

การแยกส่วนถนนทางความหมายออกจากภาพถ่ายระยะไกลโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน
นอลเชิงลึกและตัวชี้วัดภูมิทัศน์



นายธีรพงศ์ ปานบุญยืน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2559

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Semantic Road Segmentation on Remotely-Sensed Images Using Deep Convolutional
Neural Networks and Landscape Metrics

Mr. Teerapong Panboonyuen



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2016

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การแยกส่วนถนนทางความหมายออกจากภาพถ่าย ระยะไกลโดยใช้โครงข่ายประสาทแบบคอนโวลูชันนอลเชิง ลึกและตัวชี้วัดภูมิทัศน์
โดย	นายธีรพงศ์ ปานบุญยืน
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ดร.พีรพล เวทีกุล
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร.กุลสวัสดิ์ จิตรขจรวานิช ดร.สยาม ลววิโรจน์วงศ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ดร.พีรพล เวทีกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร.กุลสวัสดิ์ จิตรขจรวานิช)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร.สยาม ลววิโรจน์วงศ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สีบสกุล พิภพมงคล)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ภาณุ เศรษฐฐเสถียร)

ธีรพงศ์ ปานบุญยืน : การแยกส่วนถนนทางความหมายออกจากภาพถ่ายระยะไกลโดยใช้
 โครงข่ายประสาทแบบคอนโวลูชันนอลเชิงลึกและตัวชี้วัดภูมิทัศน์ (Semantic Road
 Segmentation on Remotely-Sensed Images Using Deep Convolutional Neural
 Networks and Landscape Metrics) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ดร.พีรพล เวทีกุล, อ.
 ที่ปรึกษาวิทยานิพนธ์ร่วม: ดร.กุลสวัสดี จิตรขจรวานิช, ดร.สยาม ลววิโรจน์วงศ์, 77 หน้า.

การจำแนกถนนออกจากภาพถ่ายระยะไกล ได้แก่ ภาพถ่ายทางอากาศ และภาพถ่าย
 ดาวเทียม นำมาใช้งานได้หลากหลายรูปแบบโปรแกรมประยุกต์ โดยเฉพาะอย่างยิ่งกับการสกัดถนนซึ่ง
 จัดเป็นขั้นหนึ่งในการเก็บฐานข้อมูลเชิงพื้นที่ ปัจจุบันการเรียนรู้เชิงลึกถูกนำมาใช้ในการสกัดถนนออก
 จากภาพถ่ายระยะไกล แต่ยังมีข้อจำกัดเรื่องค่าความถูกต้องในการจำแนกถนนซึ่งยังไม่มากพอแก่การ
 นำไปใช้งานเชิงโปรแกรมประยุกต์ วิทยานิพนธ์นี้เสนอวิธีการปรับปรุงโมเดลและเพิ่มประสิทธิภาพการ
 เรียนรู้เชิงลึกเพื่อสกัดถนนออกจากภาพถ่ายระยะไกลร่วมกับการประยุกต์ใช้ตัวชี้วัดภูมิทัศน์และ
 คอนดิชันนอลแรนดอมฟิลด์ส ผู้วิจัยเลือกใช้ฟังก์ชันกระตุ้นตัวใหม่ที่เรียกว่า หน่วยเอกซ์โพเนนเชียล
 เชิงเส้น เพื่อเพิ่มจำนวนผลลัพธ์ของวัตถุที่เป็นถนนให้มากขึ้น จากนั้นลดจำนวนวัตถุที่ไม่ใช่ถนนโดยใช้
 ตัวชี้วัดภูมิทัศน์เข้ามาตรวจสอบ สุดท้ายเพิ่มความคมและความถูกต้องให้แก่วัตถุที่เป็นถนนมากขึ้น
 ด้วยวิธีคอนดิชันนอลแรนดอมฟิลด์ส ชุดข้อมูลที่ใช้ในงานวิจัยนี้ คือ ชุดข้อมูลภาพถ่ายทางอากาศ
 (ชุดข้อมูลถนนรัฐแมสซาชูเซตส์) และชุดข้อมูลภาพถ่ายดาวเทียมจากดาวเทียมไทยโชตหรือธีออส ผล
 การทดลองด้วยวิธีที่นำเสนอแสดงให้เห็นว่าประสิทธิภาพของการจำแนกที่ได้ดีกว่าวิธีการมาตรฐาน
 (เซกเมนต์) ทั้งค่า Precision Recall และ F1

จุฬาลงกรณ์มหาวิทยาลัย
 CHULALONGKORN UNIVERSITY

ภาควิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2559

ลายมือชื่อ อ.ที่ปรึกษาร่วม

ลายมือชื่อ อ.ที่ปรึกษาร่วม

5870406821 : MAJOR COMPUTER ENGINEERING

KEYWORDS: DEEP CONVOLUTIONAL NEURAL NETWORKS / REMOTE SENSING / IMAGE PROCESSING / DEEP LEARNING / ROAD SEGMENTATION / SEMANTIC SEGMENTATION

TEERAPONG PANBOONYUEN: Semantic Road Segmentation on Remotely-Sensed Images Using Deep Convolutional Neural Networks and Landscape Metrics. ADVISOR: PEERAPON VATEEKUL, Ph.D., CO-ADVISOR: KULSAWASD JITKAJORNWANICH, Ph.D., SIAM LAWAWIROJWONG, Ph.D., 77 pp.

Road segmentation on remote sensing images: aerial (or very high resolution) images and satellite (or high resolution) images, has been employed to various application domains, particularly road extraction in which the segmented objects are served as a mandatory layer in geospatial databases. Several attempts in applying deep convolutional neural network (DCNN) to extract roads from remotely-sensed images have been made; nevertheless, the accuracy is still restricted. This thesis presents an enhanced DCNN framework specifically tailored for road extraction on remotely-sensed images by applying landscape metrics (LMs) and conditional random fields (CRFs). To improve DCNN, a modern activation function; called exponential linear unit (ELU), is engaged in our architecture resulting in a higher number of and yet more accurate extracted roads. To further alleviate falsely classified road objects, a solution based on an adoption of LMs is proposed. Lastly, to sharpen the extracted roads, a CRF method is added to our framework. The experiments were conducted on Massachusetts road aerial imagery as well as Thaichote/THEOS satellite imagery data sets. The results demonstrated that our proposed framework outperformed SegNet, the state-of-the-art object segmentation technique on any kinds of remotely-sensed imagery, in most of the cases in terms of precision, recall, and F1 scores.

Department:	Computer Engineering	Student's Signature
Field of Study:	Computer Engineering	Advisor's Signature
Academic Year:	2016	Co-Advisor's Signature
		Co-Advisor's Signature

กิตติกรรมประกาศ

ขอบพระคุณ ศ. ดร.บุญเสริม กิจศิริกุล และ ผศ. ดร.สีบสกุล พิภพมงคล ที่เป็นประธาน และกรรมการในการสอบวิทยานิพนธ์ ซึ่งอาจารย์นั้นแสนใจดีมากๆ ครับผม

ขอบพระคุณ อาจารย์ ดร.พีรพล เวทีกุล, อาจารย์ ดร.กุลสวัสดิ์ จิตขจรวานิช, ดร.สยาม ลาวีโรจน์วงศ์ และดร.ภานุ เศรษฐเสถียร มากๆ ครับผม อาจารย์ทุกท่านจะอยู่ในใจของผม ตลอดไปครับ

ขอบพระคุณ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์ มหาวิทยาลัย สถานที่ที่เปรียบเสมือนบ้านในสวนที่แสนอบอุ่น และขอบพระคุณอาจารย์ประจำ ภาควิชาฯ และพี่ๆ แม่บ้านทุกท่านที่คอยเอาใจใส่อีกทั้งยังหมั่นคอยถามไถ่ว่าเมื่อไหร่จะเรียนจบ และเรื่องต่างๆ ไปในชีวิตประจำวันครับ

ขอขอบคุณทุนอุดหนุนการศึกษาระดับบัณฑิตศึกษาจากบัณฑิตวิทยาลัย จุฬาลงกรณ์ มหาวิทยาลัย เพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัวทรงเจริญพระชนมายุครบ 72 พรรษา ให้กับผมเพื่อได้ศึกษาเล่าเรียนในระดับปริญญาวิทยาศาสตรมหาบัณฑิตครับ

ขอขอบคุณชุดข้อมูลภาพถ่ายดาวเทียม (THEOS Data Sets) ของสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน) (สทอภ.) (Geo-Informatics and Space Technology Development Agency (Public Organization) - GISTDA) สำหรับนำมาใช้ในการทดลองของวิทยานิพนธ์นี้ครับ

ขอบคุณพี่ๆ และเพื่อนๆ ในห้องปฏิบัติการการทำเหมืองข้อมูล (Data Mining Lab) และห้องปฏิบัติการอัจฉริยภาพเครื่องจักรและการค้นพบความรู้ (MIND Lab) ครับ สำหรับ กำลังใจและคำแนะนำดีๆ เพื่อนำมาปรับใช้ในวิทยานิพนธ์ครับผม

สุดท้ายนี้ ขอบพระคุณพ่อกับแม่ ลุง ป้า น้า อา และครอบครัวที่ให้กำลังใจ สนับสนุนให้ผมศึกษาต่อระดับปริญญาโทมหาบัณฑิต สนับสนุนผมตลอดมา และจะสนับสนุนผมตลอดไปนะครับ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญรูป	ฎ
สารบัญตาราง.....	ฅ
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	4
1.3 ขอบเขตของการวิจัย.....	4
1.4 ประโยชน์ที่ได้รับ	5
1.5 วิธีดำเนินการวิจัย.....	5
1.6 ผลงานตีพิมพ์จากการวิจัย	6
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	7
2.1 การจำแนกทางความหมาย (Semantic Segmentation).....	7
2.2 นิวรอลเน็ตเวิร์ก (Neural Network)	7
2.2.1 เพอร์เซ็ปตรอน (Perceptron)	8
2.2.2 นิวรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า (Feedforward Neural Network).....	9
ฟังก์ชันกระตุ้น (Activation Function).....	10
ฟังก์ชันต้นทุน (Cost Function หรือ Lost Function หรือ Objective Function).....	11
การหาค่าที่เหมาะสมที่สุด (Optimization).....	11
2.2.3 การแพร่กระจายย้อนกลับ (Backpropagation).....	12

2.3	นิรอรลเน็เตเว็ร็กคอรนโวลูชัน (Convolutional Neural Network).....	13
2.3.1	ชั้นคอรนโวลูชัน (Convolutional Layer).....	14
2.3.2	ขนาดของตัวกรอง (Filter Size).....	15
2.3.3	ชนิดของการทำคอรนโวลูชัน (Convolution Type).....	15
2.3.4	ชั้นพูลลิ่ง (Pooling Layer) และชั้นอันพูลลิ่ง (Unpooling Layer).....	18
2.3.5	ชั้นการเชื่อมต่อเต็มรูปแบบ (Fully Connected Layer).....	19
2.3.6	ชั้นดีคอรนโวลูชัน (Deconvolutional Layer).....	20
2.4	การวัดประสิทธิภาพ (Performance Evaluation).....	20
2.4.1	คอรนฟิวชันเมทริกซ์ (Confusion Matrix).....	20
2.4.2	ตัววัดประสิทธิภาพจำแนกโดยคลาส.....	21
2.5	การกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน (Connected-Component Labeling).....	21
2.6	การเพิ่มความราบเรียบโดยใช้ตัวกรองเกาส์เซียน (Smoothing with Gaussian Filter).....	22
2.7	การทำให้เป็นค่านอร์มัลไลเซชัน (Normalization).....	23
2.8	แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส (Conditional Random Fields: CRFs).....	24
2.9	ภาพถ่ายระยะไกล (Remotely-Sensed Images).....	25
บทที่ 3	งานวิจัยที่เกี่ยวข้อง.....	26
3.1.	งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกทางความหมาย (Deep Learning for Semantic Segmentation).....	26
3.2.	งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกถนนทางความหมาย (Deep Learning for Semantic Road Segmentation).....	31
3.3.	งานวิจัยเกี่ยวกับเทคนิคใหม่สำหรับการเรียนรู้เชิงลึก (Recent Techniques in Deep Learning).....	33
บทที่ 4	แนวคิดและวิธีการดำเนินงาน.....	38

4.1 การเตรียมข้อมูล	38
4.1.1 ชุดข้อมูลถนนของรัฐแมสซาชูเซตส์	38
4.1.2 ชุดข้อมูลถนนของประเทศไทย.....	39
4.2 การประมวลผลก่อน (Pre Processing).....	49
4.2.1 การทำนอร์มัลไลเซชัน (Normalization)	49
4.2.2 การเพิ่มปริมาณข้อมูลให้กับภาพถ่ายระยะไกล	49
4.2.3. การแปลงชุดข้อมูลภาพถ่ายระยะไกลให้เก็บอยู่ในรูปเวกเตอร์นัมพาย	49
4.3 การจำแนกถนนทางความหมายด้วยการเรียนรู้เชิงลึก	51
4.4 การประมวลผลหลัง (Post Processing).....	52
4.5 การสอน (Training)	55
4.6 การทดสอบ (Testing).....	55
บทที่ 5 การทดลองและผลการทดลอง	56
5.1 ผลการทดลองกับชุดข้อมูลภาพถ่ายทางอากาศ	56
5.1.1 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ (ELU-DCED)	57
5.1.2 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์ (ELU-DCED-LMs).....	57
5.1.3 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส (ELU-DCED-LMs-CRFs)	57
5.2 ผลการทดลองกับชุดข้อมูลภาพถ่ายดาวเทียม	62
5.2.1 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ (ELU-DCED).....	62

5.2.2 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์ (ELU-DCED-LMs).....	62
5.2.3 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส (ELU-DCED-LMs-CRFs)	63
บทที่ 6 สรุปการวิจัยและแนวทางการวิจัยในขั้นถัดไป	71
6.1 สรุปการวิจัย.....	71
6.2 ข้อเสนอแนะและข้อจำกัดในงานวิทยานิพนธ์นี้	71
6.3 แนวทางวิจัยในขั้นถัดไป.....	71
รายการอ้างอิง	73
ประวัติผู้เขียนวิทยานิพนธ์	77

สารบัญรูป

รูปที่ 1 ปัญหาที่พบเมื่อใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอน.....	3
รูปที่ 2 การเกิดปัญหา FP และ FN	3
รูปที่ 3 ลักษณะการจำแนกทางความหมาย	8
รูปที่ 4 โครงสร้างของเพอร์เซ็ปตรอน ข้อมูลรับเข้าและส่งออก.....	8
รูปที่ 5 โครงสร้างของนิรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า.....	10
รูปที่ 6 นิรอลเน็ตเวิร์กคอนโวลูชัน.....	14
รูปที่ 7 ตัวอย่างการทำคอนโวลูชัน โดยมีข้อมูลรับเข้าขนาด 6×6 และเมทริกตัวกรองขนาด 3×3 ...	14
รูปที่ 8 การทำคอนโวลูชันแบบกว้างและการเสริมเติม.....	15
รูปที่ 9 การทำคอนโวลูชันโดยมีข้อมูลรับเข้าขนาด 5×5 ตัวกรองขนาด 3×3	16
รูปที่ 10 การทำคอนโวลูชันโดยมีจำนวนตัวกรองเท่ากับ 3	16
รูปที่ 11 การทำขั้นพูลลิ่งและขั้นอันพูลลิ่ง	18
รูปที่ 12 ขั้นการรวมโดยใช้ค่าเฉลี่ย (ซ้าย) และค่ามากที่สุด (ขวา).....	19
รูปที่ 13 การทำขั้นคอนโวลูชันและขั้นดีคอนโวลูชัน	20
รูปที่ 14 การพิจารณาแบบ 4 จุดเพื่อนบ้าน และ 8 จุดเพื่อนบ้านตามลำดับ	22
รูปที่ 15 ตัวอย่างภาพการกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน	22
รูปที่ 16 ตัวอย่างภาพการเพิ่มความราบเรียบให้กับวัตถุโดยใช้ตัวกรองเกาส์เซียน	23
รูปที่ 17 รูปเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบพูลลิ่ง.....	26
รูปที่ 18 การข้ามชั้นเลเยอร์ของเน็ตเวิร์กในส่วนของการสกัดคุณลักษณะสำคัญของเน็ตเวิร์ก FCN	27
รูปที่ 19 เน็ตเวิร์กการเรียนรู้เชิงลึกแบบดีคอนโวลูชัน (Deconvolution network)	28
รูปที่ 20 รูปแบบของการทำดีคอนโวลูชันและอันพูลลิ่ง (Unpooling).....	28

รูปที่ 21 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ (Deep Convolutional Encoder-Decoder Networks หรือ DCED) 29

รูปที่ 22 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์โดยมีข้อมูลนำเข้าเป็นภาพ เอกซเรย์ 30

รูปที่ 23 เน็ตเวิร์กชนิดคอนโวลูชันเชิงลึกที่เสนอโดย J. Wang, J. Song, M. Chen และ Z. Yang . 31

รูปที่ 24 รูปแบบของถนนทั้ง 5 รูปแบบ ได้แก่ ถนนทางตรง, ถนนทางซ้าย, ถนนทางขวา, สีแยก และทางตัน..... 32

รูปที่ 25 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบพูลลี 32

รูปที่ 26 กราฟของ ELU เมื่อเทียบกับฟังก์ชันกระตุ้นที่นิยมใช้กันในปัจจุบันอย่าง ReLU, LReLU (leaky ReLU) และ SReLU (S-shaped ReLU)..... 34

รูปที่ 27 กราฟ Cross Entropy Loss ระหว่าง ELU กับ ReLU และ Leaky ReLU โดยค่าแกน x เป็นจำนวน Epoch และแกน y เป็นค่า Loss ของ ครอสเอนโทรปี (Cross Entropy)..... 35

รูปที่ 28 ภาพรวมทั้งหมดของการประยุกต์การเรียนรู้เชิงลึกกับโมเดล CRFs 35

รูปที่ 29 การปรับผลลัพธ์ที่ได้ในแต่ละรอบ CRFs (CRF Iteration) ซึ่งผลลัพธ์ที่เมื่อสิ้นสุดรอบที่ 10 จะมีลักษณะใกล้เคียงกับผลเฉลย 36

รูปที่ 30 ตัวอย่างผลลัพธ์ที่ได้จากการประยุกต์การเรียนรู้เชิงลึกร่วมกับการใช้ CRFs 36

รูปที่ 31 ผลลัพธ์ที่ได้จากการประเมินโดยเทียบกันระหว่างโมเดลที่ใช้ CRFs และโมเดลการเรียนรู้เชิงลึกมาตรฐาน 37

รูปที่ 32 ภาพรวมของขั้นตอนทั้งหมดที่วิทยานิพนธ์นี้นำเสนอ 38

รูปที่ 33 สองตัวอย่างภาพถ่ายทางอากาศและภาพผลเฉลยของชุดข้อมูลรัฐแมสซาชูเซตส์ 39

รูปที่ 34 ตำแหน่งที่ตั้งของเขตพื้นที่พุทธมณฑล (นครปฐม) ในชุดข้อมูลภาพถนนประเทศไทย 42

รูปที่ 35 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่พุทธมณฑล (นครปฐม) ในชุดข้อมูลภาพถนนประเทศไทย 43

รูปที่ 36 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่ศรีราชา (ชลบุรี) ในชุดข้อมูลภาพถนนประเทศไทย 44

รูปที่ 37 ตำแหน่งที่ตั้งของเขตพื้นที่ศรีราชา (ชลบุรี) ในชุดข้อมูลภาพถนนประเทศไทย..... 44

รูปที่ 38 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่สงขลาในชุดข้อมูลภาพถนนประเทศไทย	45
รูปที่ 39 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่สุรินทร์ในชุดข้อมูลภาพถนนประเทศไทย....	46
รูปที่ 40 ตำแหน่งที่ตั้งของเขตพื้นที่สงขลาในชุดข้อมูลภาพถนนประเทศไทย	46
รูปที่ 41 สามตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่อุบลราชธานีในชุดข้อมูลภาพถนนประเทศไทย	47
รูปที่ 42 ตำแหน่งที่ตั้งของเขตพื้นที่สุรินทร์ในชุดข้อมูลภาพถนนประเทศไทย	48
รูปที่ 43 ตำแหน่งที่ตั้งของเขตพื้นที่อุบลราชธานีในชุดข้อมูลภาพถนนประเทศไทย	48
รูปที่ 44 การนำภาพถ่ายระยะไกล 1 ภาพมาเพิ่มปริมาณข้อมูลโดยการหมุนภาพไปที่ละ 45 องศา (1 ภาพเพิ่มปริมาณขึ้นอีก 7 ภาพ).....	50
รูปที่ 45 การนำภาพผลเฉลย 1 ภาพมาเพิ่มปริมาณข้อมูลโดยการหมุนภาพไปที่ละ 45 องศา (1 ภาพเพิ่มปริมาณขึ้นอีก 7 ภาพ).....	50
รูปที่ 46 การแปลงภาพถ่ายระยะไกล (1) และภาพผลเฉลย (2) ให้อยู่ในรูปแบบของเวกเตอร์นัมพาย.....	51
รูปที่ 47 ภาพสถาปัตยกรรมการเรียนรู้เชิงลึกที่งานวิทยานิพนธ์นี้นำเสนอ (ELU-DCED).....	52
รูปที่ 48 ตัวอย่างวัตถุที่มีดัชนีความซับซ้อนรูปแบบต่างๆ	53
รูปที่ 49 ผลลัพธ์จากการใช้เทคนิค Connected-Component Labeling และ Gaussian Filter .	54
รูปที่ 50 ผลลัพธ์การคำนวณดัชนีความซับซ้อนแต่ละวัตถุที่ได้จากหลังจากการเรียนรู้เชิงลึก	54
รูปที่ 51 กราฟค่าความสูญเสียในแต่ละรอบอีพอค (รูปซ้าย) และกราฟของการประเมินค่า Precision Recall และ F1 บนชุดข้อมูลวาลิดเดชั่นเพื่อเลือกโมเดลที่ดีที่สุด (รูปขวา).....	59
รูปที่ 52 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายทางอากาศ	61
รูปที่ 53 กราฟค่าความสูญเสียในแต่ละรอบอีพอค (รูปซ้าย) และกราฟของการประเมินค่า Precision Recall และ F1 บนชุดข้อมูลวาลิดเดชั่นเพื่อเลือกโมเดลที่ดีที่สุด (รูปขวา).....	65
รูปที่ 54 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัดนครปฐม)	66
รูปที่ 55 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัดชลบุรี)	67

