงแก้ไขผลลัพธ์ที่ได้ใช้กับงาน Segmentation

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.วิธีกำเนินงานวิจัย

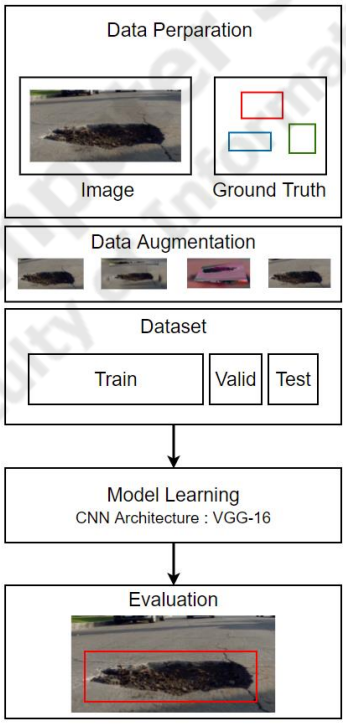
3.1 ภาพรวมระบบ



ภาพประกอบที่ 7 ภาพรวมระบบ

แผนผังการทำงานของระบบโดยที่มีผู้ใช้เป็นคนที่ถ่ายภาพส่งมาเผื่อรายงานถนนและแอดมินมีหน้าที่ถ่ายภาพถนนแล้วส่งมาที่ระบบ ในส่วนของภาพที่ถูกส่งมาจากผู้ใช้จะถูกทำนายผลด้วย Deep Learning ทุกภาพและแอดมินสามรถใช้โปรแกรมเพื่อทำนายผลได้ทันทีแล้วส่งภาพและผลทำนายมาที่ระบบ

3.2 การดำเนินการระบบตรวจจับถนน



ภาพประกอบที่ 8 ระบบตรวจจับถนน

3.2.1 ระยะการรวบรวมข้อมูล

นการเก็บข้อมูลในการวิจัยนี้ ได้ทำการเก็บข้อมูลจากเว็บไซต์ต่าง ๆ จาก Google Search, Google Maps และ Kaggle และยังมีชุดข้อมูลที่มีการแจกให้ใช้งานดังนี้

Annotated Potholes Image Dataset ข้อมูลโดย Atikur Rahman Chitholian จากเว็บ Kaggle เป็นชุดข้อมูลหลุมพร้อมกับ Ground Truth จำนวน 665 ภาพ

PID-Pavement-Image-Dataset ข้อมูลโดย hanshenChen จาก Github เป็นชุดข้อมูลที่ใช้ Google API ดึงภาพออกมาจาก Google maps มีชนิดถนนแตกพร้อมกับ Ground Truth จำนวน 7,237 ภาพ

Road Pothole Images for Pothole detection ข้อมูลโดย Sovit Ranjan Rath จาก Kaggle เป็นข้อมูลที่ใช้กล้องติดรถยนต์เก็บภาพที่เป็นหลุมและทำ Ground Truth เก็บเป็นไฟล์ CSV มีข้อมูลหลุมจำนวน 1,119 ภาพ

โดยข้อมูลทั้งหมดที่เก็บมาได้มีประเภทของหลุม 3 ชนิด หลุม, ถนนแตก, ถนนซ่อมปะ แบ่งเป็นจำนวนชนิดละ 1,990, 8,048, 1,088 ตามลำดับ รวมทั้งสิ้น 11,126 ภาพ

ตารางที่ 2 ชุดข้อมูลถนนเสียหายจำนวน 3 ชนิด

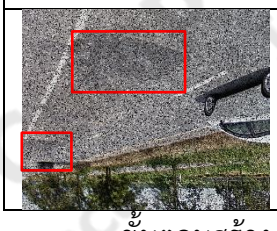
รหัส	ชนิด	จำนวนภาพ
1	หลุม	1,990
2	ถนนแตก	8,048
3	ถนนซ่อมปะ	1,088
รวม		11,126