รูปที่ 56 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูล จังหวัดสงขลา).....	68
รูปที่ 57 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูล จังหวัดสุรินทร์).....	69
รูปที่ 58 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูล จังหวัดอุบลราชธานี).....	70



สารบัญตาราง

ตารางที่ 1 คอนฟิวชันเมทริกซ์ของการจำแนกแบบ 2 คลาส	20
ตารางที่ 2 แสดงคุณลักษณะดาวเทียมธีออส	40
ตารางที่ 3 จำนวนชุดข้อมูลในข้อมูลการสอน, ข้อมูลตรวจสอบ และข้อมูลทดสอบ	58
ตารางที่ 4 จำนวนชุดข้อมูลในข้อมูลการสอนหลังจากการหมุนภาพ, ข้อมูลตรวจสอบและข้อมูลทดสอบ	58
ตารางที่ 5 ประสิทธิภาพการจำแนกโดยทดสอบกับชุดข้อมูลภาพถ่ายทางอากาศ	59
ตารางที่ 6 ประสิทธิภาพการจำแนกโดยวัดจากค่า Precision	63
ตารางที่ 7 แสดงประสิทธิภาพการจำแนกโดยวัดจากค่า Recall.....	64
ตารางที่ 8 แสดงประสิทธิภาพการจำแนกโดยวัดจากค่า F1	64

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ภาพถ่ายระยะไกล (Remotely-sensed Image) มีความสำคัญมากในการนำมาใช้งานในรูปแบบโปรแกรมประยุกต์ (Application) โดยเฉพาะอย่างยิ่งการนำมาใช้ปรับปรุงข้อมูลบนระบบสารสนเทศภูมิศาสตร์ (Geographic Information System) [1, 2] ให้ทันสมัย ซึ่งตัวอย่างรูปแบบโปรแกรมประยุกต์ที่ยังคงได้รับความนิยมในเทคโนโลยีด้านการสำรวจระยะไกล (Remote Sensing) คือการสกัดพื้นที่ถนนออกจากภาพถ่ายระยะไกล โดยในวิทยานิพนธ์นี้ได้ใช้ภาพถ่ายระยะไกล 2 ชนิด ได้แก่ ภาพถ่ายดาวเทียม (Satellite Image) ซึ่งคุณสมบัติของภาพจะเป็นภาพที่มีความละเอียดสูง (High Resolution) และภาพถ่ายทางอากาศ (Aerial Image) ซึ่งคุณสมบัติของภาพชนิดนี้จะเป็นภาพที่มีความละเอียดสูงมาก (Very High Resolution)

ปัจจุบันการสกัดถนนออกจากภาพถ่ายระยะไกลยังคงมีความสำคัญมากและเป็นหัวข้อที่มีความท้าทาย [1] เนื่องจากเส้นทางถนนมีการเปลี่ยนแปลงเพิ่มเติมอยู่ตลอดเวลา การที่คนหรือรถจะเข้าไปสำรวจบางพื้นที่อาจจะทำได้ยากหรือไม่อาจทำได้เลย เช่น ในเขตพื้นที่สู้รบหรือพื้นที่ที่กำลังทำสงคราม รวมถึงพื้นที่ภัยพิบัติ ซึ่งพื้นที่เหล่านี้จะจัดอยู่ในหมวดที่เรียกว่า พื้นที่ที่ไม่ยินดียินร้าย (Impervious Surface Extraction) [1] ทำให้เข้าไปอัพเดทแผนที่ถนนไม่ได้ ดังนั้นการสกัดพื้นที่ถนนโดยใช้ภาพถ่ายระยะไกลจึงเป็นทางเลือกที่ดีที่สุดสำหรับการนำมาอัพเดทข้อมูลถนน ประโยชน์ที่ได้นอกเหนือจากการสกัดพื้นที่ถนนสำหรับพื้นที่ที่เข้าไปสำรวจไม่ได้ เช่น การอัพเดทพื้นที่ถนนที่ได้จากภาพถ่ายระยะไกลจะนำไปใช้ประโยชน์ในการวางผังเมือง (Urban Planning) การเพิ่มประสิทธิภาพของเส้นทางและระบบนำทาง (Route Optimization and Navigation) รวมไปถึงนำมาใช้ประโยชน์ในการหาเส้นทางในการจัดการภัยพิบัติได้ (Disaster Management) [1-6]

งานวิจัยด้านการจำแนกถนนออกจากภาพถ่ายระยะไกลแบ่งวิธีการจำแนกถนนออกจากภาพถ่ายระยะไกลได้เป็น 2 วิธีการ คือ ใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning) และใช้วิธีการเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning)

วิธีการแรก คือ วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning) เช่น วิธีการสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้เคมีนส์ (K-means) [6, 7] วิธีการสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้ค่าขีดแบ่งแบบโฮโมแกรม (Homogram Thresholding) [8] และวิธีการสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้กราฟคัท (Graph Cut) [5] โดยวิธีการเหล่านี้มีข้อดี คือ สกัดถนนได้

รวดเร็ว แต่มีข้อจำกัด คือ เรียนรู้คุณลักษณะสำคัญ (Feature) ได้น้อย สกัดออกมาได้เพียงแคคุณลักษณะสำคัญของสี ซึ่งจะมีผลทำให้เมื่อนำวิธีการเหล่านี้ไปใช้กับภาพถ่ายระยะไกลที่มีสีของถนนที่แตกต่างกัน หรือในกรณีของพื้นที่ถนนและพื้นที่ที่ไม่ใช่ถนนอยู่ติดกัน โดยมีคุณลักษณะของสีหรือเนื้อผิวของพื้นที่ทั้งสองที่เหมือนกัน เมื่อใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอนอย่างวิธีการจำแนกโดยการแบ่งกลุ่ม (Clustering) ด้วยวิธีการเคมีนส์ (K-means) จากวิธีการที่ R. Maurya, P. R. Gupta และ A. S. Shukla [4] นำเสนอในปี 2009 หรือจากวิธีการที่ C. Sujatha และ D. Selvathi [9] ในปี 2015 นำเสนอวิธีการโดยใช้การหาค่าขีดแบ่งมาจำแนกพื้นที่ถนนออกจากภาพถ่ายระยะไกล ปัญหาที่พบในงานวิจัยของพวกเขาคือกลุ่มของพื้นที่ถนนกับกลุ่มของพื้นที่ที่ไม่ใช่ถนนที่อยู่ติดกันจะถูกมองว่าเป็นกลุ่มเดียวกัน แต่แท้จริงแล้วต้องอยู่แยกกัน รูปที่ 1 แสดงตัวอย่างของปัญหาที่ใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอนจำแนกถนนออกจากภาพถ่ายระยะไกล แต่ด้วยข้อจำกัดของวิธีการเรียนรู้แบบไม่มีผู้ฝึกสอนตามที่กล่าวไปข้างต้นทำให้ไม่อาจแก้ปัญหากรณีเหล่านี้ได้ และทำให้ประสิทธิภาพในการสกัดถนนออกจากภาพถ่ายระยะไกลนั้นไม่ดี

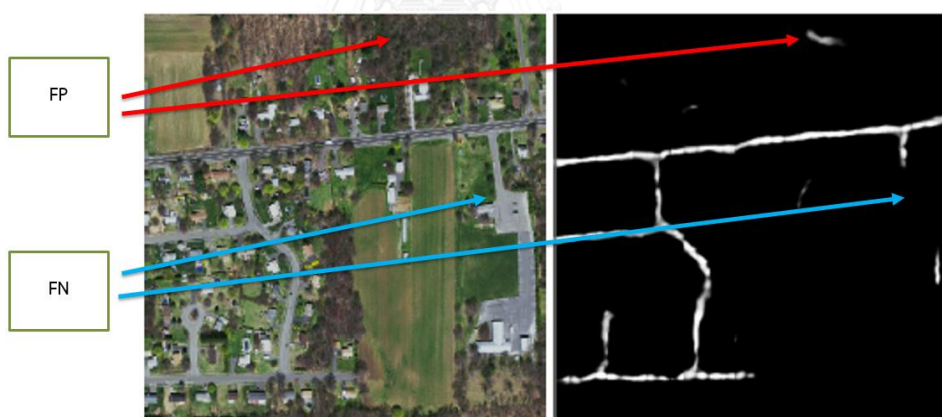
วิธีการที่สอง คือวิธีการเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning) เช่น การใช้การเรียนรู้เชิงลึก (Deep Learning) เข้ามาสกัดถนนออกจากภาพถ่ายระยะไกล ปัจจุบันการเรียนรู้เชิงลึกถูกนำมาแก้ไขปัญหาที่เกิดจากวิธีการเรียนรู้แบบไม่มีผู้ฝึกสอนได้ ซึ่งการเรียนรู้เชิงลึกเรียนรู้คุณลักษณะสำคัญได้มากมาย เช่น คุณลักษณะสำคัญของสี (Color Feature) คุณลักษณะสำคัญของรูปร่าง (Shape Feature) รวมไปถึงคุณลักษณะสำคัญของเนื้อผิว (Texture Feature) โดยการเรียนรู้เชิงลึกที่นิยมนำมาใช้ในการสกัดถนนออกจากภาพถ่ายระยะไกลในปัจจุบัน ได้แก่ นิรอรอลเน็ตเวิร์กคอนโวลูชันเชิงลึก (Deep Convolutional Neural Network) นิรอรอลเน็ตเวิร์กคอนโวลูชันเชิงลึกที่นำมาสกัดถนนแตกแขนงออกได้เป็นนิรอรอลเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี (Fully Convolutional Networks หรือ FCN) [3, 10] นิรอรอลเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ (Deep Convolutional Encoder-Decoder Neural Networks หรือ DCED) [4, 11, 12] และนิรอรอลเน็ตเวิร์กแบบดีคอนโวลูชันเชิงลึก (Deep Deconvolutional Neural Networks หรือ DeCNN) [13, 14]

ปัจจุบันแม้ว่าการเรียนรู้เชิงลึกจะสกัดคุณลักษณะสำคัญได้มาก และให้ประสิทธิภาพการสกัดถนนออกจากภาพถ่ายระยะไกลได้ดี อีกทั้งเรียนรู้และสกัดคุณลักษณะสำคัญจากรูปภาพได้โดยอัตโนมัติ ไม่จำเป็นต้องใช้ความรู้ความเชี่ยวชาญเฉพาะด้านของมนุษย์ แต่ยังมีความต้องการในเรื่องของการเพิ่มประสิทธิภาพให้มากขึ้นกว่าเดิม เนื่องจากปัญหาที่พบในการสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้การเรียนรู้แบบลึก ยังมีปัญหาที่เรียกว่า ปัญหาผลบวกลวง (False Positive หรือ FP) คือ ปัญหาที่ทำนายว่าพื้นที่นั้นเป็นถนน แต่แท้จริงแล้วพื้นที่นั้นไม่ใช่ถนน เช่น พื้นที่ลานจอดรถ และปัญหาที่เรียกว่า ปัญหาผลลบลวง (False Negative หรือ FN) คือ ปัญหาที่ทำนายว่าพื้นที่นั้นไม่ใช่

ถนน แต่แท้จริงแล้วเป็นพื้นที่ถนน เช่น ถนนในซอยแคบๆ รูปที่ 2 แสดงรูปปัญหาที่พบในการสกัดถนนออกจากภาพถ่ายทางอากาศ



รูปที่ 1 ปัญหาที่พบเมื่อใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอน โดยภาพขาวสุดจะแสดงให้เห็นว่ากลุ่มของพื้นที่ถนนกับกลุ่มของพื้นที่ที่ไม่ใช่ถนนที่อยู่ติดกันจะมองว่าอยู่กลุ่มเดียวกันซึ่งเป็นคำตอบที่ผิดเมื่อเปรียบเทียบกับภาพกลางที่เป็นผลเฉลยที่ถูกต้อง



รูปที่ 2 การเกิดปัญหา FP และ FN โดยภาพซ้ายเป็นภาพถ่ายระยะไกลที่มีคุณสมบัติของภาพเป็นภาพที่มีความละเอียดสูงมาก และภาพขวาเป็นภาพที่ทำนายถนนโดยใช้การเรียนรู้เชิงลึก¹

วิทยานิพนธ์นี้เสนอวิธีการเรียนรู้เชิงลึกแบบนิรอรเน็ตเวิร์กคอนโวลูชันเชิงลึกเพื่อสกัดถนนออกจากภาพถ่ายระยะไกลทั้งภาพถ่ายดาวเทียมและภาพถ่ายทางอากาศร่วมกับการใช้ตัวชี้วัดภูมิทัศน์ (Landscape Metrics) [15] โดยเลือกตัวชี้วัดภูมิทัศน์ประเภทตัวชี้วัดรูปร่าง (Shape Metrics)

¹ อ้างอิงจาก Fig. 7.9 ใน [13]

เพื่อนำมาคำนวณหาดัชนีค่าความซับซ้อน (Shape Index) และใช้แบบจำลองทางสถิติคอนดิชันนอล แรนดอมฟิลด์ส (Conditional Random Fields: CRFs) [16-18] เข้ามาช่วยในการประมวลผลหลัง (Post Processing) เพื่อเพิ่มความคม (Sharpening) ให้กับผลลัพธ์ที่ได้หลังจากการสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้การเรียนรู้เชิงลึกแบบนิรอรอนเน็ตเวิร์กคอนโวลูชันเชิงลึก ทั้งนี้สิ่งที่วิทยานิพนธ์นี้มีจุดมุ่งหมายเพื่อแก้ปัญหาผลบวกหลง (False Positive) และปัญหาผลลบหลง (False Negative) โดยนำไปสู่การเพิ่มประสิทธิภาพให้ดีขึ้นกว่าเดิมเมื่อนำไปเปรียบเทียบกับวิธีการมาตรฐาน (Baseline) ได้แก่ นิรอรอนเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี นิรอรอนเน็ตเวิร์กแบบดีคอนโวลูชันเชิงลึก และนิรอรอนเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์

1.2 วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอสถาปัตยกรรมการเรียนรู้เชิงลึกสำหรับสกัดถนนออกจากภาพถ่ายระยะไกลโดยใช้วิธีการเรียนรู้เชิงลึกแบบนิรอรอนเน็ตเวิร์กคอนโวลูชันเชิงลึกร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส สกัดถนนได้ทั้งภาพถ่ายดาวเทียมและภาพถ่ายทางอากาศโดยให้ประสิทธิภาพการจำแนกที่ดีขึ้นกว่าใช้วิธีการเรียนรู้เชิงลึกมาตรฐาน

1.3 ขอบเขตของการวิจัย

ในการสกัดถนนออกจากภาพถ่ายระยะไกล วิทยานิพนธ์นี้ใช้ชุดข้อมูลภาพถ่ายระยะไกล 2 ชุดข้อมูลสำหรับการสอนและการทดสอบ ได้แก่ (1) ชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์ (Massachusetts Roads Data Set) [19] ซึ่งชุดข้อมูลนี้เป็นชุดข้อมูลภาพถ่ายทางอากาศ โดยคุณสมบัติของภาพชนิดนี้จะเป็นภาพที่มีความละเอียดสูงมาก และ (2) ชุดข้อมูลของประเทศไทย (Thailand Roads Data Sets) โดยเป็นชุดข้อมูลภาพถ่ายดาวเทียม (Satellite Image) ซึ่งคุณสมบัติของภาพจะเป็นภาพที่มีความละเอียดสูง (High Resolution) สำหรับชุดข้อมูลของประเทศไทยเป็นชุดข้อมูลที่ถ่ายจากดาวเทียมธีออส (THEOS) โดยผู้ให้บริการ คือ สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (GISTDA) ชุดข้อมูลนี้ประกอบด้วย 5 สถานที่ ได้แก่ อำเภอพุทธมณฑล จังหวัดนครปฐม, อำเภอศรีราชา จังหวัดชลบุรี, จังหวัดสงขลา, จังหวัดสุรินทร์ และจังหวัดอุบลราชธานี

1.4 ประโยชน์ที่ได้รับ

สกัดถนนออกจากภาพถ่ายดาวเทียมและภาพถ่ายทางอากาศได้โดยมีความแม่นยำและประสิทธิภาพการจำแนกมากกว่าวิธีการเดิม (วิธีการมาตรฐาน) ซึ่งการสกัดถนนออกจากภาพถ่ายระยะไกลนี้นำไปช่วยในการวางผังเมือง การสำรวจพื้นที่โดยไม่ต้องใช้คนหรือรถ การอัปเดตแผนที่ และเพิ่มประสิทธิภาพของระบบนำทางได้

1.5 วิธีดำเนินการวิจัย

1. ศึกษาวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับหัวข้อที่จะทำ เช่น การเรียนรู้เชิงลึก การสกัดถนนออกจากภาพถ่ายระยะไกล การประมวลผลภาพ
2. รวบรวมข้อมูลตัวอย่างภาพถ่ายระยะไกล (ภาพถ่ายดาวเทียมและภาพถ่ายทางอากาศ) สำหรับทำการทดลอง
3. ศึกษาการใช้งานเครื่องมือการเรียนรู้เชิงลึก โดยในงานวิจัยเลือกใช้เครื่องมือการเรียนรู้เชิงลึกที่ชื่อว่า ลาซานญา (Lasagne) โดยใช้ภาษาไพธอน (Python) ในการเขียนโค้ด
4. ศึกษาและเขียนโค้ดเพื่อจำลองการทำงานวิจัยที่เกี่ยวข้องเพื่อใช้เป็นมาตรฐานเปรียบเทียบในการประเมินประสิทธิภาพการทำงาน
5. ทดลองแนวทางวิจัยที่จะนำเสนอในเบื้องต้น รวมถึงประเมินประสิทธิภาพสิ่งที่น่าสนใจกับวิธีการมาตรฐาน
6. เขียนและตีพิมพ์เพื่อนำไปเสนอผลงานทางวิชาการ
7. จัดทำวิทยานิพนธ์

1.6 ผลงานตีพิมพ์จากการวิจัย

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ ได้รับการตีพิมพ์เป็นบทความทางวิชาการ 4 เรื่อง ดังนี้

1. **“Road Segmentation on Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields”** โดย ชีรพงศ์ ปานบุญยืน, พีรพล เวทีกุล, กุลสวัสดิ์ จิตขจรวานิช, สยาม ลววิโรจน์ และภานุ เศรษฐเสถียร ในวารสารระดับนานาชาติ Remote Sensing (ISSN 2072-4292) อยู่ในฐานข้อมูล ISI Journal/SCImago Journal Rank (SJR): Q1/ Impact Factor (2016), 3.244
2. **“An Enhanced Deep Convolutional Encoder-Decoder Network for Road Segmentation on Aerial Imagery”** โดย ชีรพงศ์ ปานบุญยืน, พีรพล เวทีกุล, กุลสวัสดิ์ จิตขจรวานิช และ สยาม ลววิโรจน์วงศ์ ในงานประชุมวิชาการ “The 13th International Conference on Computing and Information Technology, Recent Advances in Information and Communication Technology 2017 series, Volume 566. The proceedings will also be indexed by ISI Proceedings, DBLP, Ulrich's, EI-Compendex, SCOPUS, Zentralblatt Math, MetaPress, and Springerlink” จัดขึ้น ณ โรงแรมอโนมา เซ็นทรัลเวิลด์ กรุงเทพมหานคร ประเทศไทย ระหว่างวันที่ 6 – 7 กรกฎาคม 2560
3. **“Image Vectorization of Road Satellite Data Sets”** โดย ชีรพงศ์ ปานบุญยืน, พีรพล เวทีกุล และ กุลสวัสดิ์ จิตขจรวานิช ในวารสารระดับชาติ Journal of Remote Sensing and GIS Association of Thailand (วารสารสมาคมสำรวจระยะไกลและสารสนเทศภูมิศาสตร์แห่งประเทศไทย), Vol.17 (ISSN 1513-4261), Special Issue (Page 310-321), 2016. (อยู่ในฐานข้อมูล TCI Tier 2)
4. **“Adapting Landscape Metrics for Connected Component-Based Road Extraction on Satellite Images”** โดย กุลสวัสดิ์ จิตขจรวานิช, ชีรพงศ์ ปานบุญยืน และ พีรพล เวทีกุล ในวารสารระดับนานาชาติ International Review on Computers and Software (IRECOS) อยู่ในฐานข้อมูล Scopus (SCImago Journal Rank (SJR): Q2) (อยู่ในช่วงพิจารณาการตอบรับบทความวิชาการ)