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.2.2 Data Augmentation

การขยายชุดข้อมูลภาพให้มีความหลากหลายจะช่วยให้มีการสร้างข้อมูลเพิ่มจากเดิมได้มากกว่าสองเท่า และช่วยให้โมเดลได้เรียนรู้ข้อมูลหลาย ๆ แบบ [1] โดยการทำ Data Augmentation ได้ใช้ Source code โปรแกรม Bounding Box Augmentation ที่สร้างโดย mukopikmin และ tonybjorkman จาก Github สามารถสุ่มสร้างภาพใหม่โดยมีการคำนวณ Ground Truth ให้ใหม่โดยอัตโนมัติ โดยมีฟังก์ชันให้เลือกในการสุ่มสร้างภาพใหม่มากมายเช่น เปลี่ยนสี ใส่ข้อมูลรบกวน บิดภาพ เบลอ เป็นต้น แต่งานวิจัยเลือกที่จะใช้วิธีตามตารางดังนี้

ตารางที่ 3 Data Augmentation

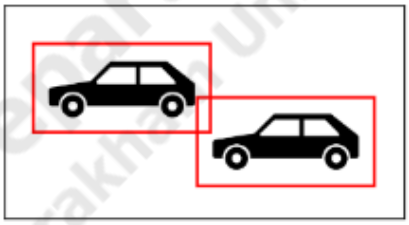
Affine	Blur
Arithmetic	**Color**
	

ขั้นตอนสร้างภาพใหม่จะสร้างเฉพาะข้อมูลที่มีน้อยมากเกินไปเช่น หลุมและถนนซ่อมปะ ที่มีข้อมูลน้อยกว่าถนนแตก จึงสร้างภาพใหม่แค่หลุมและถนนซ่อมปะ กำหนดจำนวนดังนี้

(1) หลุม 1 ภาพ กำหนดให้สร้างภาพใหม่เป็น 3 ภาพ ทำให้ได้ข้อมูลเพิ่มเป็น 5,970 ภาพ

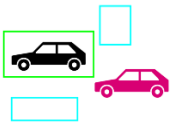
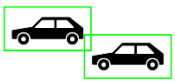
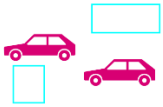
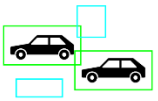
(2) ถนนซ่อมปะ 1 ภาพ กำหนดให้สร้างภาพใหม่เป็น 7 ภาพ ทำให้ได้ข้อมูลเพิ่มเป็น 7,616 ภาพ

ตัวอย่างการคำนวณหา Precision และ Recall ของการตรวจจับวัตถุ กำหนดค่าความจริง (Ground Truth) ดังรูปต่อไปนี้



ภาพประกอบที่ 9 กำหนดค่าความจริง

ตารางที่ 4 ตัวอย่างการ Predict

	ตัวอย่างการ Predict	Precision $\frac{TP}{(TP+FP)}$	Recall $\frac{TP}{(TP+FN)}$
A		$\frac{1}{1+2}=0.33$	$\frac{1}{1+1}=0.5$
B		$\frac{2}{2+0}=1$	$\frac{2}{2+0}=1$
C		$\frac{0}{0+2}=0$	$\frac{0}{0+2}=0$
D		$\frac{2}{2+2}=0.5$	$\frac{2}{2+0}=1$
E		$\frac{1}{1+0}=1$	$\frac{1}{1+1}=0.5$

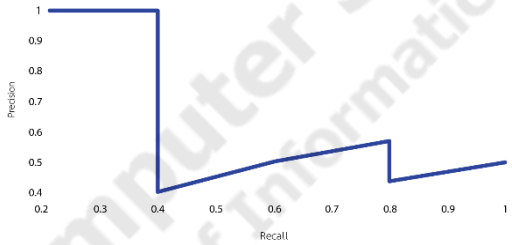
การหาค่า Precision และ Recall คำนวณด้วยตารางทีละขั้นตอนตามจำนวนที่

ทำนายออกมาเรียงความมั่นใจ (Confident) จากมากไปหาน้อย ในระหว่างนั้นก็มีการนับค่าที่ทำนายถูกและผิดและมีการหาค่า Precision และ Recall ไปพร้อม ๆ กัน

ตารางที่ 5 วัดความแม่นยำ

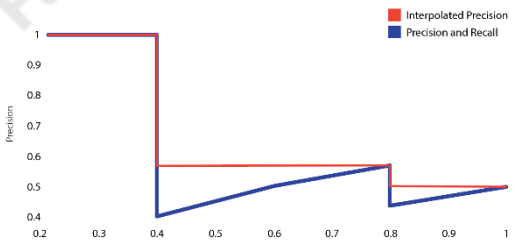
Confidence	TP	FP	ALL TP	ALL FP	Precision	Recall
99%	1	0	1	0	1.0	0.2
98%	1	0	2	0	1.0	0.4
95%	0	1	2	1	0.67	0.4
87%	0	1	2	2	0.5	0.4
75%	0	1	2	3	0.4	0.4
60%	1	0	3	3	0.5	0.6
55%	1	0	4	3	0.57	0.8
45%	0	1	4	4	0.5	0.8
33%	0	1	4	5	0.44	0.8
30%	1	0	5	0	0.5	1.0

เมื่อนำค่า Precision และ Recall ไปสร้างเป็นกราฟจะได้ตามรูปภาพดังนี้



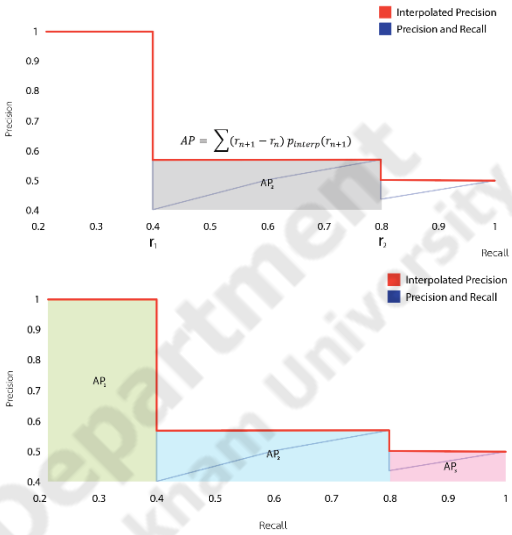
ภาพประกอบที่ 10 Precision และ Recall

จัดเรียงค่า Precision ใหม่ด้วยฟังก์ชัน Interpolated Precision



ภาพประกอบที่ 11 Interpolated Precision

เลือกเฉพาะเส้น Precision ที่มีค่าเท่ากันและแบ่งกราฟออกเป็นส่วน ๆ เพื่อหาพื้นที่ใต้กราฟ (AUC) ทำได้ดังนี้



ภาพประกอบที่ 12 พื้นที่ใต้กราฟ (AUC)

เมื่อนำค่า A1, A2, A3 มารวมกันคำตอบจะเป็นค่าความแม่นยำเฉลี่ย (Average Precision) ดังนั้น AP = 0.5664 หรือ AP = 56%

4.วิธีการทดลอง

จากชุดข้อมูลที่ผ่านการ Data Augmentation ทำให้ได้ข้อมูลใหม่เพิ่มขึ้นตามตารางต่อไปนี้

ตารางที่ 6 ชุดข้อมูลจำนวนทั้งหมด

รหัส	ชนิด	จำนวนภาพ
1	หลุม	7,960
2	ถนนแตก	8,048
3	ถนนซ่อมปะ	8,704
รวม		24,712

ทำการแบ่งข้อมูลออกเพื่อใช้ตรวจสอบความถูกต้องระหว่างเรียนรู้ (Training) และใช้

Computer Science Department
Faculty of Informatics, Mahasarakham University

ทดสอบเพื่อวัดประสิทธิภาพแบ่งออกได้เป็นสามส่วนตามตารางดังต่อไปนี้

ตารางที่ 7 การแบ่งข้อมูล

จำนวนข้อมูลสำหรับเรียนรู้ (90%)		**จำนวนข้อมูลสำหรับวัดประสิทธิภาพ (10%)**		
22,241		2,471		
เรียนรู้ (90%)	ตรวจสอบ (10%)	หลุม	ถนนแตก	ถนนซ่อมปะ
20,016	2,225	824	824	823
ข้อมูลทั้งหมด 24,712 ภาพ				