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องกับวิทยานิพนธ์นี้ประกอบด้วย การจำแนกทางความหมาย นิวรอลเน็ตเวิร์ก (Neural Network) นิวรอลเน็ตเวิร์กคอนโวลูชัน (Convolutional Neural Network) การวัดประสิทธิภาพ (Performance Evaluation) การเพิ่มความราบเรียบโดยใช้ตัวกรองเกาส์เซียน (Smoothing with Gaussian Filter) การทำให้เป็นค่าอนอร์มัลไลเซชัน (Normalization) แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส (Conditional Random Fields: CRFs) และภาพถ่ายระยะไกล (Remotely-Sensed Images)

2.1 การจำแนกทางความหมาย (Semantic Segmentation)

เป็นการจำแนกวัตถุภายในรูปภาพโดยสกัดออกมาเป็นรูปร่างของวัตถุที่เราสนใจ ซึ่งปัจจุบัน การเรียนรู้เชิงลึกในทางการประมวลผลภาพ (Image Processing) นิยมทำการจำแนกทางความหมาย มากกว่าการจำแนกรูปภาพ (Image Classification) หรือการตรวจจับรูปภาพ (Image Detection) โดยจะแสดงความแตกต่างระหว่างการจำแนกรูปภาพ การตรวจจับรูปภาพและการจำแนกทางความหมายไว้ดังรูปที่ 3

2.2 นิวรอลเน็ตเวิร์ก (Neural Network)

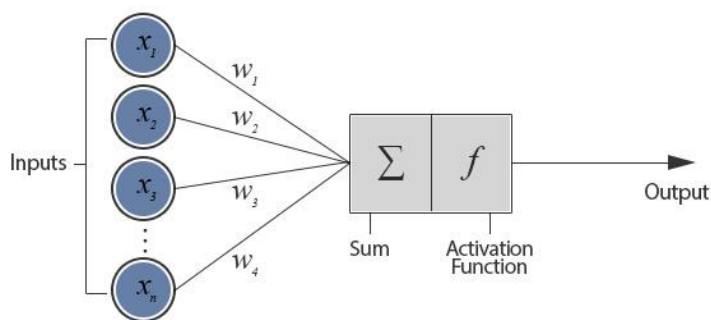
ด้วยแรงบันดาลใจจากสมองของมนุษย์ทำให้เกิดเป็นแบบจำลองที่เรียกว่า นิวรอลเน็ตเวิร์ก ซึ่งทำให้คอมพิวเตอร์รู้จักคิดและจดจำในทำนองเดียวกับสมอง การทำงานของนิวรอลเน็ตเวิร์ก เปรียบเสมือนการทำงานของโครงข่ายประสาทในสมอง คือ การเรียนรู้จากข้อมูลที่มีอยู่แล้ว เพื่อใช้ทำนายข้อมูลในลักษณะเดียวกัน² โดยในหัวข้อนี้จะมีหัวข้อย่อยเรียงตามลำดับดังต่อไปนี้ หน่วยพื้นฐานของนิวรอลเน็ตเวิร์กที่เรียกว่า เพอร์เซ็ปตรอน ถัดมาเป็นนิวรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า และสุดท้าย คือ การแพร่กระจายย้อนกลับ

² อ้างอิงจาก [20]

1			
2	แมว	ลูกแมว ลูกเปิด ลูกสุนัข	ลูกแมว ลูกเปิด ลูกสุนัข
3	การจำแนก (Classification)	การตรวจจับ (Detection)	การจำแนกทางความหมาย (Semantic Segmentation)

รูปที่ 3 ลักษณะการจำแนกทางความหมาย³ โดยแถวที่ 1 เป็นการแสดงภาพต้นฉบับและแสดงผลลัพธ์ที่เกิดขึ้นจากการใช้แต่ละวิธีตามแถวที่ 3 โดยคำตอบที่ได้จะอยู่ในแถวที่ 2

2.2.1 เพอร์เซ็ปตรอน (Perceptron)



รูปที่ 4 โครงสร้างของเพอร์เซ็ปตรอน ข้อมูลรับเข้าและส่งออก⁴

ส่วนประกอบที่เล็กที่สุดของนิวรอลเน็ตเวิร์ก คือเพอร์เซ็ปตรอน (รูปที่ 4) เปรียบได้กับเซลล์ประสาทหนึ่งเซลล์ที่เรียกว่านิวรอล เพอร์เซ็ปตรอนคือขั้นตอนวิธีการจำแนกที่ให้ผลลัพธ์เป็นสองกลุ่มคือ 1 และ 0 กำหนดให้ฟังก์ชันของเพอร์เซ็ปตรอนแทนด้วย $f(x)$ โดยมีข้อมูลรับเข้า คือ x และข้อมูลส่งออก คือ y สมการเขียนได้โดย

³ อ้างอิงจาก Fig. 18 ใน [21]

⁴ อ้างอิงจาก <http://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-1/7>

$$\hat{y} = f(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^m w_i x_i + b > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

โดย w คือเวกเตอร์ของน้ำหนัก (Weights) ถัดมา b คือค่าไบแอส (Bias) และสุดท้าย m คือขนาดของข้อมูลรับเข้า สำหรับกระบวนการเรียนรู้ของเพอร์เซ็ปตรอน

กำหนดให้ชุดข้อมูลตัวอย่างแทนด้วย x และผลลัพธ์จริงของตัวอย่างแทนด้วย y โดยจะมีสมการดังต่อไปนี้

$$w_i \leftarrow w_i + \Delta w_i \quad (2)$$

$$\text{โดยที่ } \Delta w_i = \alpha(\hat{y} - y)x_i \quad (3)$$

โดยพารามิเตอร์ α คืออัตราการเรียนรู้ (Learning rate) เป็นค่าที่ใช้บอกว่าในแต่ละรอบของการเรียนรู้ จะมีการเปลี่ยนแปลงน้ำหนักเทียบกับอัตราส่วนของผลต่างของผลลัพธ์ไปมากเท่าใด

2.2.2 นิวรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า (Feedforward Neural Network)

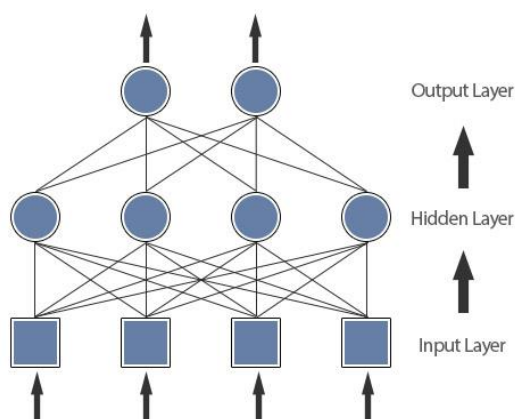
นิวรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า (รูปที่ 5)⁵ มีลำดับในการคำนวณและส่งผ่านของข้อมูลไปในทิศทางเดียว โดยโครงสร้างจะแบ่งออกเป็นลำดับชั้น ในแต่ละลำดับชั้นจะมีเพอร์เซ็ปตรอนจำนวนหนึ่งซึ่งไม่มีเส้นเชื่อมถึงกันภายในชั้นเดียวกัน แต่จะมีเส้นเชื่อมกับเพอร์เซ็ปตรอนตัวอื่นที่อยู่ในลำดับชั้นที่ติดกันทั้งหมด โดยข้อมูลส่งออกของเพอร์เซ็ปตรอนในชั้นก่อนหน้าเป็นข้อมูลรับเข้าของเพอร์เซ็ปตรอนในชั้นปัจจุบัน

กำหนดให้สัญลักษณ์แทนการคำนวณไปข้างหน้า (Feedforward) โดยให้ a_k^{l-1} คือผลลัพธ์ของเพอร์เซ็ปตรอนตัวที่ k ในลำดับชั้น $l-1$ และ w_{jk}^l คือ น้ำหนักสำหรับเพอร์เซ็ปตรอนตัวที่ j ในลำดับชั้น l ที่มีเส้นเชื่อมมาจากเพอร์เซ็ปตรอนตัวที่ k ในลำดับชั้นก่อนหน้า และ b_j^l คือ ค่าไบแอส นอกจากนี้ ให้ g แทนฟังก์ชันกระตุ้น และให้ n แทนจำนวนเพอร์เซ็ปตรอนในลำดับชั้นที่ $l-1$ จะเขียนสมการของการคำนวณ a_j^l ได้เป็น

$$z_j^l = \sum_{k=1}^n w_{jk}^l a_k^{l-1} + b_j^l \quad (4)$$

⁵ อ้างอิงจาก [20]

$$a_j = g(z_j) \quad (5)$$



รูปที่ 5 โครงสร้างของนิเวรอลเน็ตเวิร์กแบบป้อนไปข้างหน้า⁶

ฟังก์ชันกระตุ้น (Activation Function)

ข้อมูลส่งออกของแต่ละเพอร์เซ็ปตรอน จะมีการใช้ฟังก์ชันกระตุ้น $g(z)$ เพื่อให้ให้นิเวรอลเน็ตเวิร์กแก้ปัญหาได้หลากหลายมากขึ้น⁷ ฟังก์ชันกระตุ้นมีหลายรูปแบบ โดยฟังก์ชันกระตุ้นที่นำมาใช้ในวิทยานิพนธ์มีดังต่อไปนี้

- ฟังก์ชันค่าสูงสุดอย่างอ่อน (Softmax Function)

เป็นฟังก์ชันที่ให้ค่าผลลัพธ์ออกมาอยู่ในช่วง 0 ถึง 1 กำหนดให้ในชั้นของเน็ตเวิร์กที่ต้องการคำนวณมีผลลัพธ์ทั้งหมด K ตัว ค่าของผลลัพธ์นั้นแทนด้วยสัญลักษณ์ z จะได้ว่าฟังก์ชันค่าสูงสุดอย่างอ่อนของผลลัพธ์ตัวที่ j หรือแทนด้วยสัญลักษณ์ f_j โดยคำนวณได้จากสมการดังต่อไปนี้

$$f(z)_j = \frac{e^{z_j}}{\sum_{i=1}^K e^{z_i}} \quad (6)$$

⁶ อ้างอิงจาก <http://www.theprojectspot.com/tutorial-post/introduction-to-artificial-neural-networks-part-1/7>

⁷ อ้างอิงจาก [20]

- ฟังก์ชันเรกติไฟด์เชิงเส้น (Rectified Linear Unit Function หรือ ReLU)

เป็นฟังก์ชันที่ให้ผลลัพธ์ออกมาเป็นจำนวนบวก หรือเป็นศูนย์เสมอ สมการของฟังก์ชันเรกติไฟด์เชิงเส้น f คำนวณได้จากสมการที่ 7

$$f(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } z \geq 0 \end{cases} \quad (7)$$

ฟังก์ชันต้นทุน (Cost Function หรือ Lost Function หรือ Objective Function)

ฟังก์ชันต้นทุน⁸ เป็นฟังก์ชันของนิเวศวิทยาที่แสดงถึงต้นทุนของเน็ตเวิร์ก กล่าวคือ ใน การเรียนรู้ของเน็ตเวิร์กนั้น จะทำการปรับค่าน้ำหนักเพื่อที่จะลดค่าผลลัพธ์ของฟังก์ชันต้นทุนนี้ ฟังก์ชันต้นทุนที่เป็นที่นิยมจะแสดงไว้ดังต่อไปนี้

ทั้งนี้สัญลักษณ์ของสมการในแต่ละข้อจะใช้ J แทนฟังก์ชันต้นทุน n คือจำนวนข้อมูลทั้งหมด ที่ใช้ในการเรียนรู้ y_i แทนผลลัพธ์จริงที่ต้องการของข้อมูลชุดที่ i และ \hat{y}_i แทนผลลัพธ์ที่ทำนายได้ของ ข้อมูลชุดที่ i

- ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Squared Error หรือ MSE)

$$J = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (8)$$

- ค่าเฉลี่ยครอสเอนโทรปีแบบทวิภาค (Binary Cross-entropy)

$$J = -\frac{1}{n} \sum_{i=1}^n y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (9)$$

การหาค่าเหมาะที่สุด (Optimization)

การเรียนรู้ของนิเวศวิทยาเน็ตเวิร์กเป็นการเรียนรู้เพื่อลดค่าของฟังก์ชันต้นทุนให้น้อยที่สุดโดยใช้ การปรับปรุงน้ำหนักของเส้นเชื่อมในเน็ตเวิร์ก⁹ สำหรับวิธีการปรับปรุงน้ำหนักของเส้นเชื่อมที่ใช้ในงาน วิทยานิพนธ์นี้ คือ สโตแคสติกเกรเดียนต์เดสเซนท์ (Stochastic Gradient Descent หรือ SGD)

⁸ อ้างอิงจาก [20]

⁹ อ้างอิงจาก [20]

กำหนดให้ w แทนพารามิเตอร์ซึ่งเป็นน้ำหนักที่ต้องการจะปรับค่า ถัดมาตัว α คือ อัตราการเรียนรู้ และ $\frac{\partial J}{\partial w}$ คือ เกรเดียนของฟังก์ชันต้นทุนเทียบกับ w

การเรียนรู้โดยสโตแคสติกเกรเดียนเดสเซนส์จะมีการปรับค่าของ w ดังสมการต่อไปนี้

$$w_t = w_{t-1} - \alpha \frac{\partial J_t}{\partial w} \quad (10)$$

สำหรับการใช้โมเมนตัม (Momentum) มีจุดประสงค์เพื่อทำให้การเรียนรู้มีการลู่เข้าที่ดีขึ้นจากการหลีกเลี่ยงการติดอยู่ที่โลคอลออปติมา (Local Optima) กำหนดให้ v แทนค่าความเร็วซึ่งมีการปรับค่าพร้อมกับ w และ γ แทนค่าสัมประสิทธิ์ของโมเมนตัม (Momentum Coefficient) เขียนสมการในการเรียนรู้ได้ดังนี้

$$v_t = \gamma v_{t-1} + \alpha \frac{\partial J_t}{\partial w} \quad (11)$$

$$w_t = w_{t-1} - v_t \quad (12)$$

2.2.3 การแพร่กระจายย้อนกลับ (Backpropagation)

ในขั้นตอนการป้อนไปข้างหน้า การหาค่าความผิดพลาดของเพอร์เซ็ปตรอนในลำดับชั้นสุดท้ายทำได้โดยคำนวณเกรเดียนของฟังก์ชันต้นทุนเทียบกับค่าผลลัพธ์ในชั้นสุดท้ายเท่านั้น แต่ในการหาค่าความผิดพลาดของเพอร์เซ็ปตรอนเพื่อใช้ในการเรียนรู้ของลำดับชั้นก่อนหน้านั้นไม่อาจหาได้โดยตรง จึงต้องอาศัยวิธีการที่เรียกว่า การแพร่กระจายย้อนกลับ¹⁰

กำหนดให้

δ_l^j คือ ค่าความผิดพลาดของเพอร์เซ็ปตรอนตัวที่ j ในลำดับชั้น l

j คือ ฟังก์ชันต้นทุน

z คือ ค่าที่คำนวณได้ก่อนจะผ่านฟังก์ชันกระตุ้น g

¹⁰ อ้างอิงจาก [20]

เขียนสมการค่าความผิดพลาดได้เป็น

$$\delta_j^l = \frac{\partial J}{\partial z_j^l} = \frac{\partial J}{\partial a_j^l} \frac{\partial a_j^l}{\partial z_j^l} = \frac{\partial J}{\partial a_j^l} g'(z_j^l) \quad (13)$$

การหาค่า $\frac{\partial J}{\partial a_j^l}$ ในลำดับชั้นสุดท้ายคำนวณหาได้โดยตรงจากฟังก์ชันต้นทุนที่เลือกใช้ สำหรับลำดับชั้นก่อนหน้าต้องหาโดยวิธีการแพร่กระจายย้อนกลับ โดยจะทำคล้ายกับการป้อนไปข้างหน้า เพียงแต่กลับทิศทางเท่านั้น โดยคำนวณได้จาก

$$\frac{\partial J}{\partial a_j^l} = \sum_{k=1}^m \frac{\partial J}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial a_j^l} = \sum_{k=1}^m \delta_k^{l+1} w_{kj}^{l+1} \quad (14)$$

กำหนดให้ m คือจำนวนเพอร์เซปตรอนในลำดับชั้นที่ $l+1$

เมื่อคำนวณค่าความผิดพลาดของแต่ละระดับชั้นได้ จะหาค่าผิดพลาดเทียบกับน้ำหนักและค่าไบแอสใดๆ ได้ดังสมการ

$$\frac{\partial J}{\partial w_{jk}^l} = \frac{\partial J}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l} = \delta_j^l a_k^{l-1} \quad (15)$$

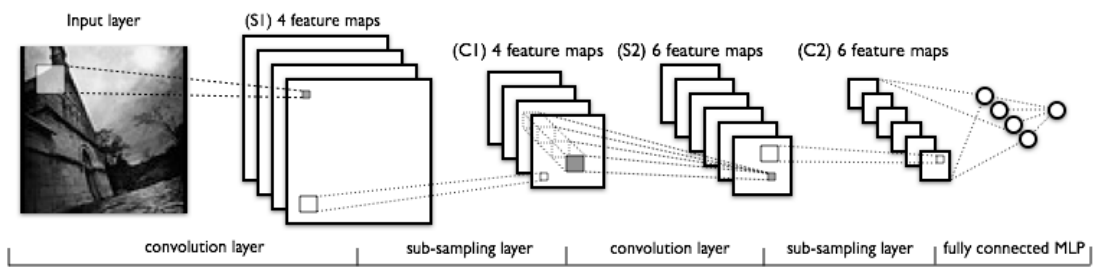
$$\frac{\partial J}{\partial b_j^l} = \frac{\partial J}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} = \delta_j^l \quad (16)$$

เนื่องจากวิทยานิพนธ์นี้ใช้สโตแคสติกเกรเดียนเดสเซนท์ ดังนั้นการปรับปรุงค่าน้ำหนัก w_{jk}^l ทำได้โดยการใช้สมการดังต่อไปนี้

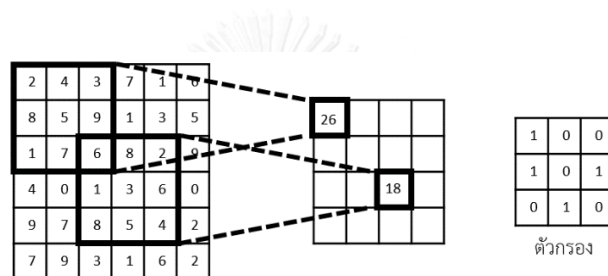
$$w_{jk,t}^l = w_{jk,t-1}^l - \alpha a_{k,t-1}^{l-1} \delta_{j,t}^l \quad (17)$$

2.3 นิเวรอลเน็ตเวิร์กคอนโวลูชัน (Convolutional Neural Network)

นิเวรอลเน็ตเวิร์กคอนโวลูชันเป็นนิเวรอลเน็ตเวิร์กเชิงลึกรูปแบบหนึ่ง มีจุดเริ่มต้นมาจากงานวิจัยทางการรู้จำภาพตัวอักษร โดยส่วนมากจะใช้ข้อมูลรับเข้าเป็นเมทริกซ์ที่แปลงมาจากรูปภาพ ซึ่งปัจจุบันรับข้อมูลนำเข้าได้ทั้งภาพ 2 มิติและ 3 มิติ โครงสร้างของนิเวรอลเน็ตเวิร์กคอนโวลูชันมีได้หลายรูปแบบจากการนำชั้นต่างๆ มาประกอบเข้าด้วยกันดังรูปที่ 6

รูปที่ 6 นิเวรอลเน็ตเวิร์กคอนโวลูชัน¹¹

2.3.1 ชั้นคอนโวลูชัน (Convolutional Layer)

รูปที่ 7 ตัวอย่างการทำคอนโวลูชัน โดยมีข้อมูลรับเข้าขนาด 6×6 และเมทริกตัวกรองขนาด 3×3 ¹²

เป็นชั้นที่ทำหน้าที่หาคูณลักษณะจากกลุ่มของข้อมูลรับเข้าที่อยู่ใกล้ๆ กัน โดยใช้วิธีการดอทเมทริกซ์กับตัวกรอง (Filter) น้ำหนักของตัวกรองนั้นจะใช้ร่วมกันในทุกๆ การทำคอนโวลูชันของข้อมูลรับเข้า (รูปที่ 7) กำหนดให้ข้อมูลรับเข้าแทนด้วยเมทริกซ์ a^{l-1} ขนาด $N \times N$ และมีตัวกรองที่มีน้ำหนัก w ขนาด $m \times m$ ผลลัพธ์ a^l ของการทำคอนโวลูชันคำนวณได้ดังสมการต่อไปนี้

$$z_{ij}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{a,b}^l a_{i+a,j+b}^{l-1} + b^l \quad (18)$$

$$a_{ij}^l = g(z_{ij}^l) \quad (19)$$

ในชั้นคอนโวลูชัน มีองค์ประกอบที่ใช้ดังต่อไปนี้

¹¹ อ้างอิงจาก Fig. 5 ใน [20]

¹² อ้างอิงจาก Fig. 2.5 ใน [20]

2.3.2 ขนาดของตัวกรอง (Filter Size)

คือ ความกว้างและความสูงของตัวกรองที่จะนำมาใช้ในการทำคอนโวลูชัน

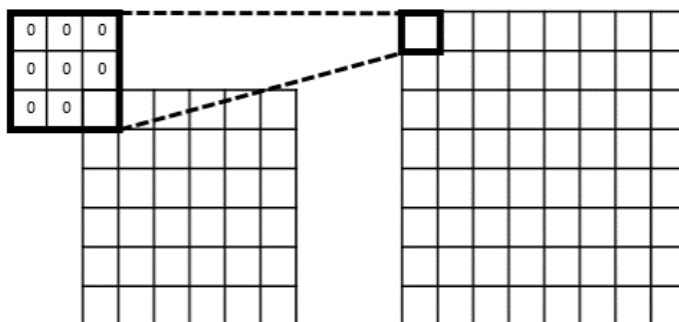
2.3.3 ชนิดของการทำคอนโวลูชัน (Convolution Type)

- คอนโวลูชันแบบแคบ (Narrow Convolution)

การทำคอนโวลูชันโดยทั่วไป มักจะเป็นการทำคอนโวลูชันแบบแคบ ในการทำคอนโวลูชัน ตัวกรองที่นำมาทำการต่อเมทริกซ์นั้นจะไม่มีผลกระทบเลยขอบของเมทริกซ์รับเข้า ส่งผลให้ผลลัพธ์ของการทำคอนโวลูชันที่มีข้อมูลรับเข้าขนาด $N \times N$ กับตัวกรองขนาด $M \times M$ จะได้เมทริกซ์ขนาด $(N-M+1) \times (N-M+1)$

- คอนโวลูชันแบบกว้าง (Wide Convolution)

เป็นการทำคอนโวลูชันที่มีการกระทำเลยขอบของเมทริกซ์รับเข้าออกไป โดยพื้นที่ที่เกินออกไปนั้น จะมีการแทนค่าของข้อมูลช่องนั้นๆ ด้วย 0 เรียกว่า การเสริมเติม (padding) ผลลัพธ์ของการทำคอนโวลูชันแบบกว้างที่มีข้อมูลรับเข้าขนาด $N \times N$ กับตัวกรองขนาด $M \times M$ จะได้เมทริกซ์ขนาด $(N+M-1) \times (N+M-1)$ ทั้งนี้การทำคอนโวลูชันแบบกว้างมีขึ้นเพื่อป้องกันการสูญเสียข้อมูลตรงบริเวณขอบของข้อมูลรับเข้าดังรูปที่ 8

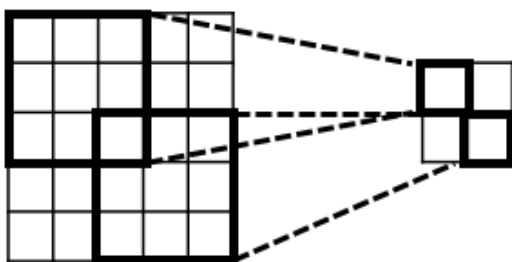


รูปที่ 8 การทำคอนโวลูชันแบบกว้างและการเสริมเติม¹³

¹³ อ้างอิงจาก Fig. 2.6 ใน [20]

- **ขนาดของการก้าวข้าม (Stride Size)**

ขนาดของการก้าวข้าม คือจำนวนช่องของข้อมูลรับเข้าที่จะเลื่อนไปเมื่อทำการหาผลลัพธ์ของคอนโวลูชันในแต่ละครั้ง โดยทั่วไปขนาดของการก้าวข้ามจะมีค่าเท่ากับหนึ่งหรือสอง ดังรูปที่ 9

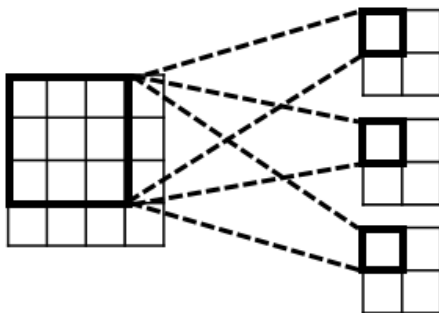


รูปที่ 9 การทำคอนโวลูชันโดยมีข้อมูลรับเข้าขนาด 5x5 ตัวกรองขนาด 3x3 และมีขนาดของการก้าวข้ามเป็น 2^{14}

- **จำนวนตัวกรอง (Number of Filters)**

ในการดำเนินการสำหรับแต่ละชั้นคอนโวลูชันมีตัวกรองได้มากกว่าหนึ่งตัว และน้ำหนักของตัวกรองแต่ละตัวอาจต่างกันได้ การกำหนดจำนวนตัวกรองในชั้นคอนโวลูชันใดๆ จะเป็นการกำหนดจำนวนช่องสัญญาณ (Channel) ของข้อมูลรับเข้าในชั้นถัดไปดังรูปที่ 10

CHULALONGKORN UNIVERSITY



รูปที่ 10 การทำคอนโวลูชันโดยมีจำนวนตัวกรองเท่ากับ 3^{15}

¹⁴ อ้างอิงจาก Fig. 2.7 ใน [20]

¹⁵ อ้างอิงจาก Fig. 2.8 ใน [20]

- จำนวนช่องสัญญาณ (Channel)

จำนวนช่องสัญญาณ หรือเรียกได้อีกอย่างหนึ่งว่า ความลึกของข้อมูลรับเข้า มีค่าได้มากกว่าหนึ่งค่า ยกตัวอย่างเช่น ในงานวิจัยทางการใช้รูปภาพมีการใช้ช่องสัญญาณทั้งหมด 3 ช่องสัญญาณแทนค่าของแม่สี หรือเกิดจากจำนวนของตัวกรองในชั้นคอนโวลูชันก่อนหน้า กำหนดให้จำนวนช่องสัญญาณมีค่าเป็น k เขียนสมการในการคำนวณผลลัพธ์ของชั้นคอนโวลูชัน¹⁶ ได้ดังสมการต่อไปนี้

$$z_{ij}^l = \sum_{c=0}^{k-1} \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{a,b}^l a_{c,i+a,j+b}^{l-1} + b^l \quad (20)$$

$$a_{ij}^l = g(z_{ij}^l) \quad (21)$$

- การแพร่กระจายย้อนกลับ (Backpropagation)

เป็นไปในลักษณะเดียวกับนิรอรลเน็ตเวิร์กโดยทั่วไป คือ การหาค่าความผิดพลาดเทียบกับค่า z_{ij}^l ในระดับชั้นใดๆ จะทำได้จาก

$$\delta_{ij}^l = \frac{\partial J}{\partial z_{ij}^l} = \frac{\partial J}{\partial a_{ij}^l} \frac{\partial a_{ij}^l}{\partial z_{ij}^l} = \frac{\partial J}{\partial a_{ij}^l} g'(z_{ij}^l) \quad (22)$$

และหา $\frac{\partial J}{\partial a_{ij}^l}$ ได้จากการแพร่กระจายย้อนกลับดังสมการต่อไปนี้

$$\begin{aligned} \frac{\partial J}{\partial a_{ij}^l} &= \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial J}{\partial z_{i-a,j-b}^{l+1}} \frac{\partial z_{i-a,j-b}^{l+1}}{\partial a_{ij}^l} \\ &= \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \delta_{i-a,j-b}^{l+1} w_{a,b}^{l+1} \end{aligned} \quad (23)$$

โดยที่ m คือ ขนาดของตัวกรอง

¹⁶ อ้างอิงจาก [20]

ถัดไปเมื่อกำหนดค่าความผิดพลาดของแต่ละระดับชั้นได้ จะหาค่าผิดพลาดเทียบกับน้ำหนักและค่าไบแอสใดๆ ได้จากสมการดังต่อไปนี้

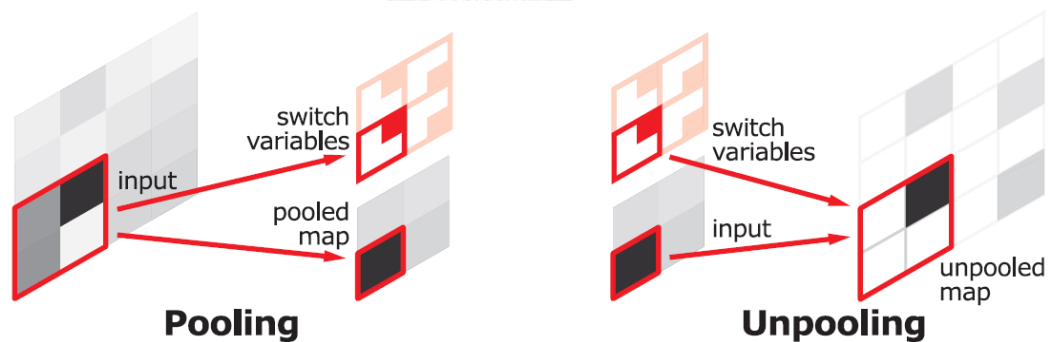
$$\frac{\partial J}{\partial w_{ab}^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial J}{\partial z_{ij}^l} \frac{\partial z_{ij}^l}{\partial w_{ab}^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \delta_{ij}^l a_{i+a, j+b}^{l-1} \quad (24)$$

$$\frac{\partial J}{\partial b^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial J}{\partial z_{ij}^l} \frac{\partial z_{ij}^l}{\partial b^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \delta_{ij}^l \quad (25)$$

โดย N คือขนาดของข้อมูลรับเข้าในชั้นที่ l

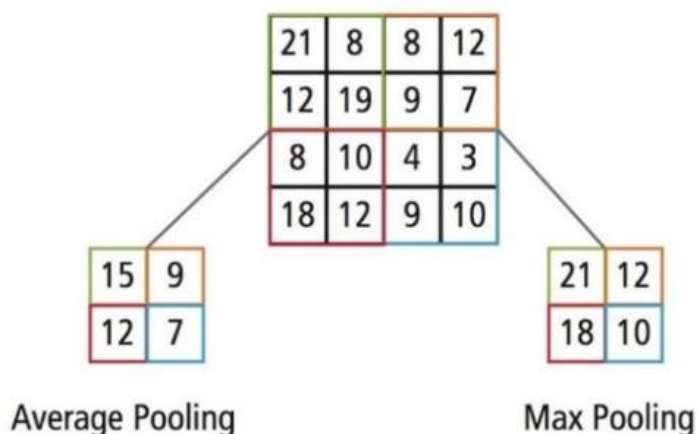
2.3.4 ชั้นพูลลิ่ง (Pooling Layer) และชั้นอันพูลลิ่ง (Unpooling Layer)

สำหรับชั้นพูลลิ่งเป็นชั้นที่ใช้เพื่อทำการลดขนาดของรูปภาพที่ผ่านการคอนโวลูชันมาแล้ว โดยทั่วไปจะนิยมนำมาต่อกับชั้นคอนโวลูชันทันที และสำหรับชั้นอันพูลลิ่งเป็นชั้นที่ใช้เพื่อทำการเพิ่มขนาดของรูปภาพที่ผ่านการคอนโวลูชันมาแล้ว โดยแสดงวิธีการทำพูลลิ่งและอันพูลลิ่งไว้ดังรูปที่ 11 ถัดมาเป็นการแสดงการทำพูลลิ่งโดยการเลือกค่ามากที่สุดซึ่งเป็นชนิดที่นิยมนำมาใช้มากที่สุด และการทำพูลลิ่งโดยใช้ค่าเฉลี่ยไว้ดังรูปที่ 12



รูปที่ 11 การทำชั้นพูลลิ่งและชั้นอันพูลลิ่ง¹⁷

¹⁷ อ้างอิงจาก Fig. 3 ใน [14]



รูปที่ 12 ชั้นการรวมโดยใช้ค่าเฉลี่ย (ซ้าย) และค่ามากที่สุด (ขวา)¹⁸

2.3.5 ชั้นการเชื่อมโยงเต็มรูปแบบ (Fully Connected Layer)

หลังจากการประกอบกันของชั้นคอนโวลูชันและชั้นพูลลิ่งจำนวนหนึ่งแล้ว ในขั้นสุดท้ายของนิเวศน์เน็ตเวิร์กคอนโวลูชัน จะเป็นการเชื่อมโยงเต็มรูปแบบ กล่าวคือ ในขั้นนี้ประกอบด้วยชั้นย่อยๆ ที่มีเพอร์เซ็ปตรอนอยู่จำนวนหนึ่ง โดยที่เพอร์เซ็ปตรอนแต่ละตัว จะมีเส้นเชื่อมกับเพอร์เซ็ปตรอนทุกตัวในชั้นก่อนหน้า และเพอร์เซ็ปตรอนทุกตัวในชั้นถัดไป ทั้งนี้การคำนวณการป้อนไปข้างหน้าและการแพร่กระจายย้อนกลับทำได้ด้วยวิธีการปกติ [20] ในปัจจุบันมีวิธีที่นิยมนำมาใช้เพื่อป้องกันปัญหาการเกิดโอเวอร์ฟิต (Overfitting) กับชั้นการเชื่อมโยงเต็มรูปแบบ เรียกว่า วิธีดรอพเอาท์ (Dropout) ซึ่งเป็นวิธีที่ใช้การกำหนดค่าความน่าจะเป็นแบบแบร์นูลลี (Bernoulli) ให้กับแต่ละเส้นเชื่อมของค่าน้ำหนัก (Weight)

สำหรับการเรียนรู้ของวิธีดรอพเอาท์ แบ่งออกเป็น 2 ขั้นตอน ได้ดังนี้

(1) **ขั้นตอนการเรียนรู้** ในแต่ละการเชื่อมต่อของค่าน้ำหนักจะมีการสุ่มค่าตามความน่าจะเป็นที่ผู้วิจัยกำหนดในขั้นตอนการเรียนรู้ซึ่งน้ำหนักแต่ละค่าจะถูกสุ่มค่าระหว่าง 0 หรือ 1 ถ้าน้ำหนักมีค่าเป็น 0 จะไม่นำไปใช้ในการเรียนรู้

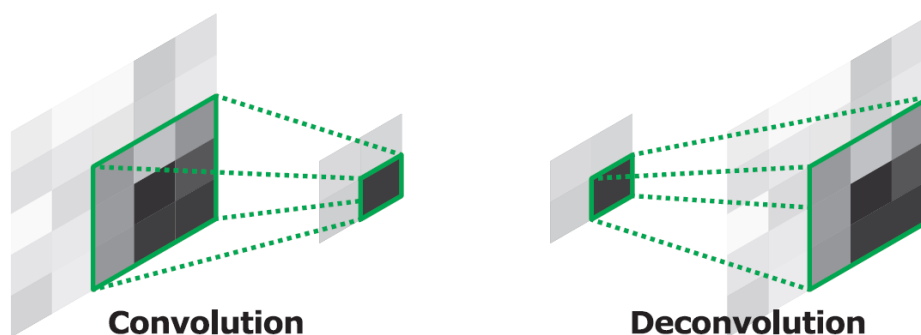
(2) **ขั้นตอนการทดสอบ** ค่าน้ำหนักในสถาปัตยกรรมทุกค่าจะถูกคูณด้วยค่าความน่าจะเป็นที่กำหนด เช่น กำหนดค่าความน่าจะเป็นเท่ากับ 0.5 เป็นต้น

วิธีดรอพเอาท์ได้รับการพิสูจน์จากงานวิจัยหลายๆ งานแล้วว่า ป้องกันการเกิดปัญหาโอเวอร์ฟิตและเพิ่มประสิทธิภาพการจำแนกในการเรียนรู้ของเครื่องได้ [22]

¹⁸ อ้างอิงจาก <http://www.embedded-vision.com/platinum-members/cadence/embedded-vision-training/documents/pages/neuralnetworksimagerecognition>

2.3.6 ชั้นดีคอนโวลูชัน (Deconvolutional Layer)

เป็นชั้นที่จะมีลักษณะเดียวกันกับการทำคอนโวลูชัน โดยมีส่วนประกอบที่สำคัญได้แก่ เคอร์เนลฟิลเตอร์ (Kernel Filter) ซึ่งในขั้นตอนการเรียนรู้เริ่มต้น เคอร์เนลฟิลเตอร์แต่ละเคอร์เนลจะถูกกำหนดโดยการสุ่มค่าเพื่อนำไปคูณกับข้อมูลนำเข้า เรียกว่า ฟีเจอร์แมพ (Feature Map) หลังจากนั้นจะมีการปรับค่าโดยการเรียนรู้แบบแพร่กระจายย้อนกลับ (Backpropagation) จำนวนผลลัพธ์ที่ได้จากชั้นดีคอนโวลูชันจะเท่ากับจำนวนผลลัพธ์จากชั้นคอนโวลูชัน รูปที่ 13 แสดงวิธีการทำคอนโวลูชันและดีคอนโวลูชัน



รูปที่ 13 การทำชั้นคอนโวลูชันและชั้นดีคอนโวลูชัน¹⁹

2.4 การวัดประสิทธิภาพ (Performance Evaluation)

การวัดประสิทธิภาพการจำแนกแบบสองคลาส (Binary-Class Classification) แสดงได้ดังนี้

2.4.1 คอนฟิวชันเมทริกซ์ (Confusion Matrix)

ตารางที่ 1 คอนฟิวชันเมทริกซ์ของการจำแนกแบบ 2 คลาส

		คลาสที่ทำนาย	
		ถนน	ไม่ใช่ถนน
คลาสจริง	ถนน	TP	FN
	ไม่ใช่ถนน	FP	TN

¹⁹ อ้างอิงจาก Fig. 3 ใน [14]

คอนฟิวชันเมตริกซ์ของการจำแนกจากข้อมูลทั้งหมด 2 คลาส ประกอบด้วย คลาสที่ตอบว่าเป็นถนนและคลาสที่ตอบว่าไม่ใช่ถนน ค่าในแต่ละแถวแสดงจำนวนข้อมูลที่มีคลาสนั้นเป็นคำตอบที่ถูกต้อง ส่วนค่าในแต่ละหลักแสดงจำนวนข้อมูลที่ทำนายได้คลาสนั้น กำหนดให้ สำหรับคลาสใดๆ

- (1) TP คือ จำนวนข้อมูลที่ทำนายได้คลาสนั้นและผลลัพธ์คือคลาสนั้น (True Positive)
- (2) FP คือ จำนวนข้อมูลที่ทำนายได้คลาสนั้นแต่ผลลัพธ์คือคลาสอื่น (False Positive)
- (3) TN คือ จำนวนข้อมูลที่ทำนายได้คลาสอื่นและผลลัพธ์คือคลาสอื่น (True Negative)
- (4) FN คือ จำนวนข้อมูลที่ทำนายได้คลาสอื่นแต่ผลลัพธ์คือคลาสนั้น (False Negative)

2.4.2 ตัววัดประสิทธิภาพจำแนกโดยคลาส

การคำนวณหาค่า Precision Recall และ F1 คำนวณได้จาก

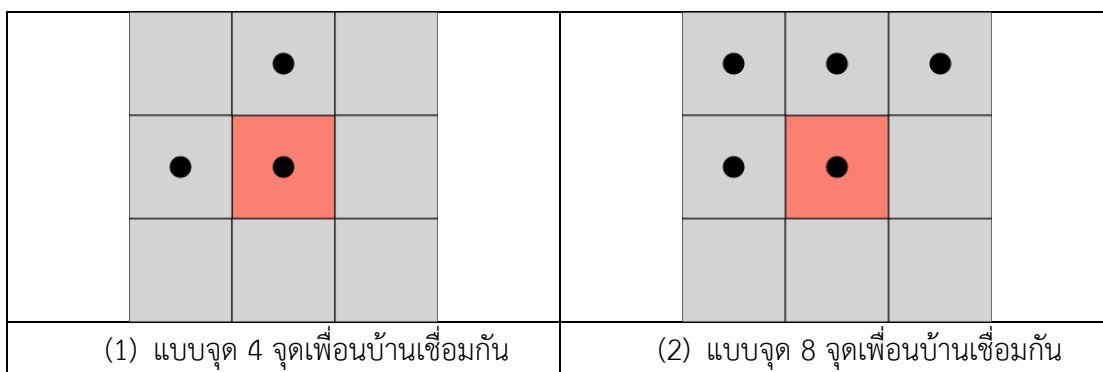
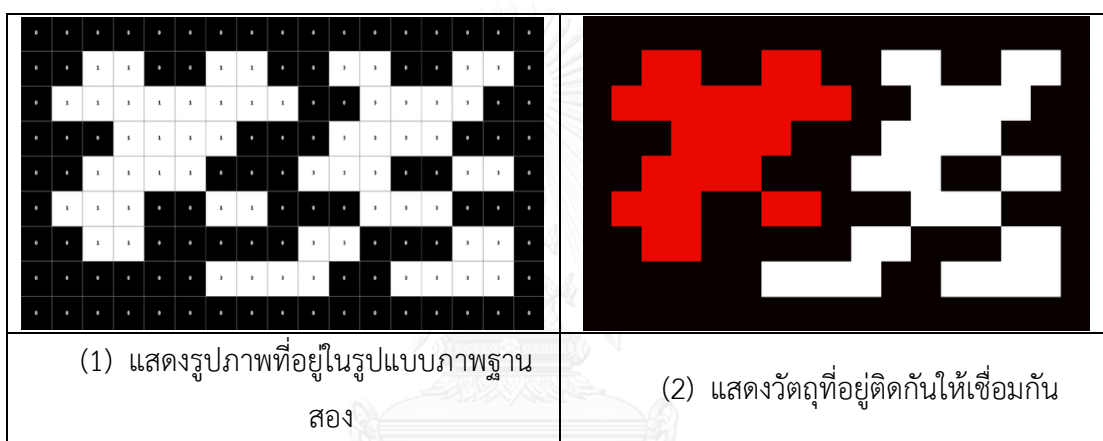
$$Precision = \frac{TP}{TP + FP} \quad (26)$$

$$Recall = \frac{TP}{TP + FN} \quad (27)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (28)$$