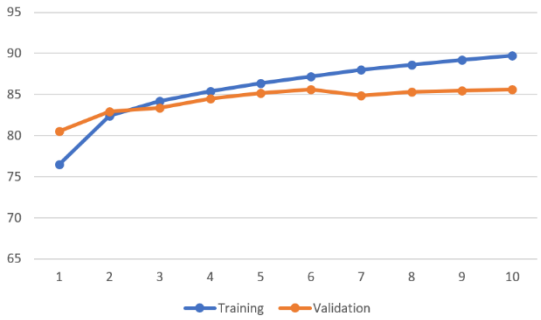
4.1 ตั้งค่าการเรียนรู้

ตารางที่ 8 การตั้งค่าการฝึกการเรียนรู้

Function	Description
Epoch กำหนดรอบการเรียนรู้	10
Train RPN ฝึกการเรียนรู้ โมเดล RPN	True
Train final classifier ฝึกการเรียนรู้ โมเดล VGG-16 ส่วนที่ทำการคัดเลือกคำตอบ (Classification)	True
Train base NN ฝึกการเรียนรู้ โมเดล VGG-16 ส่วนที่หาคุณลักษณะเด่นของรูปภาพ (Feature Extraction)	True
Anchor box scales กำหนดขนาดของ Anchor box ทั้งสามขนาด	128, 256, 512
Anchor box ratios กำหนดสัดส่วนของ Anchor box ทั้งสามขนาด	[1 : 1], [0.7 : 1.4], [1.4 : 0.7]
Image size กำหนดขนาดภาพที่จะทำการฝึกทั้งหมด	600
Optimizer	Adam, Learning rate 0.00001
Augment สุ่มสร้างรูปภาพใหม่	False
Model training APIs ฟังก์ชันการฝึกการเรียนรู้ (Training)	train_on_batch

4.2 ผลการทดลอง

การฝึกโมเดล 10 รอบ ด้วยข้อมูลสำหรับเรียนรู้ 80 เปอร์เซ็นต์จากข้อมูลทั้งหมด และข้อมูลตรวจสอบความถูกต้องอีก 10 เปอร์เซ็นต์จากข้อมูลทั้งหมด ซึ่งการทดสอบจากข้อมูลตรวจสอบความถูกต้องได้ผลลัพธ์ที่ 80 ถึง 85 เปอร์เซ็นต์ แสดงผลการฝึกทั้ง 10 รอบด้วยรูปต่อไปนี้



ภาพประกอบที่ 13 ผลลัพธ์การฝึกการเรียนรู้

4.3 ประเมินและวิเคราะห์ผล

การประเมินผลวัดจากข้อมูลสำหรับวัดประสิทธิภาพ 10 เปอร์เซ็นต์จากข้อมูลทั้งหมด และมีจำนวนข้อมูลที่เท่ากัน มีเพียงถนนซ่อมปะที่มีจำนวน 823 ภาพ ผู้วิจัยจึงเพิ่มข้อมูลหนึ่งภาพเข้าไปด้วยตัวเองเพื่อให้จำนวนที่เท่ากันทั้งสามชนิด จากนั้นนำข้อมูลสำหรับวัดประสิทธิภาพแต่ละชนิด หาค่าความแม่นยำ

Computer Science Department Faculty of Informatics, Mahasarakham University

(Precision), ค่าความระลึก (Recall) ด้วย IoU ที่ได้ 0.5 ขึ้นไปและถือเป็นค่าวัดที่อยู่ในระดับกลางที่ใช้กันทั่วไปในงานตรวจจับวัตถุ เมื่อได้ผลลัพธ์แล้วนำค่าทั้งสองหาความแม่นยำเฉลี่ย (Average Precision)

ผลลัพธ์การตรวจจับในแต่ละหมวดหมู่จะแสดงรายละเอียดการตรวจจับทั้งหมดและจำนวนผลเฉลย (Ground Truth) ดังนี้