2.5 การกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน (Connected-Component Labeling)

วิธีนี้เป็นวิธีสำหรับนำมาพิจารณาวัตถุที่มีส่วนเชื่อมกันให้นับรวมกันเป็นหนึ่งวัตถุ โดยวิธีการที่นิยมนำมาพิจารณาว่า วัตถุที่อยู่ใกล้กันจะนับรวมเป็นวัตถุเดียวกันได้ คือ การพิจารณาวัตถุที่อยู่รอบวัตถุที่สนใจแบบ 4 จุดเพื่อนบ้าน และ 8 จุดเพื่อนบ้านตามลำดับ โดยจะแสดงตำแหน่งเพื่อนบ้านสำหรับพิจารณาดังรูปที่ 14 และตัวอย่างสำหรับการกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน (Connected-Component Labeling) ดังรูปที่ 15 ซึ่งวิธีประเภทนี้เป็นวิธีที่นิยมใช้กันมากในกลุ่มของคอมพิวเตอร์วิทัศน์ (Computer Vision)

รูปที่ 14 การพิจารณาแบบ 4 จุดเพื่อนบ้าน และ 8 จุดเพื่อนบ้านตามลำดับ²⁰รูปที่ 15 ตัวอย่างภาพการกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน²¹

2.6 การเพิ่มความราบเรียบโดยใช้ตัวกรองเกาส์เซียน (Smoothing with Gaussian Filter)

ในรูปถนนวนจะมีวัตถุที่อยู่บนถนนปะปนอยู่มาก เช่น รถยนต์ เงาของต้นไม้ เสาไฟฟ้า เป็นต้น วัตถุเหล่านี้เสมือนเป็นสิ่งรบกวน (Noise) บนท้องถนน ในวิทยานิพนธ์นี้ได้นำตัวกรองเกาส์เซียนมาใช้เพื่อกำจัดสิ่งรบกวนเหล่านี้ เนื่องจากตัวกรองเกาส์เซียนช่วยเพิ่มความราบเรียบให้กับวัตถุได้ด้วยการเบลอ (Blur) และปรับค่าพารามิเตอร์ σ ได้ดังสมการที่ 29

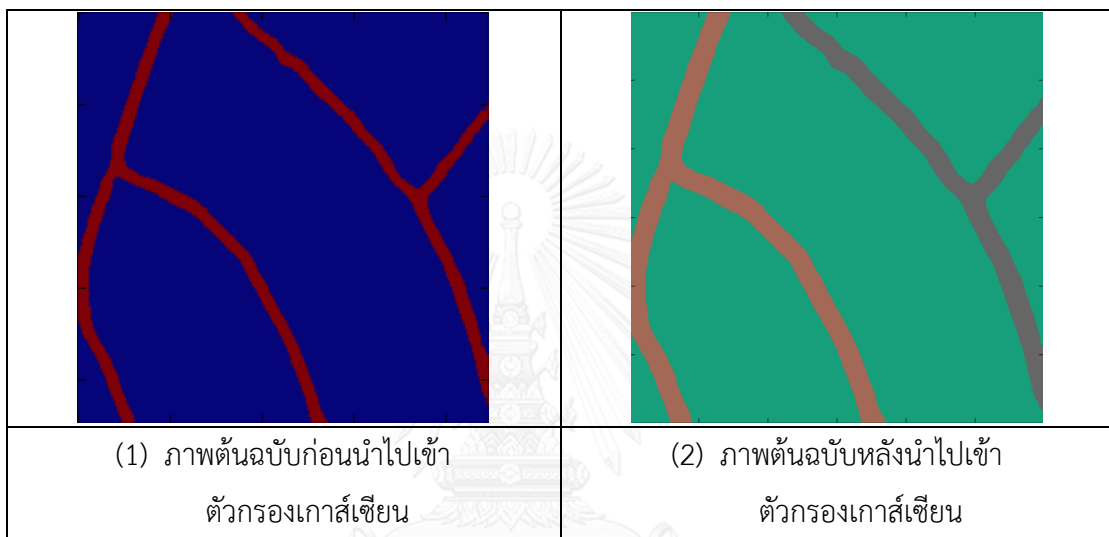
²⁰ อ้างอิงจาก https://en.wikipedia.org/wiki/Connected-component_labeling

²¹ อ้างอิงจาก https://en.wikipedia.org/wiki/Connected-component_labeling

$$S_{i,j} = I(i,j) * G(i,j,\sigma) \quad (29)$$

โดย $S_{i,j}$ คือ ภาพภายหลังการกรอง $I(i,j)$ คือ ภาพที่ต้องการหาขอบ และ $G(i,j,\sigma)$ คือ ฟังก์ชันเกาส์เซียนที่มีค่าเบี่ยงเบนมาตรฐานของการกระจาย σ ผลลัพธ์เบื้องต้นที่ได้จากการทดลองกับภาพถ่ายแสดงไว้ดัง

รูปที่ 16



รูปที่ 16 ตัวอย่างภาพการเพิ่มความราบเรียบให้กับวัตถุโดยใช้ตัวกรองเกาส์เซียน

2.7 การทำให้เป็นค่านอร์มัลไลเซชัน (Normalization)

ภาพถ่ายระยะไกลที่ใช้ในวิทยานิพนธ์นี้มีช่องสัญญาณ 3 ช่องสัญญาณ คือช่องสัญญาณสีแดง, ช่องสัญญาณสีเขียว และช่องสัญญาณสีน้ำเงิน ในงานวิจัยนี้คำนวณค่านอร์มัลไลเซชันของภาพถ่ายระยะไกลด้วยการปรับค่าแต่ละช่องสัญญาณจากช่วง $[0, 255]$ ให้อยู่ในช่วง $[0, 1]$ โดยการนำแต่ละช่องสัญญาณมาลบด้วยค่า 128 และหารด้วย 128 ตามลำดับ ในกรณีของภาพเฉลยจะมีเพียง 1 ช่องสัญญาณ คือ ช่องสัญญาณสีเทาซึ่งอยู่ในช่วง $[0, 255]$ งานวิจัยนี้นำช่องสัญญาณสีเทามาหารด้วย 255 เพื่อให้ช่องสัญญาณมี 2 ค่า คือ 0 และ 1 โดย 0 คือ พิกเซลที่ไม่ใช่ถนน และ 1 คือ พิกเซลที่เป็นถนนตามลำดับ

2.8 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส (Conditional Random Fields: CRFs)

ในปัจจุบันแบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส (CRFs) เป็นแบบจำลองที่นักวิจัยยอมรับว่ามีประสิทธิภาพมากกว่าแบบจำลอง Hidden Markov Models (HMMs) CRFs เป็น Generative Models ที่อาศัยค่าความน่าจะเป็นร่วม (Joint Probability) ระหว่างข้อมูลภาพถ่ายกับผลที่ได้หลังจากการจำแนก หรือข้อมูลภาพถ่ายกับภาพผลเฉลย งานวิจัยที่ผ่านมาได้มีการนำ CRFs มาประยุกต์ใช้ในส่วนการประมวลผลหลัง เพื่อกำจัดสิ่งรบกวนหลังการจำแนกจากการเรียนรู้เชิงลึก (Smooth Noisy Segmentation Maps) ในวิทยานิพนธ์นี้จึงเลือกโมเดล CRFs เข้ามาช่วยทำให้วัตถุถนนที่เป็นผลลัพธ์จากโมเดลการเรียนรู้เชิงลึกมีความคมและความถูกต้องมากขึ้น (Road Object Sharpening) เช่นเดียวกัน

โมเดล CRFs ประกอบไปด้วยฟังก์ชันพลังงาน (Energy Function) ซึ่งเป็นตัวหาค่าความน่าจะเป็นร่วมระหว่างข้อมูลภาพถ่ายกับผลที่ได้หลังจากการจำแนก หรือข้อมูลภาพถ่ายกับภาพผลเฉลย ฟังก์ชันพลังงาน (Energy Function) ของ CRFs แสดงได้ดังสมการที่ 30

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \quad (30)$$

โดยที่ x คือ ผลเฉลยของแต่ละพิกเซล โดยที่สมการ Unary Potential คือ $\theta_i(x_i) = -(\log P(x_i))$ ขณะที่ $P(x_i)$ คือ ค่าความน่าจะเป็นของผลเฉลยของพิกเซลที่ i ที่คำนวณจากการเรียนรู้เชิงลึก

สมการ Pairwise Potential กระจายได้เป็น $\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K w_m k^m(f_i, f_j)$ โดยที่ $\theta_{ij}(x_i, x_j) = 1$ และ $x_i \neq x_j$ โดยเมื่อรวมกับ Kernel ที่ประยุกต์มาจากตำแหน่งทวิภาคี (Bilateral Position) และเงื่อนไขของสี (Color Terms) แสดงสมการเต็มรูปแบบได้ดังสมการที่ 31

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right) \right] \quad (31)$$

โดยกำหนดให้ p คือ ตำแหน่งของพิกเซล และ I คือตำแหน่งของสี RGB และไฮเปอร์พารามิเตอร์ σ_α σ_β และ σ_γ เป็นตัวควบคุมค่าสเกลของเกาส์เซียนเคอร์เนล (Gaussian Kernels)

2.9 ภาพถ่ายระยะไกล (Remotely-Sensed Images)

ภาพถ่ายระยะไกลสำหรับงานวิจัยชิ้นนี้ แบ่งออกเป็น 2 ชนิด ได้แก่ ภาพถ่ายที่มีคุณสมบัติเป็นภาพถ่ายที่มีความละเอียดสูงมาก (Very High Resolution Images) บันทึกภาพโดยดาวเทียม (Satellite) และภาพถ่ายที่มีความละเอียดสูง (High Resolution Images) บันทึกภาพโดยกล้องที่อยู่บนอากาศยาน ภาพสองชนิดมีความคมชัดแตกต่างกันเนื่องจากระยะทางที่บันทึกภาพต่างกัน

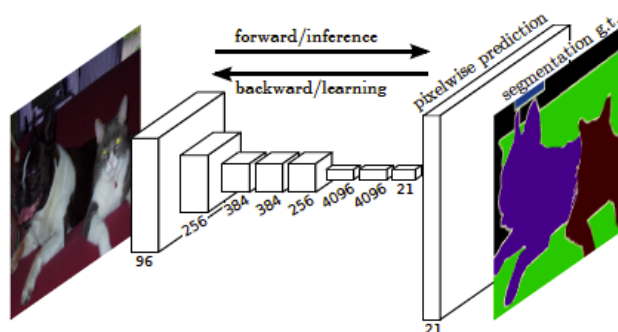


บทที่ 3 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องในวิทยานิพนธ์นี้แบ่งออกเป็น 3 กลุ่ม ได้แก่ งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกทางความหมาย (Deep Learning for Semantic Segmentation) งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกถนนทางความหมาย (Deep Learning for Semantic Road Segmentation) และงานวิจัยเกี่ยวกับเทคนิคใหม่สำหรับการเรียนรู้เชิงลึก (Recent Techniques in Deep Learning)

3.1. งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกทางความหมาย (Deep Learning for Semantic Segmentation)

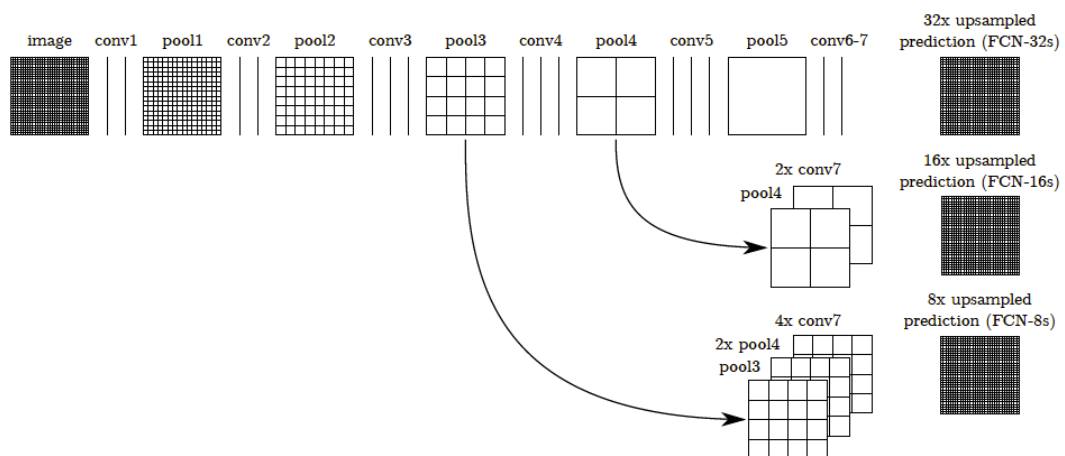
ปัจจุบันการจำแนกภาพทางความหมายเป็นที่นิยมทำกันมากโดยใช้การเรียนรู้เชิงลึกเข้ามาช่วยในการจำแนก ในปี 2015 J. Long, E. Shelhamer, และ T. Darrell [10] ได้นำเสนอวิธีการเรียนรู้เชิงลึกเพื่อนำมาจำแนกภาพทางความหมายโดยใช้โครงข่ายประสาทเทียมคอนโวลูชันเชิงลึกแบบฟูลลี (FCN) ซึ่งเน็ตเวิร์กที่ใช้จะเหมือนกับโครงข่ายประสาทเทียมคอนโวลูชันเชิงลึก (Deep Convolutional Neural Network) โดยทั่วไป เพียงแต่ขั้นสุดท้ายหลังจากชั้นการจำแนก (Classification Layer) คำตอบจะกระจายไปบนทุกพิกเซลบนภาพ (Pixel-wise Class Prediction Map) โดยใช้วิธีการอัพแซมปิง (Upsampling) เพิ่มเข้ามาอีกหนึ่งชั้นหลังจากชั้น Fully Connected ซึ่งแสดงได้ดังรูปที่ 17



รูปที่ 17 รูปเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี²²

²² อ้างอิงจาก Fig. 1 ใน [10]

จากนั้น J. Long, E. Shelhamer, และ T. Darrell ได้ทำการเสนอการข้ามชั้นเลเยอร์ของเน็ตเวิร์กในส่วนของการทำการสกัดคุณลักษณะสำคัญ โดยเสนอวิธีการข้ามเลเยอร์ถึง 3 รูปแบบ ดังรูปที่ 18 คือข้ามเลเยอร์ที่ 4 , 5 และ 6 ส่งไปทำการจำแนกที่ชั้น Fully Connected และทำการตอบแบบกระจายทุกพิกเซลบนภาพ เรียกว่า FCN-8s ถัดมา คือข้ามเลเยอร์ที่ 5 และ 6 ส่งไปทำการจำแนกที่ชั้น Fully Connected และทำการตอบแบบกระจายทุกพิกเซลบนภาพเรียกว่า FCN-16s และสุดท้ายคือการทำครบทุกเลเยอร์เรียกว่า FCN-32s โดยสิ่งที่เขานำเสนอถูกนำไปทดสอบกับชุดข้อมูล PASCAL VOC [24] พบว่า FCN-8s ให้ค่าความแม่นยำเฉลี่ย (Mean Accuracy) สูงถึง 75.9 ซึ่งประสิทธิภาพที่ได้สูงกว่าวิธีการ FCN-16s และ FCN-32s ตามลำดับ และเมื่อนำไปเทียบกับวิธีการมาตรฐาน (Baseline) พบว่าวิธีการ FCN-8s มีประสิทธิภาพสูงที่สุด

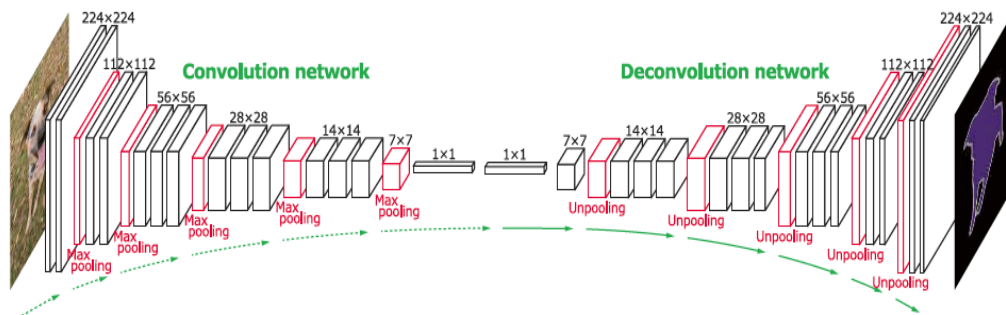


รูปที่ 18 การข้ามชั้นเลเยอร์ของเน็ตเวิร์กในส่วนของการสกัดคุณลักษณะสำคัญของเน็ตเวิร์ก FCN²³

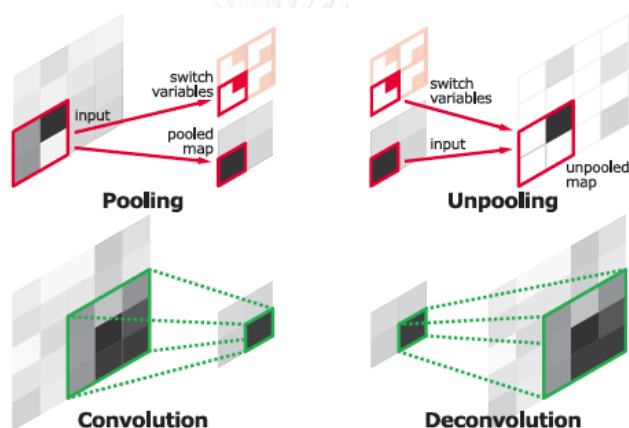
ถัดมาในปี 2015 H. Noh, S. Hong, และ B. Han [14] ได้เสนอวิธีการเรียนรู้เชิงลึกเพื่อนำมาจำแนกภาพทางความหมายโดยใช้การเรียนรู้เชิงลึกแบบดีคอนโวลูชัน (Deconvolution Network) โดยจะแบ่งเน็ตเวิร์กออกเป็น 2 ฝั่ง คือ ฝั่งแรกจะเป็นการนำเน็ตเวิร์ก VGG16 [25] มาทำในส่วนคอนโวลูชันเน็ตเวิร์กต่างๆ ไป โดยที่ฝั่งที่สองจะเป็นการทำ Upsampling (ดีคอนโวลูชัน) ด้วยเน็ตเวิร์ก VGG16 เช่นกัน เพื่อกลับขึ้นไปตอบแบบการกระจายทุกพิกเซลบนภาพ (Pixel-wise Class Prediction Map) โดยรูปของเน็ตเวิร์กทั้งหมดจะถูกแสดงไว้ยังรูปที่ 19 และรูปแบบของการทำดีคอนโวลูชันและอันพูลลิ่ง (Unpooling) จะแสดงไว้รูปที่ 20 โดยสิ่งที่นำเสนอ H. Noh, S. Hong, และ B. Han ถูกนำไปใช้เทียบประสิทธิภาพกับชุดข้อมูลที่เป็นการแข่งขันในรายการ PASCAL VOC [24]

²³ อ้างอิงจาก Fig. 3 ใน [10]

พบว่าประสิทธิภาพสิ่งที่พวกเขานำเสนอนั้นดีกว่าวิธีการอย่าง FCN-8s [10] และ DeepLab-CRF [16]



รูปที่ 19 เน็ตเวิร์กการเรียนรู้เชิงลึกแบบดีคอนโวลูชัน (Deconvolution network)²⁴



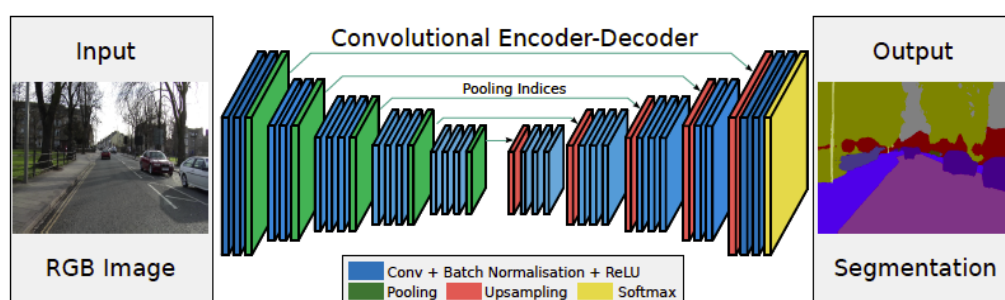
รูปที่ 20 รูปแบบของการทำดีคอนโวลูชันและอันพูลลิง (Unpooling)²⁵

ในช่วงปลายปี 2015 V. Badrinarayanan, A. Kendall, R. Cipolla และ S. Member [4, 11, 12] ได้เสนอวิธีการเรียนรู้เชิงลึกเพื่อนำมาจำแนกภาพทางความหมายโดยใช้การเรียนรู้เชิงลึกรูปแบบใหม่โดยการเลียนแบบดีคอนโวลูชัน (Deconvolution network) โดยตัดขั้นตอนการจำแนก (Classification Layer) คือชั้น Fully Connected ออก โดยเหลือเพียงแค่ส่วนการทำารสกัดคุณลักษณะสำคัญ (Feature Extraction) ทั้ง 2 ฝั่งเน็ตเวิร์ก โดยเรียกการจำแนกภาพทางความหมายโดยใช้การเรียนรู้เชิงลึกแบบนี้ว่า เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดเดอร์ (Deep Convolutional Encoder-Decoder Networks หรือ DCED) โดยใช้ชื่อเน็ตเวิร์กที่เขานำเสนอว่า

²⁴ อ้างอิงจาก Fig. 2 ใน [14]

²⁵ อ้างอิงจาก Fig. 3 ใน [14]

SegNet ดังรูปที่ 21 เมื่อเปรียบเทียบกับรูปที่ 19 จากการนำเสนอของ H. Noh, S. Hong, และ B. Han [5] จะพบว่าจากเดิมฝั่ง Convolution Network ที่ถูกตัดชั้นการจำแนกออกจะเหลือเพียง 13 ชั้น จากเดิม 16 ชั้น จะเรียกฝั่งนี้ใหม่ว่า ฝั่งเอนโคดเดอร์ (Encoder) และฝั่ง Deconvolution Network ที่ถูกตัดชั้นการจำแนกออกจะเหลือเพียง 13 ชั้น จากเดิม 16 ชั้น จะเรียกฝั่งนี้ใหม่เช่นกันว่า ฝั่งดีโคดเดอร์ (Decoder) ซึ่งข้อดีของการตัดชั้นการจำแนก คือ ลดจำนวนพารามิเตอร์ในเน็ตเวิร์ก จาก 134 ล้านพารามิเตอร์เหลือเพียง 14.7 ล้านพารามิเตอร์ หลังจากสิ้นสุดชั้นของการสกัดคุณลักษณะสำคัญจากฝั่งดีโคดเดอร์ (Decoder) แล้ว จึงจะเข้าไปสู่ชั้นการจำแนกโดยในบทความนี้เสนอการจำแนกโดยใช้ Softmax ซึ่งจะมีหน้าเป็นตัวตอบแบบการกระจายทุกพิกเซลบนภาพ โดยผลลัพธ์ที่ได้จาก SegNet จะตอบออกมาเป็นหลายคำตอบ (Multi class) สำหรับรูปเน็ตเวิร์กทั้งหมดจะแสดงไว้ที่รูปที่ 21 โดยเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ ถูกนำไปใช้เทียบประสิทธิภาพกับชุดข้อมูลที่เป็นการแข่งขันในรายการ PASCAL VOC [24] พบว่าประสิทธิภาพสิ่งที่คุณภาพนำเสนอดีกว่าวิธีที่ของ H. Noh, S. Hong, และ B. Han เสนออย่างวิธีการเรียนรู้เชิงลึกแบบดีคอนโวลูชัน

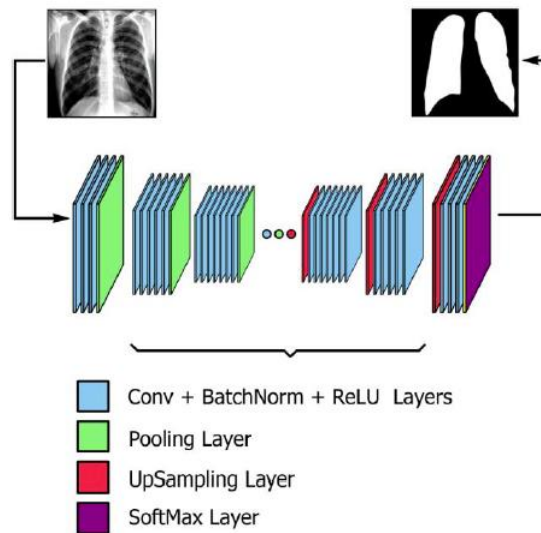


รูปที่ 21 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ (Deep Convolutional Encoder-Decoder Networks หรือ DCED)²⁶

และถัดมาในปี 2016 มีการนำเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ไปใช้จำแนกภาพทางความหมายในวงการแพทย์ โดย A. Kalinovsky, A. Kalinovsky, และ V. Kovalev [26] ได้นำเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโคดเดอร์-ดีโคดเดอร์ (รูป 21) เข้ามาจำแนกภาพทางความหมายโดยใช้ชุดข้อมูลภาพเอ็กซเรย์หน้าอก (X-Ray Chest Images) ดังรูปที่ 22 โดยในชั้นสุดท้ายหลังจากชั้นการจำแนกโดยใช้ Softmax ซึ่งจะมีหน้าเป็นตัวตอบแบบการกระจายทุกพิกเซล

²⁶ อ้างอิงจาก Fig. 2 ใน [4]

บนภาพ (Pixel-wise Class Prediction Map) สำหรับบทความนี้จะตอบออกมาเป็นสองคำตอบ (Binary Class) คือคำตอบที่เป็นปอด (พิกเซลจะเป็นสีขาว) และ คำตอบที่ไม่ใช่ปอด (พิกเซลจะเป็นสีดำ) โดยจำแนกหน้าออกออกจากภาพถ่ายเอกซเรย์ได้อย่างแม่นยำ เมื่อเปรียบเทียบประสิทธิภาพการทำงานโดยใช้ค่า Dice Score พบว่าค่าประเมินที่ได้มีค่า Dice ที่สูงมากซึ่งค่าที่ได้จากการทดสอบจะอยู่ในช่วง 0.926 ถึง 0.974



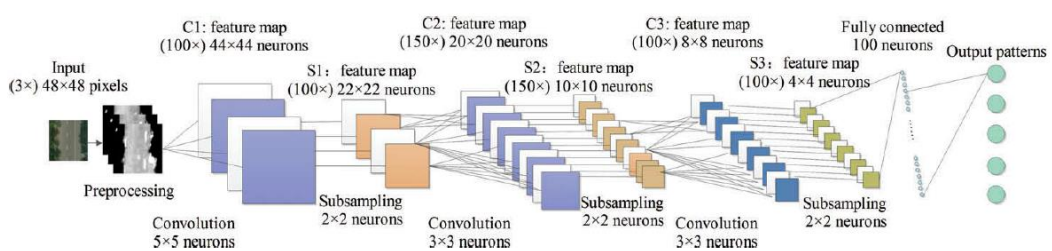
รูปที่ 22 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโค้ดเดอร์-ดีโค้ดเดอร์โดยมีข้อมูลนำเข้าเป็นภาพเอกซเรย์²⁷

จากงานวิจัยที่ผ่านมาพบว่าตั้งแต่ปี 2015 เป็นต้นมา แสดงให้เห็นว่าเริ่มมีบทความที่ใช้การเรียนรู้เชิงลึกเข้ามาช่วยในการจำแนกภาพทางความหมาย (Semantic Segmentation) โดยแบ่งชนิดของการเรียนรู้เชิงลึกสำหรับการจำแนกทางความหมายได้ 3 ชนิด ได้แก่ เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี (FCN) [10], เน็ตเวิร์กแบบดีคอนโวลูชันเชิงลึก (DeCNN) [14] และสุดท้ายเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโค้ดเดอร์-ดีโค้ดเดอร์ (DCED) [4] โดยจาก 3 เน็ตเวิร์กที่กล่าวมาข้างต้น พบว่าเมื่อนำไปประเมินประสิทธิภาพในรายการแข่งขันการจำแนกรูปภาพของ PASCAL VOC [21] พบว่าเน็ตเวิร์กคอนโวลูชันเชิงลึกแบบเอนโค้ดเดอร์-ดีโค้ดเดอร์ให้ประสิทธิภาพดีที่สุด และยังสามารถนำไปใช้กับการจำแนกภาพทางความหมายที่ต้องการตอบหลายคำตอบและต้องการตอบเพียงสองคำตอบได้อย่างมีประสิทธิภาพ

²⁷ อ้างอิงจาก Fig. 2 ใน [26]

3.2. งานวิจัยเกี่ยวกับการเรียนรู้เชิงลึกสำหรับการจำแนกถนนทางความหมาย (Deep Learning for Semantic Road Segmentation)

ตั้งแต่ปี 2015 ที่ผ่านมามีความตื่นตัวในการใช้การเรียนรู้เชิงลึกเข้ามาจำแนกภาพทางความหมายมากขึ้น โดยในส่วนของงานจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกล เริ่มมีนักวิจัยใช้วิธีการที่กล่าวไปข้างต้นเข้ามาจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกล เริ่มจากช่วงกลางปี 2015 J. Wang, J. Song, M. Chen และ Z. Yang [1] ได้เสนอเสนอการเรียนรู้เชิงลึกโดยใช้โครงข่ายประสาทเทียมคอนโวลูชันเชิงลึก (Deep Convolutional Neural Network) (แสดงรูปที่ 23) เข้ามาจำแนกรูปแบบของถนนซึ่งจะมีอยู่ 5 รูปแบบ คือ ถนนทางตรง, ถนนทางซ้าย, ถนนทางขวา, สีแยก, และทางตัน (แสดงรูปที่ 24) ซึ่งบทความนี้จะยังเป็นเพียงแต่ใช้การเรียนรู้เชิงลึกเข้ามาช่วยทำงานร่วมกับวิธีที่เขานำเสนออีกหนึ่งวิธี เรียกว่า Finite State Machine (FSM) โดยตัว FSM จะทำหน้าที่เหมือน Tracker คอยลากเส้นในส่วนที่เป็นพื้นที่ถนนบนภาพถ่ายระยะไกล โดยจะพิจารณาจากคำตอบที่ได้จากการเรียนรู้เชิงลึกว่าควรจะไปทิศทางใด ซึ่งจากการร่วมกันทำงานโดยใช้การเรียนรู้เชิงลึกร่วมกับ FSM นี้เองทำให้ผลลัพธ์ที่ได้หลังจากการประเมินประสิทธิภาพในการสกัดถนนออกจากภาพถ่ายดาวเทียมสูงกว่าวิธีการที่เป็นที่ยอมรับในการสกัดถนนออกจากภาพถ่ายระยะไกลวิธีหนึ่ง (ดีที่สุด ณ ขณะนั้นโดยใช้วิธีการเรียนรู้แบบไม่มีผู้ฝึกสอน) ที่เรียกว่า Tensor-Cuts [5] ซึ่งนำเสนอโดย C. Poullis ปี 2014 ที่ตีพิมพ์ลงในวารสาร ISPRS Journal of Photogrammetry and Remote Sensing โดยเป็นวารสารระดับนานาชาติซึ่งมีอิทธิพลสูงมากวารสารหนึ่งของวงการการสำรวจข้อมูลจากระยะไกล (Remote Sensing)



รูปที่ 23 เน็ตเวิร์กชนิดคอนโวลูชันเชิงลึกที่เสนอโดย J. Wang, J. Song, M. Chen และ Z. Yang²⁸

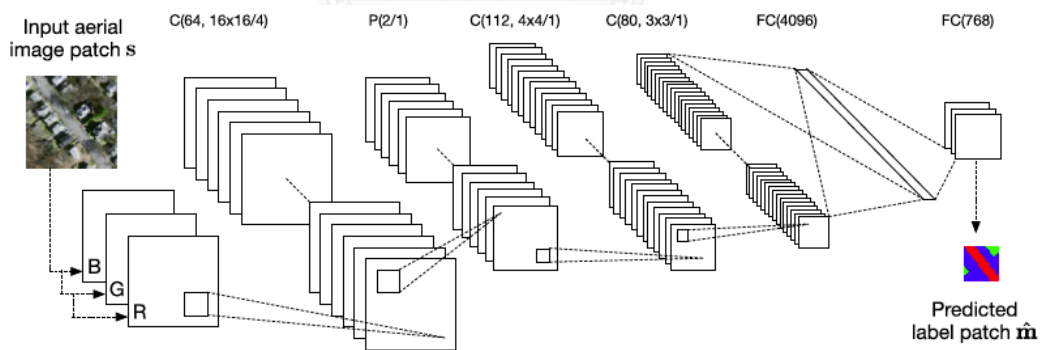
ถัดมาในปี 2016 S. Saito, T. Yamashita และ Y. Aoki [3] ได้เสนอการเรียนรู้เชิงลึกโดยใช้เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี (Fully Convolutional Networks หรือ FCN) (แสดงรูปที่

²⁸ อ้างอิงจาก Fig. 5 ใน [1]

25) เข้ามาจำแนกถนนทางความหมายออกจากภาพถ่ายทางอากาศ โดยภาพถ่ายทางอากาศนี้เป็นชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์ (Massachusetts Roads Data Set) [19] โดยข้อมูลนำเข้าจะถูกส่งเข้าไปเรียนรู้ (Training) ในเน็ตเวิร์กซึ่งจะมีขนาด 64×64 ซึ่งผลลัพธ์สุดท้ายที่ S. Saito, T. Yamashita และ Y. Aoki เสนอนั้นจะตอบออกมาอยู่ในรูปของสามคำตอบ ได้แก่ คำตอบที่เป็นถนน (พิกเซลจะแสดงสีแดง), คำตอบที่เป็นอาคาร (พิกเซลจะแสดงสีเขียว) และคำตอบที่ไม่ใช่ทั้งถนนและอาคาร (พิกเซลจะแสดงเป็นสีน้ำเงิน) โดยผลการประเมินโดยใช้ค่า F1 พบว่าอยู่ที่ 0.762



รูปที่ 24 รูปแบบของถนนทั้ง 5 รูปแบบ ได้แก่ ถนนทางตรง, ถนนทางซ้าย, ถนนทางขวา, สีแยก และ ทางตัน²⁹



รูปที่ 25 เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี (Fully Convolutional Networks หรือ FCN) ที่นำเสนอโดย S. Saito, T. Yamashita และ Y. Aoki³⁰

²⁹ อ้างอิงจาก Fig. 6 ใน [1]

³⁰ อ้างอิงจาก Fig. 1 ใน [3]

และสุดท้ายช่วงปลายปี 2016 S. Muruganandham และคณะ [13] ได้เสนอการเรียนรู้เชิงลึกเพื่อเข้ามาจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกลถึง 3 วิธี ได้แก่ เน็ตเวิร์กคอนโวลูชันเชิงลึกแบบฟูลลี (Fully Convolutional Networks หรือ FCN), FCN-8s และ เน็ตเวิร์กแบบดีคอนโวลูชันเชิงลึก (Deep Deconvolutional Neural Networks หรือ DeCNN) โดยทั้ง 3 เน็ตเวิร์กถูกนำมาทำการฝึกสอนและทดสอบเพื่อจำแนกถนนทางความหมายออกจากภาพถ่ายทางอากาศ โดยภาพถ่ายทางอากาศนี้เป็นชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์ (Massachusetts Roads Data Set) [19] เช่นเดียวกับชุดข้อมูลที่ S. Saito, T. Yamashita และ Y. Aoki [3] ใช้ โดยผลลัพธ์หลักจากที่นำไปทดสอบกับชุดข้อมูลถนนของรัฐแมสซาชูเซตส์แล้วประเมินออกมาเป็นค่า F1 อยู่ที่ 0.657 (วิธีการ DeCNN), 0.742 (วิธีการ FCN) และสุดท้าย 0.762 (วิธีการ FCN-8s)

จากงานวิจัยที่เกี่ยวข้องในส่วนของการจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกลโดยใช้การเรียนรู้เชิงลึกแสดงให้เห็นว่าเริ่มมีงานวิจัยที่นำวิธีการเรียนรู้เชิงลึกในรูปแบบมาใช้ในการจำแนกถนน โดยวิธีการที่ดีที่สุดในการจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกลโดยใช้การเรียนรู้เชิงลึก คือวิธีการ FCN-8s ซึ่งให้ค่า F1 อยู่ที่ 0.762 โดยทำการฝึกสอนและทดสอบกับชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์

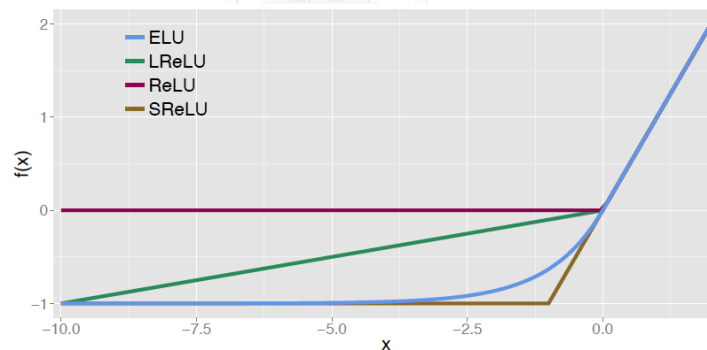
3.3. งานวิจัยเกี่ยวกับเทคนิคใหม่สำหรับการเรียนรู้เชิงลึก (Recent Techniques in Deep Learning)

ปัจจุบันเทคนิคในส่วนของฟังก์ชันกระตุ้น (Activation Function) ที่นิยมใช้ในการเรียนรู้เชิงลึกแบบนิวโรนเน็ตเวิร์กคอนโวลูชันเชิงลึก (Deep Convolutional Neural Network) คือฟังก์ชันเรกติไฟด์เชิงเส้น (Rectified Linear Unit Function หรือ ReLU) แต่ในช่วงปี 2016 ได้มีบทความหนึ่งเสนอวิธีการที่เรียกว่า ฟังก์ชันเอกซ์โพเนนเชียลเชิงเส้น (Exponential Linear Unit Function หรือ ELU) โดย Clevert และคณะ [27] โดยเขานำเสนอฟังก์ชันเอกซ์โพเนนเชียลเชิงเส้นเพื่อนำมาใช้ในการเรียนรู้ (Learning) ให้กับการเน็ตเวิร์กการเรียนรู้เชิงลึก (Deep Network) โดยจะช่วยให้เน็ตเวิร์กเรียนรู้ได้เร็วและให้ประสิทธิภาพสูงกว่าฟังก์ชันกระตุ้นอย่างฟังก์ชันเรกติไฟด์เชิงเส้น โดยงานวิจัยชิ้นนี้ได้รับการตอบรับการตีพิมพ์จาก International Conference on Learning Representations หรือ ICLR ปี 2016 โดยสมการ ELU จะแสดงดังสมการที่ 30 และ 31 และกราฟของ ELU เมื่อเทียบกับฟังก์ชันกระตุ้นที่นิยมใช้กันในปัจจุบันอย่าง ReLU, LReLU (leaky ReLU) และ SReLU (S-shaped ReLU) ไว้รูปที่ 26

$$f(x) = \begin{cases} x; & \text{if } x > 0 \\ \alpha(\exp(x) - 1); & \text{if } x < 0 \end{cases} \quad (32)$$

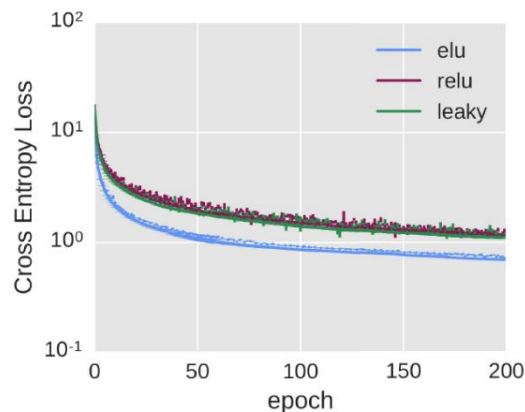
$$f'(x) = \begin{cases} x; & \text{if } x > 0 \\ f(x) + \alpha; & \text{if } x < 0 \end{cases} \quad (33)$$

โดยผลการทดลองที่ได้เมื่อนำไปวัดประสิทธิภาพกับเน็ตเวิร์กต่างๆ เช่น เน็ตเวิร์ก AlexNet หรือเน็ตเวิร์ก CNN ที่ใช้ ฟังก์ชันกระตุ้นชนิด ReLU หรือ Leaky ReLU โดยทำการฝึกสอนและทดสอบกับสองชุดข้อมูลมาตรฐาน ได้แก่ ชุดข้อมูล CIFAR-10 และชุดข้อมูล CIFAR-100 พบว่า เน็ตเวิร์กที่ใช้ ELU ให้ผลค่าความผิดพลาด (Error) อยู่ที่ 6.55 และ 24.28 ขณะที่ AlexNet มีค่าความผิดพลาดอยู่ที่ 18.04 และ 45.80 สุดท้าย เน็ตเวิร์ก CNN มีค่าความผิดพลาดอยู่ที่ 7.25 และ 33.71 เมื่อนำไปฝึกสอนและทดสอบกับชุดข้อมูล CIFAR-10 และชุดข้อมูล CIFAR-100 ตามลำดับ สำหรับงานวิจัยชิ้นนี้ได้แสดงกราฟค่าลอสของครอสเอนโทรปี (Cross Entropy Loss) ระหว่าง ELU กับ ReLU และ Leaky ReLU โดยแสดงในรูปที่ 27 จะเห็นได้ว่าการเรียนรู้โดยเน็ตเวิร์กเชิงลึกแบบ ELU-Network จะให้ Loss ลู่เข้า (Convergence) ได้รวดเร็วและมีประสิทธิภาพสูงกว่า



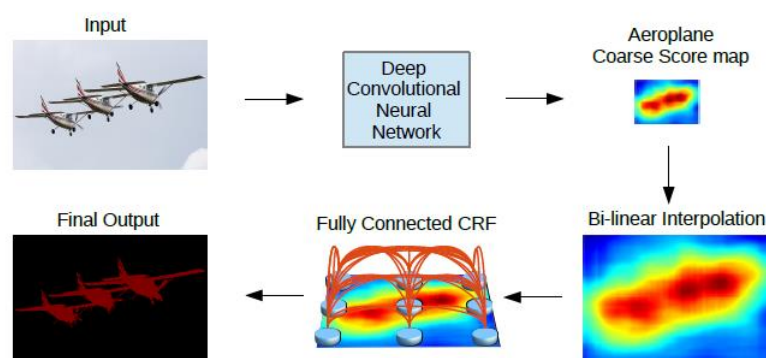
รูปที่ 26 กราฟของ ELU เมื่อเทียบกับฟังก์ชันกระตุ้นที่นิยมใช้กันในปัจจุบันอย่าง ReLU, LReLU (leaky ReLU) และ SReLU (S-shaped ReLU)³¹