ตารางที่ 9 รายละเอียดการตรวจจับของหลุม

ผลการตรวจจับของ หลุม	
จำนวนที่พบวัตถุ	5,525
จำนวนผลเฉลย	2,154
ทำนายถูก	740
ทำนายผิด	4,785
ความแม่นยำเฉลี่ย	0.23

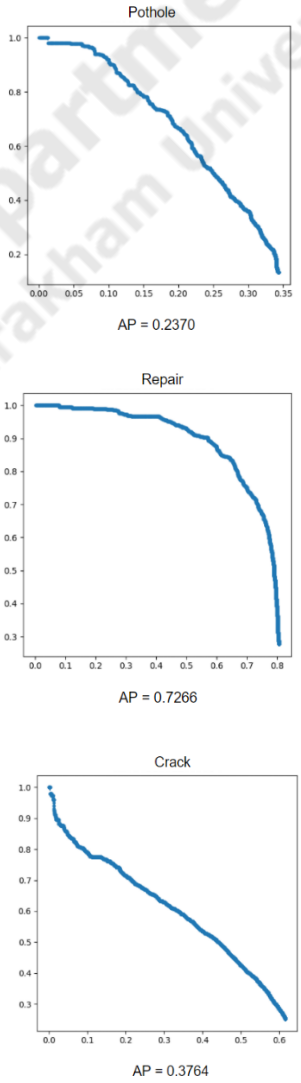
ตารางที่ 10 รายละเอียดการตรวจจับของซ่อมปะ

ผลการตรวจจับของ ซ่อมปะ	
จำนวนที่พบวัตถุ	4,590
จำนวนผลเฉลย	1,580
ทำนายถูก	1,273
ทำนายผิด	3,317
ความแม่นยำเฉลี่ย	0.72

ตารางที่ 11 รายละเอียดการตรวจจับของแตกร้าว

ผลการตรวจจับของ แตกร้าว	
จำนวนที่พบวัตถุ	17,467
จำนวนผลเฉลย	7,127
ทำนายถูก	4,389
ทำนายผิด	13,078
ความแม่นยำเฉลี่ย	0.37

และแสดงผลลัพธ์ทั้งสามชนิดด้วยความแม่นยำเฉลี่ย (Average Precision) ดังต่อไปนี้



ภาพประกอบที่ 14 ผลลัพธ์ทั้งสามชนิด

การประเมินผลจาก หลุมบ่อ ถนนซ่อมปะ ถนนแตก ได้ผลลัพธ์ตามลำดับดังนี้ 0.2370, 0.7266, 0.3764 เมื่อนำค่าทั้งสามชนิดมา

Computer Science Department
Faculty of Informatics, Mahasarakham University

รวมกันและหาค่าเฉลี่ยทั้งหมดจะได้ค่า mAP (mean Average Precision ดังนั้น mAP เท่ากับ 0.4466 หรือ 44 เปอร์เซ็นต์ ด้วย IoU ที่ 0.5

เมื่อสรุปผลจากกราฟของแต่ละหมวดหมู่ ถนนซ่อมปะ มีกราฟที่สูงกว่าหมวดอื่น ๆ และมี Recall สูงที่สุด แปลว่าถนนซ่อมปะมีการตรวจจับวัตถุที่มีอยู่ในผลเฉลย (Ground Truth) ได้มากกว่าหมวดอื่น ๆ อาจเป็นเพราะรูปร่างของถนนซ่อมปะมีลักษณะที่มองออกได้ง่ายและมีรูปทรงที่คล้าย ๆ กัน เมื่อเทียบกับ หลุมและถนนแตก รูปร่างของวัตถุจะมีความไม่แน่นอนอาจทำให้การตรวจจับได้ยาก และเมื่อดูจากกราฟ หลุมและถนนแตก จะมีการตรวจจับวัตถุที่มีอยู่ในผลเฉลย (Ground Truth) ได้น้อยแต่มีความสม่ำเสมอของกราฟกว่าถนนซ่อมปะ ความสม่ำเสมอนี้เมื่อดูของถนนซ่อมปะจะเห็นได้ว่ามีการดิ่งลงของ Precision ตั้งแต่ช่วง 0.7 เป็นต้นไป แปลว่าถนนซ่อมปะจะตอบผิดมากกว่าปกติถ้าความมั่นใจ ต่ำกว่า 70%