³¹ อ้างอิงจาก Fig. 1 ใน [27]



รูปที่ 27 กราฟ Cross Entropy Loss ระหว่าง ELU กับ ReLU และ Leaky ReLU โดยค่าแกน x เป็นจำนวน Epoch และแกน y เป็นค่า Loss ของ ครอสเอนโทรปี (Cross Entropy)³²

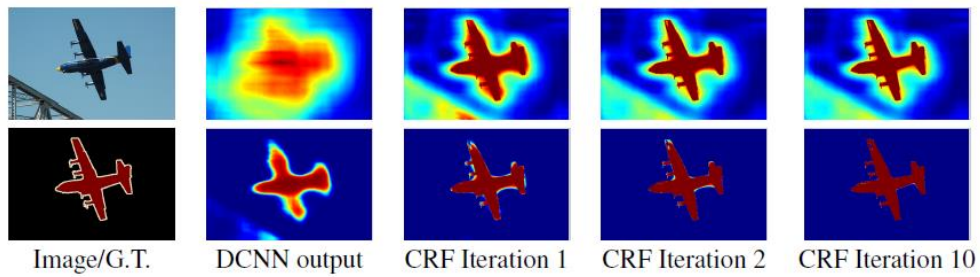
นอกจากนั้น อีกเทคนิคหนึ่งที่ใช้เข้ามาใช้ในส่วนของการประมวลผลหลัง (Post Processing) เพื่อเพิ่มประสิทธิภาพให้กับวิธีการเรียนรู้เชิงลึก คือ เทคนิคแบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส (Conditional Random Fields: CRFs) โดยในงานวิจัยที่ผ่านมา [16-18] พบว่าเทคนิค CRFs ถูกนำมาใช้ในงานของการประมวลผลหลังจากผลลัพธ์ที่ได้ของการเรียนรู้เชิงลึก โดยนำโมเดล CRFs เข้ามาปรับตารางคะแนน (Score Map) ซึ่งเสมือนเป็นข้อมูลนำเข้า (Input) ก่อนเข้าฟังก์ชัน Softmax และตารางความเชื่อ (Belief Map) เสมือนเป็นผลลัพธ์ (Output) จากฟังก์ชัน Softmax โดยแสดงผลลัพธ์ที่ใช้โมเดล CRFs ร่วมกับการเรียนรู้เชิงลึกไว้ดังรูปที่ 28, รูปที่ 29 และรูปที่ 30



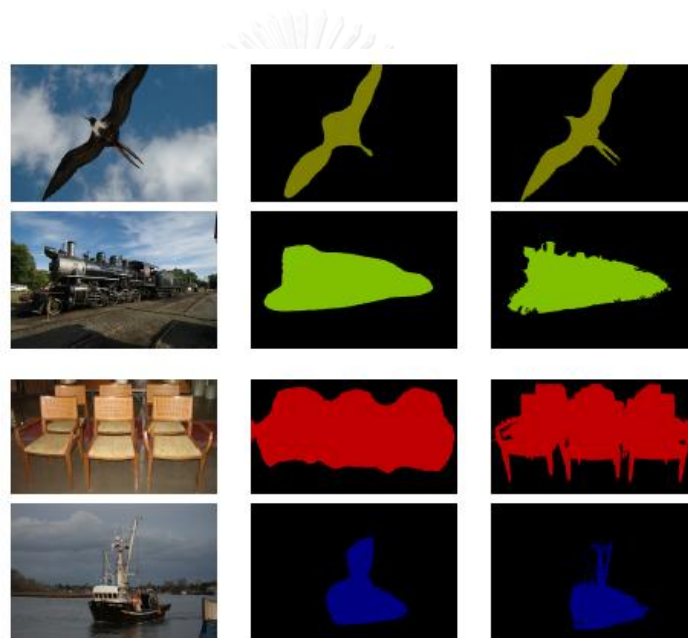
รูปที่ 28 ภาพรวมทั้งหมดของการประยุกต์การเรียนรู้เชิงลึกกับโมเดล CRFs³³

³² อ้างอิงจาก Fig. 4 ใน [27]

³³ อ้างอิงจาก Fig. 3 ใน [18]



รูปที่ 29 การปรับผลลัพธ์ที่ได้ในแต่ละรอบ CRFs (CRF Iteration) ซึ่งผลลัพธ์ที่เมื่อสิ้นสุดรอบที่ 10 จะมีลักษณะใกล้เคียงกับผลเฉลย³⁴



รูปที่ 30 ตัวอย่างผลลัพธ์ที่ได้จากการประยุกต์การเรียนรู้เชิงลึกร่วมกับการใช้ CRFs โดยภาพซ้าย คือ ข้อมูลนำเข้า ภาพกลาง คือ ผลลัพธ์ที่ทำนายจากการเรียนรู้เชิงลึก และภาพขวา คือ ผลลัพธ์ที่ทำนายจากการเรียนรู้เชิงลึกร่วมกับการใช้โมเดล CRFs³⁵

จากนั้นในงานวิจัยที่ผ่านมาได้แสดงผลการประเมินจากการที่ใช้สถาปัตยกรรมการเรียนรู้เชิงลึกร่วมกับโมเดล CRFs โดยทดลองกับชุดข้อมูล Pascal VOC 2012 ไว้ดังรูปที่ 31 แสดงให้เห็นว่า

³⁴ อ้างอิงจาก Fig. 2 ใน [18]

³⁵ อ้างอิงจาก Fig. 7 ใน [18]

การนำมาโมเดล CRFs เข้ามาประยุกต์กับการเรียนรู้เชิงลึกให้ประสิทธิภาพที่ดีกว่าการเรียนรู้เชิงลึกเพียงอย่างเดียว

Method	Mean IOU (%)
MSRA-CFM	61.2
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.3
DeepLab-MSc-CRF-LargeFOV	71.6

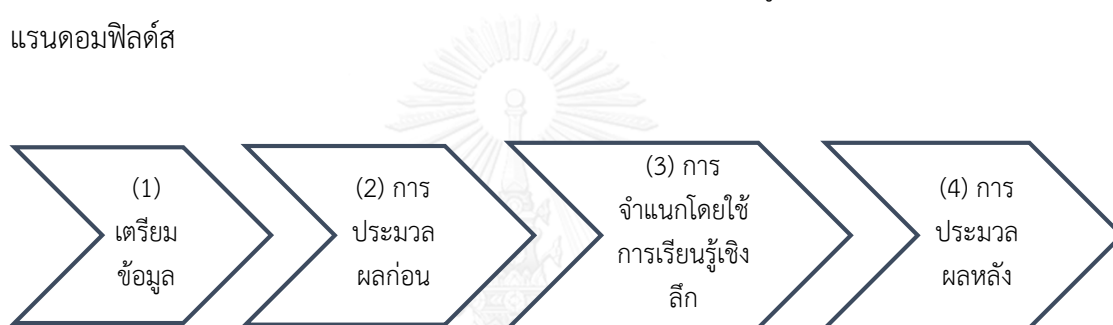
รูปที่ 31 ผลลัพธ์ที่ได้จากการประเมินโดยเทียบกันระหว่างโมเดลที่ใช้ CRFs และโมเดลการเรียนรู้เชิงลึกมาตรฐาน โดยโมเดลการเรียนรู้เชิงลึกที่ใช้ CRFs ให้ประสิทธิภาพที่สูงกว่าโมเดลการเรียนรู้เชิงลึกแบบปกติมากกว่า 7 เปอร์เซ็นต์³⁶

³⁶ อ้างอิงจาก Table. 1 ใน [18]

บทที่ 4

แนวคิดและวิธีการดำเนินงาน

ในบทนี้นำเสนอวิธีการเรียนรู้เชิงลึกเพื่อเข้ามาจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกล โดยวิธีการที่นำเสนอทั้งหมดถูกแบ่งออกเป็น 4 ส่วน ดังรูปที่ 32 ได้แก่ ในส่วนแรก (1) เป็นการอธิบายชุดข้อมูลภาพถ่ายระยะไกลทั้งหมดที่ใช้ในวิทยานิพนธ์นี้ (2) แสดงวิธีการประมวลผลก่อน (Pre Processing) ก่อนนำไปฝึกสอนให้กับการเรียนรู้เชิงลึก (3) แสดงวิธีการที่ใช้ในการจำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกลด้วยการเรียนรู้เชิงลึกที่วิทยานิพนธ์นี้นำเสนอ (4) เป็นการทำให้วิธีการประมวลผลหลัง (Post Processing) โดยการใช้ตัวชี้วัดภูมิทัศน์และวิธีการคอนดิชันนอลแรนดอมฟิลด์ส



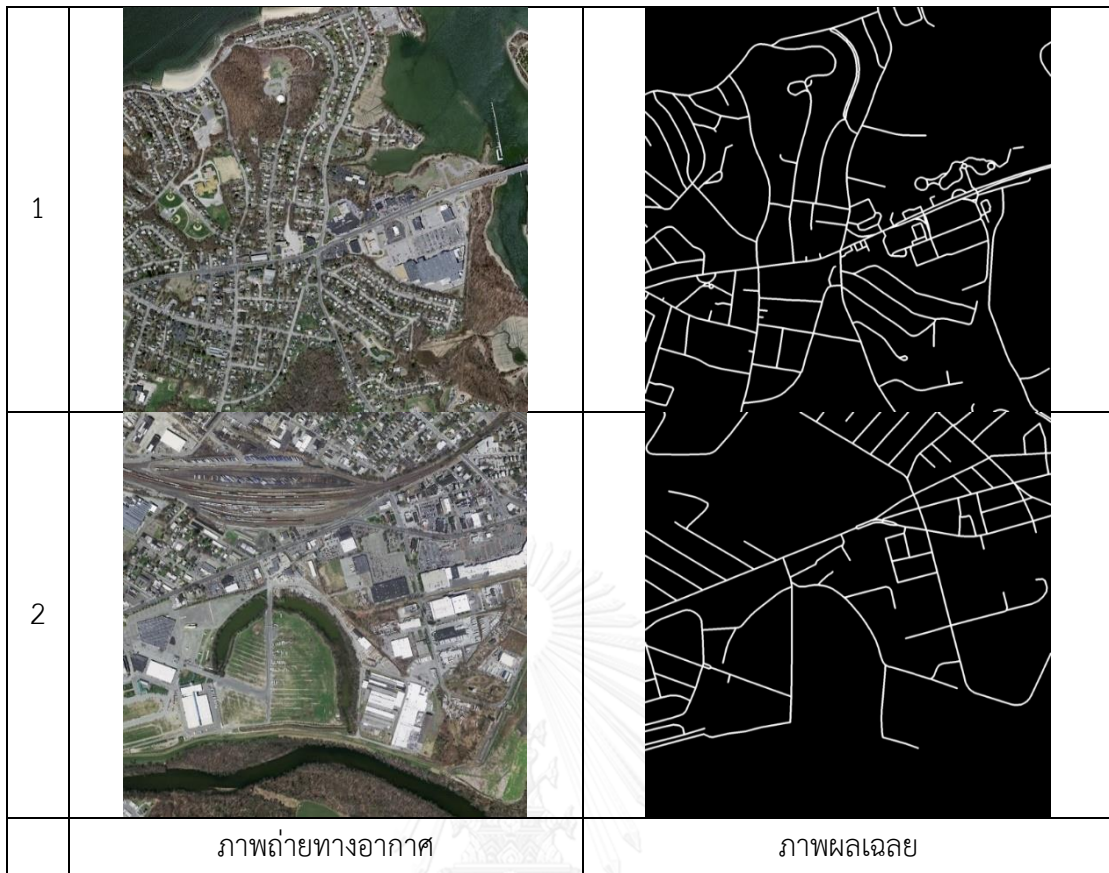
รูปที่ 32 ภาพรวมของขั้นตอนทั้งหมดที่วิทยานิพนธ์นี้นำเสนอ

4.1 การเตรียมข้อมูล

ในส่วนของการเตรียมข้อมูลแบ่งออกเป็น 2 ส่วน โดยเริ่มจากเตรียมชุดข้อมูลถนนของรัฐแมสซาชูเซตส์และเตรียมชุดข้อมูลถนนของประเทศไทย

4.1.1 ชุดข้อมูลถนนของรัฐแมสซาชูเซตส์

ชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์เป็นชุดข้อมูลภาพถ่ายทางอากาศ (Aerial Image) ซึ่งคุณสมบัติของภาพชนิดนี้จะเป็นภาพที่มีความละเอียดสูงมาก (แสดงรูปที่ 33) โดยแบ่งชุดข้อมูลย่อยออกเป็น 3 ชุด คือชุดข้อมูลสำหรับการฝึกสอนจำนวน 1,108 ภาพ ชุดข้อมูลสำหรับตรวจสอบจำนวน 14 ภาพ และ ดข้อมูลสำหรับการทดสอบจำนวน 49 ภาพ โดยแต่ละภาพจะมีขนาดอยู่ที่ 1500x1500 พิกเซล (Pixel) โดยมีค่า Resolution อยู่ที่ 1 เมตร²/พิกเซล [19]



รูปที่ 33 สองตัวอย่างภาพถ่ายทางอากาศและภาพผลเฉลยของชุดข้อมูลรัฐแมสซาชูเซตส์

4.1.2 ชุดข้อมูลถนนของประเทศไทย

ชุดข้อมูลถัดมาเป็นชุดข้อมูลของประเทศไทยโดยเป็นชุดข้อมูลภาพถ่ายดาวเทียม (Satellite Image) ซึ่งคุณสมบัติของภาพจะเป็นภาพที่มีความละเอียดสูง (High Resolution) สำหรับชุดข้อมูลของประเทศไทยเป็นชุดข้อมูลที่ถูกถ่ายจากดาวเทียมไทยโชต (Thaichote) หรือดาวเทียมธีออส (THEOS) โดยผู้ให้บริการ คือสำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน), (GISTDA)

สำหรับภาพถ่ายดาวเทียมในชุดข้อมูลนี้ถูกถ่ายโดยดาวเทียมธีออสซึ่งเป็นระบบดาวเทียมสำรวจทรัพยากรของประเทศไทยโดยถูกออกแบบให้เป็นดาวเทียมขนาดเล็กอีกทั้งมีอายุการใช้งานอย่างน้อย 5 ปี ทำงานโดยอาศัยแหล่งพลังงานจากดวงอาทิตย์ บันทึกภาพได้ครอบคลุมพื้นที่ทั่วโลก ติดตั้งอุปกรณ์ถ่ายภาพแบบออปติคัล (Optical Imagery) ทำให้บันทึกข้อมูลภาพในช่วงคลื่นแสงที่ตามองเห็น (Visible band) จนถึงช่วงคลื่นอินฟราเรดใกล้ (Near Infrared) ทั้งนี้เมื่อเปรียบเทียบช่วงคลื่นของดาวเทียมธีออสกับดาวเทียมอื่นๆ พบว่า 3 ช่วงคลื่นของดาวเทียมธีออส มีความคล้ายคลึงกับ

ช่วงคลื่นของดาวเทียม SPOT ยกเว้นช่วงคลื่นสีน้ำเงินที่มีเพิ่มมากกว่าของดาวเทียม SPOT และมีความคล้ายคลึงกันกับช่วงคลื่นของดาวเทียม Landsat ระบบ TM สำหรับในงานวิทยานิพนธ์นี้เลือกใช้ชุดข้อมูลจากผลิตภัณฑ์แบบ Pan-Sharpned ซึ่งรวมข้อมูลที่ตามองเห็นจาก 4 ช่วงคลื่น (น้ำเงิน เขียว แดง อินฟราเรดไกล) เข้ากับข้อมูลเชิงพื้นที่ของช่วงคลื่นขาว-ดำ โดยจะอธิบายคุณลักษณะดาวเทียมธีออสไว้ตารางที่ 2³⁷

ตารางที่ 2 คุณลักษณะดาวเทียมธีออส

น้ำหนัก	715 กิโลกรัม
ขนาด	2.1 เมตร x 2.1 เมตร (Sun Synchronous)
แผงรับแสงอาทิตย์	840 วัตต์
เชื้อเพลิงและความจุของถังเชื้อเพลิง (Hydrazine)	80 กก.
วงโคจรแบบ	สัมพันธ์กับดวงอาทิตย์ (Sun Synchronous)
ความสูงของวงโคจร	822 กิโลเมตร จากพื้นโลก
ความเอียงของแนวการโคจร	98.7 องศา
จำนวนวงโคจรต่อวัน	14+5/26 วงโคจรต่อวัน
เวลาท้องถิ่นที่โคจรผ่าน	10:00 น.
ระยะเวลาโคจรรอบโลก 1 รอบ	101.4 นาที
การโคจรกลับมาแนวเดิม	ทุก 26 วัน (369 วงโคจร)
ความเร็วเมื่อเทียบกับพื้นโลก	6.6 กม./วินาที
ความจุของอุปกรณ์เก็บข้อมูล	40 Gbit solid-state memory
การประมวลผลข้อมูลบนดาวเทียม	อัตราส่วนของการบีบอัดข้อมูล 2.80 or 3.75 for PAN 2.95 or 3.75 for MS
อัตราการส่งข้อมูลภาพ (Image Telemetry)	120 Mbit/s (X band)

³⁷ อ้างอิงจาก <http://www.gistda.or.th/main/th/node/90>

ช่องว่างระหว่างแนวโคจร (ที่เส้นศูนย์สูตร)	108 กิโลเมตร ระหว่างแนวการโคจร 2 แนวที่ไกลที่สุดที่ ดาวเทียมผ่าน 2800 กิโลเมตร ระหว่างแนวโคจร 2 วงที่ต่อเนื่องกัน
ขอบเขตการบันทึกข้อมูล	ทั่วทั้งโลก
ขอบเขตการรับสัญญาณ	รัศมีมากกว่า 2000 กิโลเมตร จากสถานีรับภาคพื้นดิน (ที่มุมเงย 5 องศา)
เวลาในการโคจรมาถึงเป้าหมาย	2 วัน เมื่อดาวเทียมเอียงถึง 50 องศา 5 วัน เมื่อดาวเทียมเอียงถึง 30 องศา
อุปกรณ์บันทึกข้อมูล	Panchromatic (ช่วงคลื่นเดียว) รายละเอียดภาพ 2 เมตร, ความกว้างแนวภาพ 22 กิโลเมตร Multispectral (หลายช่วงคลื่น) รายละเอียดภาพ 15 เมตร, ความกว้างแนวภาพ 90 กิโลเมตร
อายุการใช้งาน	อย่างน้อย 5 ปี

ในส่วนของผลเฉลยวิทยานิพนธ์นี้ใช้ชุดข้อมูล Shape Files จากข้อมูลภูมิสารสนเทศพื้นฐานของประเทศ (Fundamental Geographic Data Set : FGDS) ซึ่งเป็นชั้นข้อมูลที่มีศักยภาพสูงที่นำมาใช้งานร่วมกันระหว่างหน่วยงานต่างๆ ได้ และใช้เป็นข้อมูลพื้นฐานในการอ้างอิงเพื่อเพิ่มเติมชั้นข้อมูลในด้านอื่นๆ ได้เช่นเดียวกัน ข้อมูลประเภทนี้เกี่ยวข้องกับลักษณะของภูมิประเทศทั่วไป

คณะกรรมการภูมิสารสนเทศแห่งชาติได้มีกำหนดชั้นข้อมูลภูมิสารสนเทศพื้นฐานของประเทศไทยไว้ทั้งหมด 13 ชั้นข้อมูล โดยในวิทยานิพนธ์นี้ใช้ข้อมูลผลเฉลยจากชั้นข้อมูลเส้นทางคมนาคมของกระทรวงคมนาคมร่วมกับภาพถ่ายจากดาวเทียม-สตอก.³⁸

ในชุดข้อมูลประกอบด้วย 5 สถานที่ คือ พุทธมณฑล นครปฐม, ศรีราชา ชลบุรี, สงขลา, สุรินทร์ และอุบลราชธานี โดยขนาดภาพถ่ายดาวเทียมมีขนาด 1500x1500 พิกเซล

(1) **พุทธมณฑล (นครปฐม)** มีทั้งหมดจำนวน 263 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอนจำนวน 200 ภาพ ชุดข้อมูลสำหรับการตรวจสอบจำนวน 14 ภาพ และชุดข้อมูลสำหรับการทดสอบจำนวน 49 ภาพ (แสดงรูปที่ 34, 35)