ค่าเฉลี่ย mAP (mean Average Precision) เท่ากับ 0.4466 หรือ 44 เปอร์เซ็นต์ ที่ได้ออกมาไม่สามารถบอกได้ว่าโมเดลจะทำนายผลแต่ละหมวดหมู่ได้มากน้อยเพียงใด แต่เป็นการบอกค่าโดยรวมที่โมเดลสามารถทำนายผลออกมาได้ เมื่อเทียบผลลัพธ์กับการตรวจจับวัตถุด้วยวิธี Yolo แสดงตารางดังต่อไปนี้

ตารางที่ 12 ประสิทธิภาพของโมเดล Yolo

Model	Dataset	Class	IoU	mAP
YOLOv3-608	Coco Dataset	90	0.5	57.9
YOLOv2	Coco Dataset	90	0.5	48.1
Tiny YOLO	Coco Dataset	90	0.5	23.7

ประสิทธิภาพโดยรวมของโมเดล Yolo อยู่ระหว่าง 20 ถึง 60 แต่วัดด้วยจำนวนหมวดหมู่ที่มากกว่าและค่า mAP (mean Average Precision) ที่ได้ไม่สามารถบอกได้ว่าหมวดหมู่ไหนทำนายออกมาได้ดีที่สุด โดยผลลัพธ์ของโมเดลจำแนกถนนได้ mAP เท่ากับ 44 เปอร์เซ็นต์ถือว่าอยู่ในเกณฑ์ที่ยอมรับได้

แต่เมื่อดูในระดับค่า AP (Average Precision) ถนนซ่อมปะมีค่าเฉลี่ยออกมาสูงที่สุดที่ 72 เปอร์เซ็นต์ โดยหลุมและถนนแตกมีเปอร์เซ็นต์ที่ต่ำกว่ามาก การนำไปใช้อาจเกิดปัญหาการการตรวจจับวัตถุไม่พบได้มากกว่า

ตาราง 13 Model จำแนกถนน

	หลุม	**ถนนซ่อมปะ**	**ถนนแตก**
AP	23%	72%	37%
ผลลัพธ์	ต่ำ	ปกติ	ต่ำ

Computer Science Department
Faculty of Informatics, Mahasarakham University

5.สรุปและอภิปรายผล

งานวิจัยเล่มนี้ได้กล่าวถึงการตรวจจับความเสียหายบนถนนเช่น หลุม, ถนนแตก, ถนนซ่อมปะ โดยใช้การเรียนรู้เชิงลึกด้วยวิธี Faster R-CNN ซึ่งสามารถใช้งานการตรวจจับได้ผ่าน Mobile Application และ Desktop Application

จากการทดลองวัดประสิทธิภาพการตรวจจับความเสียหายบนถนนรวมทั้งสิ้น 2,471 ภาพ ผลลัพธ์ความถูกต้องของแต่ละหมวดหมู่ หลุม, ถนนแตก, ถนนซ่อมปะ ได้ผลลัพธ์ตามลำดับดังนี้ 0.2370, 0.7266, 0.3764 และความถูกต้องเฉลี่ยอยู่ที่ 44 เปอร์เซ็นต์

ชุดข้อมูลทั้งหมดที่ใช้ในการเรียนรู้มีทั้งผู้วิจัยได้เก็บจาก Google maps และดาวน์โหลดจากชุดข้อมูลที่มีการแจกให้ใช้ฟรีเมื่อสรุปผลจากการเรียนรู้ที่ได้ผลการประเมินของหลุมน้อยที่สุด เพราะชุดข้อมูลของหลุมมีความหลากหลายน้อยกว่าหมวดอื่น ๆ และมีลักษณะของถนนประเทศไทยน้อย เมื่อเทียบกับชุดข้อมูลของถนนซ่อมปะผู้วิจัยได้เก็บภาพถนนซ่อมปะทั้งหมดใน Google maps ประเทศไทย

จากการใช้งานจริงด้วยภาพที่ถ่ายด้วยกล้องติดรถยนต์ได้ผลลัพธ์ที่ไม่ดีมากนักอาจเป็นเพราะกล้องรถยนต์มีเลนส์ที่โค้งสูงมากและกว้างกว่าปกติเมื่อถ่ายภาพก็ทำให้วัตถุอยู่ไกลเกินไปกว่าที่จะมองเห็นรายละเอียด กรณีที่ใช้กล้องมือถือถ่ายจะได้ผลลัพธ์ที่ดีกว่าเพราะมือถือมีระบบอัตโนมัติที่ปรับภาพออกมาได้ดีมากกว่าและเลนส์ของกล้องมือถือจะเป็นเลนส์ไม่กว้างไม่แคบจนเกินไปจึงทำให้วัตถุที่สนใจอยู่ในระดับที่มองเห็นได้ดีและยังคงมีรายละเอียดที่คมชัดมากกว่า

5.1 ปัญหาและอุปสรรค

5.1.1 อัลกอริทึมการตรวจจับ Faster R-CNN มีความซับซ้อนสูงมากจำเป็นต้องใช้ทรัพยากรในการประมวลผลสูงและใช้ระยะเวลานาน

5.1.2 ชุดข้อมูลที่ใช้ในการเรียนรู้มีไม่เพียงพอและไม่มีความหลากหลาย ชุดข้อมูลที่ถ่ายด้วยกล้องติดรถยนต์มีน้อยเกินไป ผลลัพธ์ที่ได้จึงตรวจจับได้ไม่ดีพอ

5.1.3 อัลกอริทึมการตรวจจับวัตถุจำเป็นต้องมีการทำผลเฉลยด้วยตัวเองและในหนึ่งภาพก็อาจมีได้หลายวัตถุจึงใช้ระยะเวลานานในการสร้างผลเฉลย

5.2 ข้อเสนอแนะ

5.2.1 ควรตรวจสอบประเภทการเก็บข้อมูลตัวเลขที่ไม่สัมพันธ์กันในชุดคำสั่งการเรียนรู้

5.2.2 ควรมีชุดข้อมูลในการเรียนรู้ที่ถ่ายด้วยถนนประเทศไทยจริง ๆ ให้มากขึ้น

เอกสารอ้างอิง

1. K. Surapong, “Data Augmentation คืออะไร ประโยชน์ของ Data Augmentaion ในการเทรน Deep Learning – Regularization ep.1,” 18 กันยายน 2562. [ออนไลน์]. สืบค้นจาก: https://www.bualabs.com/archives/1170/what-is-data-augmentation-benefits-data-augmentaion-deep-

Computer Science Department
Faculty of Informatics, Mahasarakham University

learning-training-model-deep-neural-network-regularization-ep-1/. [เข้าถึงเมื่อ 16 พฤศจิกายน 2563].

2. P. Sharma, “A Step-by-Step Introduction to the Basic Object Detection Algorithms (Part 1),” 11 ตุลาคม 2561. [ออนไลน์]. สืบค้นจาก: https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/. [เข้าถึงเมื่อ 11 กันยายน 2564].
3. S. M. I. Y. W. J. L. a. W. L. Zhun Fan, "Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network," 2018.
4. E. S. U. B. T. D. Jonathan Long, "Fully Convolutional Networks for Semantic Segmentation," ไม่ทราบปี.
5. H. Zhang และ D. Qin, “Deep Learning Based Fossil-Fuel Power Plant Monitoring in High Resolution Remote Sensing Images: A Comparative Study,” Beijing, China, 2562.
6. Yelisetty and Aakarsh, "Understanding Fast R-CNN and Faster R-CNN for Object Detection.," towardsdatascience, 14 7 2562. [Online]. Available: https://towardsdatascience.com/understanding-fast-r-cnn-and-faster-r-cnn-for-object-detection-adbb55653d97. [เข้าถึงเมื่อ 4 กันยายน 2564].

Computer Science Department
Faculty of Informatics, Mahasarakham University