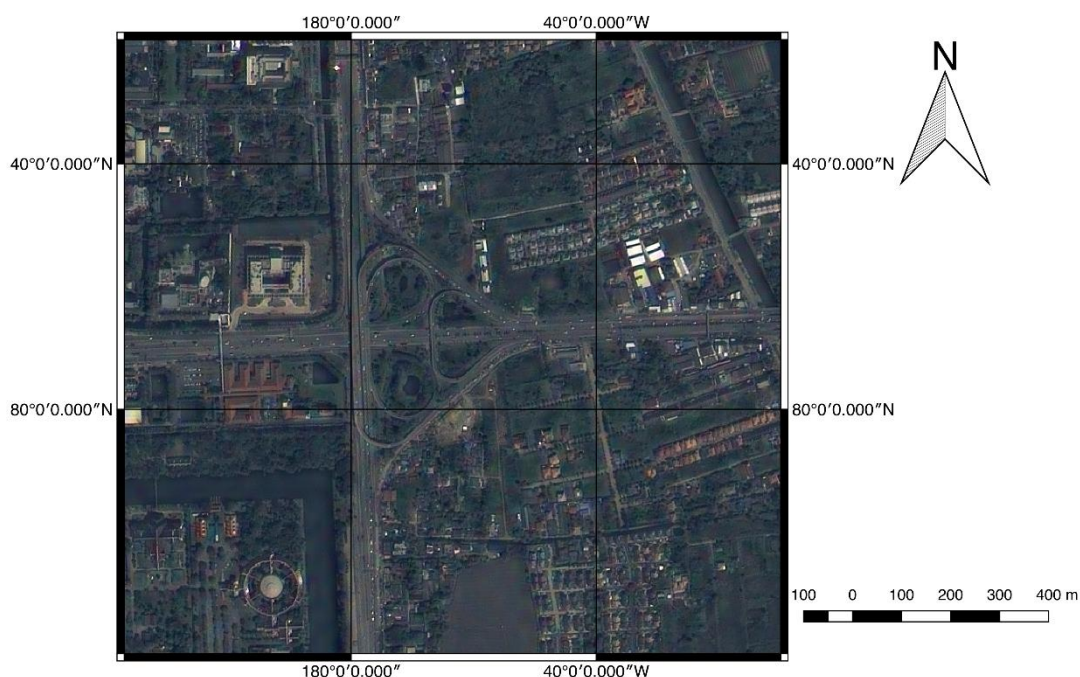
³⁸ อ้างอิงจาก <http://thaisdi.gistda.or.th/en/index.php>

(2) **ศรีราชา (ชลบุรี)** มีทั้งหมดจำนวน 100 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอนจำนวน 87 ภาพ ชุดข้อมูลสำหรับการตรวจสอบจำนวน 3 ภาพ และชุดข้อมูลสำหรับการทดสอบจำนวน 10 ภาพ (แสดงรูปที่ 36, 37)

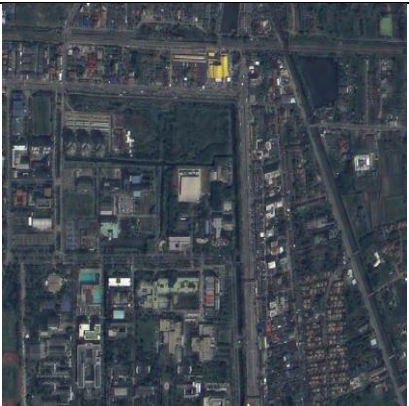

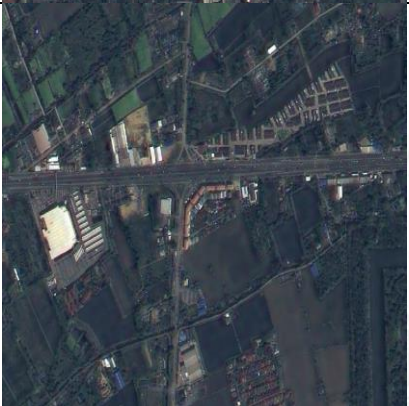
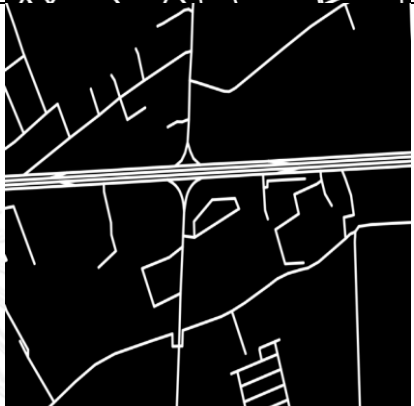
(3) **สงขลา** มีทั้งหมดจำนวน 100 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอนจำนวน 87 ภาพ ชุดข้อมูลสำหรับการตรวจสอบจำนวน 3 ภาพ และชุดข้อมูลสำหรับการทดสอบจำนวน 10 ภาพ (แสดงรูปที่ 38, 40)

(4) **สุรินทร์** มีทั้งหมดจำนวน 70 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอนจำนวน 57 ภาพ ชุดข้อมูลสำหรับการตรวจสอบจำนวน 3 ภาพ และชุดข้อมูลสำหรับการทดสอบจำนวน 10 ภาพ (แสดงรูปที่ 39, 42)

(5) **อุบลราชธานี** มีทั้งหมดจำนวน 70 ภาพ แบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอนจำนวน 57 ภาพ ชุดข้อมูลสำหรับการตรวจสอบจำนวน 3 ภาพ และชุดข้อมูลสำหรับการทดสอบจำนวน 10 ภาพ (แสดงรูปที่ 41, 43)



รูปที่ 34 ตำแหน่งที่ตั้งของเขตพื้นที่พุทธมณฑล (นครปฐม) ในชุดข้อมูลภาพถนนประเทศไทย

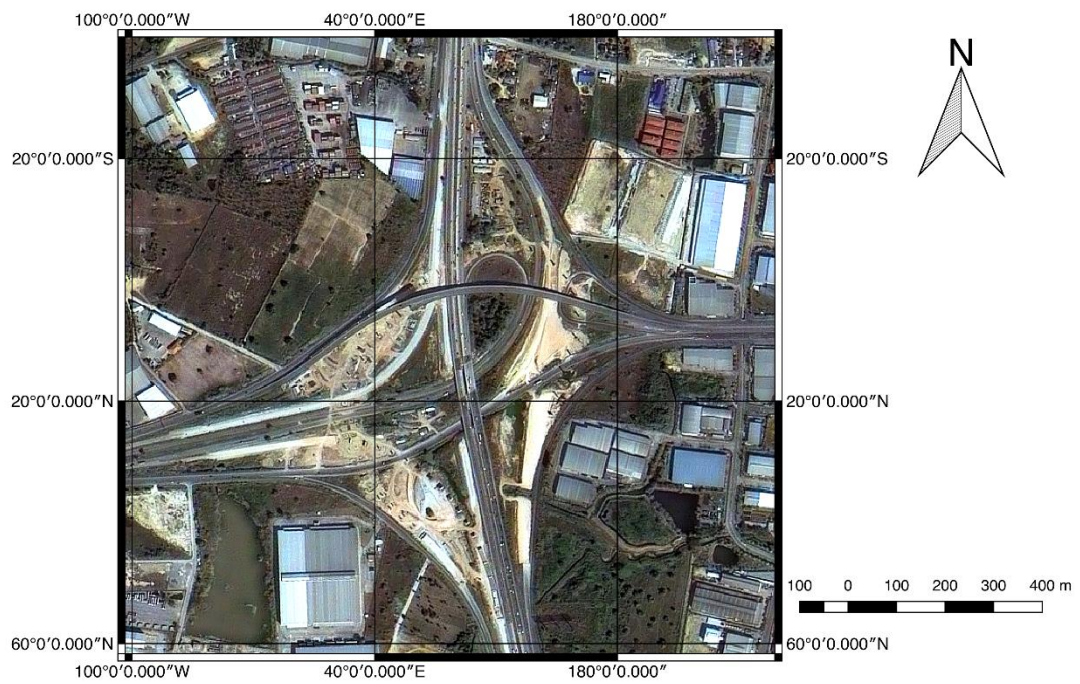
1		
2		
	<p>ภาพถ่ายดาวเทียม</p>	<p>ภาพผลเฉลย</p>

รูปที่ 35 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่พุทธมณฑล (นครปฐม) ในชุดข้อมูลภาพถนนประเทศไทย





1		
---	---	--



รูปที่ 36 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่ศรีราชา (ชลบุรี) ในชุดข้อมูลภาพถนนประเทศไทย

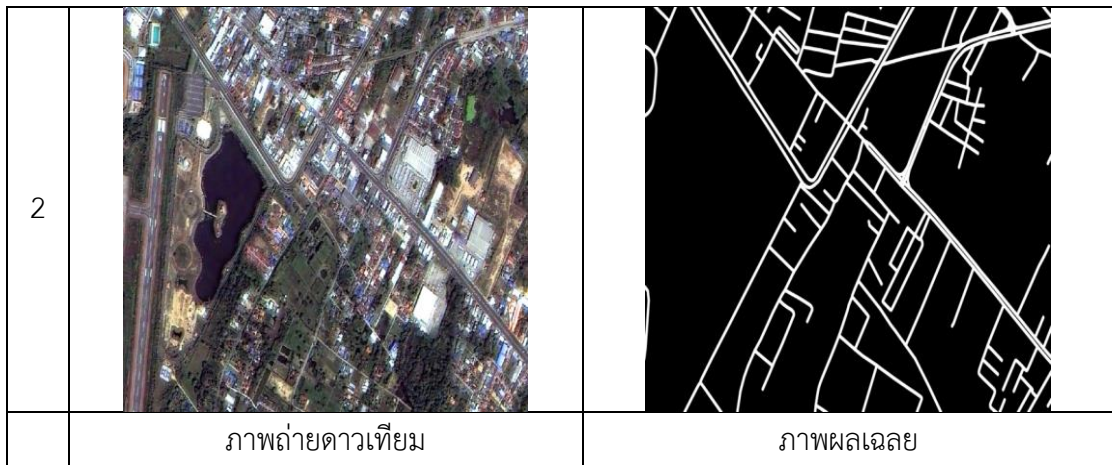


รูปที่ 37 ตำแหน่งที่ตั้งของเขตพื้นที่ศรีราชา (ชลบุรี) ในชุดข้อมูลภาพถนนประเทศไทย

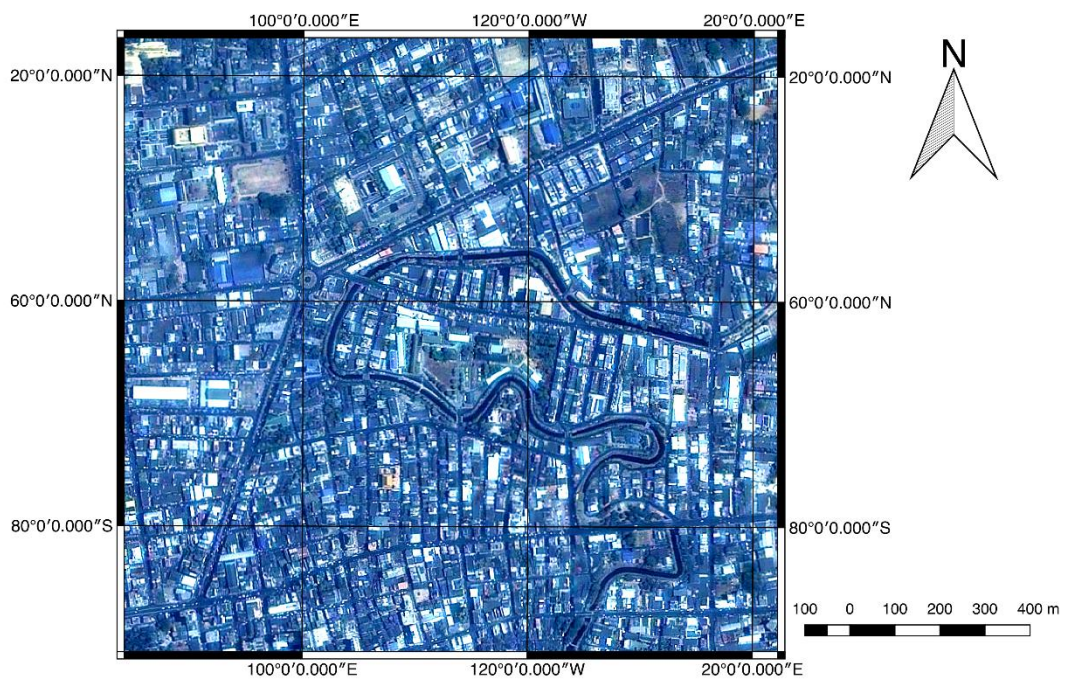
1		
2		
	ภาพถ่ายดาวเทียม	ภาพผลเฉลย

รูปที่ 38 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่สงขลาในชุดข้อมูลภาพถนนประเทศไทย




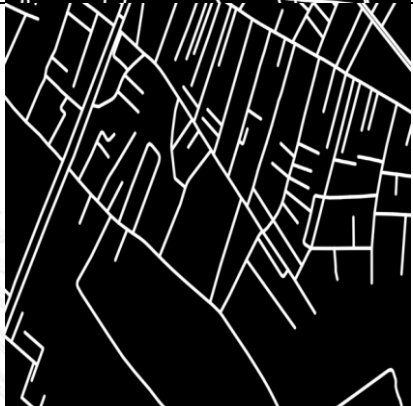


1		
---	---	--



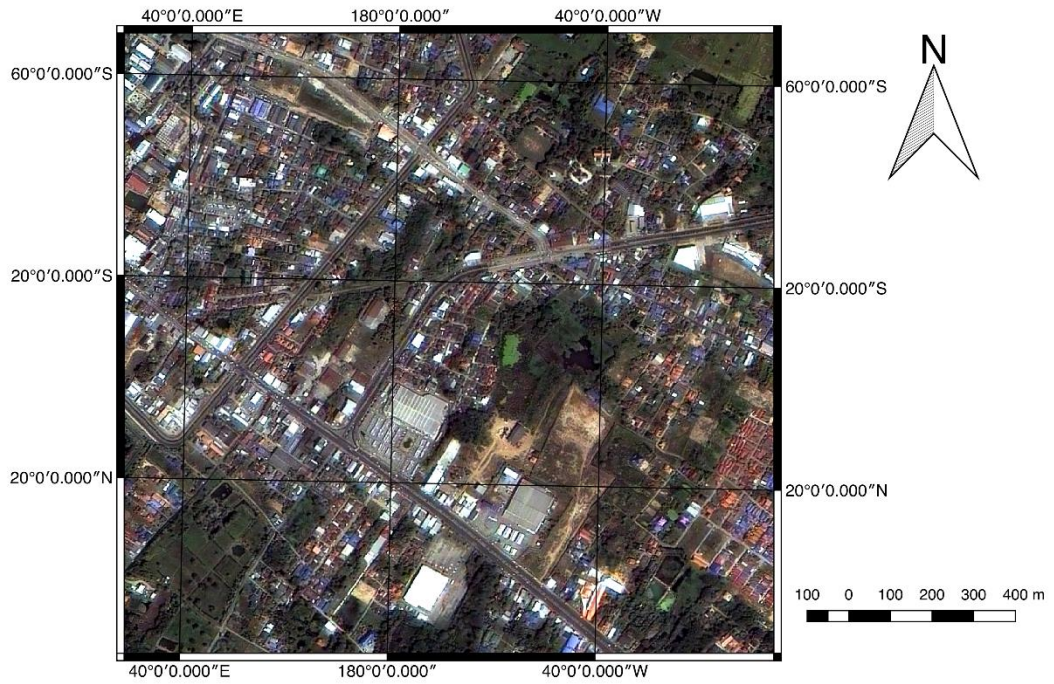
รูปที่ 39 สองตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่สุรินทร์ในชุดข้อมูลภาพถนนประเทศไทย



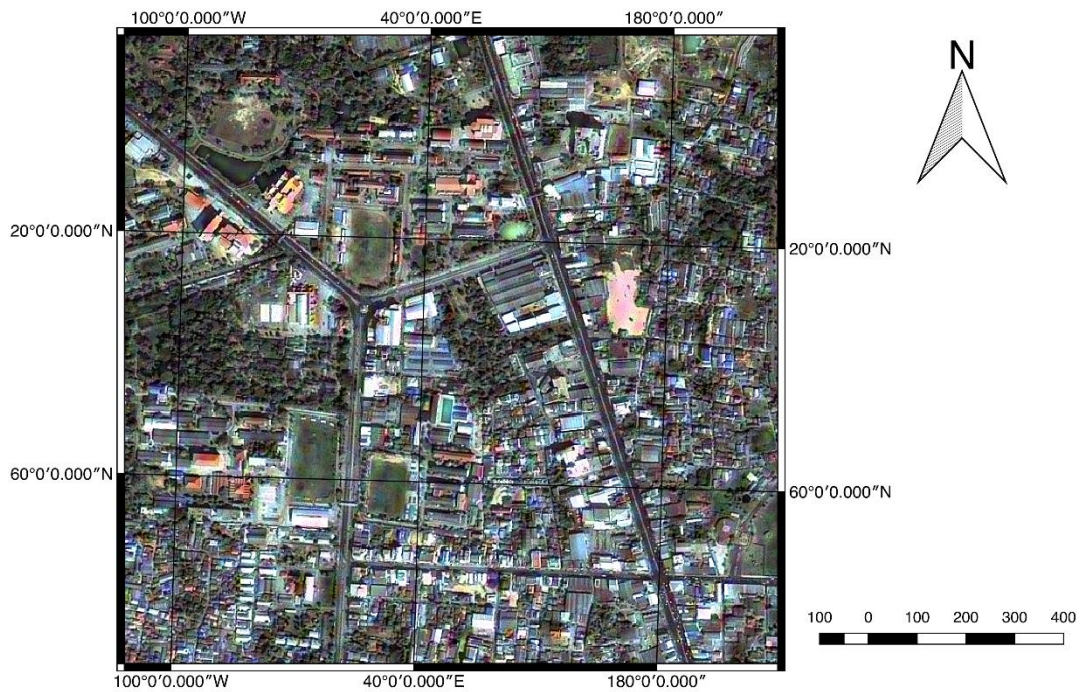
รูปที่ 40 ตำแหน่งที่ตั้งของเขตพื้นที่สงขลาในชุดข้อมูลภาพถนนประเทศไทย

1		
2		
3		
	ภาพถ่ายดาวเทียม	ภาพผลเฉลย

รูปที่ 41 สามตัวอย่างของภาพถ่ายดาวเทียมที่เขตพื้นที่อุบลราชธานีในชุดข้อมูลภาพถนนประเทศไทย



รูปที่ 42 ตำแหน่งที่ตั้งของเขตพื้นที่สุรินทร์ในชุดข้อมูลภาพถนนประเทศไทย



รูปที่ 43 ตำแหน่งที่ตั้งของเขตพื้นที่อุบลราชธานีในชุดข้อมูลภาพถนนประเทศไทย

4.2 การประมวลผลก่อน (Pre Processing)

ในส่วนของการประมวลผลก่อนแบ่งออกเป็น 3 ส่วน โดยเริ่มจากการทำนอร์มัลไลเซชัน ถัดมาเป็นการเพิ่มปริมาณข้อมูลให้กับภาพถ่ายระยะไกล และสุดท้ายคือการแปลงชุดข้อมูลภาพถ่ายระยะไกลให้เก็บอยู่ในรูปเวกเตอร์นัมพาย

4.2.1 การทำนอร์มัลไลเซชัน (Normalization)









สำหรับภาพถ่ายระยะไกลที่ใช้ในวิทยานิพนธ์นี้มีช่องสัญญาณ 3 ช่องสัญญาณ คือ ช่องสัญญาณสีแดง, ช่องสัญญาณสีเขียว และช่องสัญญาณสีน้ำเงิน สำหรับการทำให้ให้เป็นค่านอร์มัลไลเซชันกับภาพถ่ายระยะไกล งานวิจัยนี้นำแต่ละช่องสัญญาณมาทำการปรับค่าแต่ละช่องสัญญาณจากช่วง $[0, 255]$ ให้อยู่ช่วง $[0, 1]$ โดยการนำแต่ละช่องสัญญาณมาลบด้วยค่า 128 และหารด้วย 128 ตามลำดับ ส่วนในกรณีของภาพเฉดสีมี 1 ช่องสัญญาณ คือ ช่องสัญญาณสีเทาซึ่งอยู่ในช่วง $[0, 255]$ งานวิจัยนี้นำช่องสัญญาณสีเทามาหารด้วย 255 เพื่อให้ช่องสัญญาณมีเพียง 2 ค่า คือ $[23]$ โดย 0 คือพิกเซลที่ไม่ใช่ถนน และ 1 คือพิกเซลที่เป็นถนน ตามลำดับ

4.2.2 การเพิ่มปริมาณข้อมูลให้กับภาพถ่ายระยะไกล

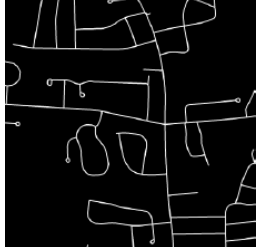

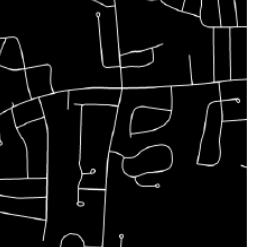


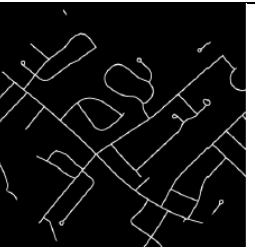

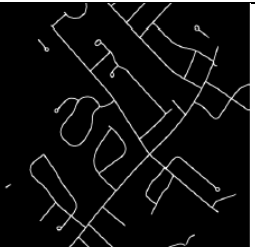
สำหรับการเพิ่มปริมาณให้กับชุดข้อมูลการฝึกสอน (Training data set) วิทยานิพนธ์นี้เลือกวิธีการหมุนภาพให้กับแต่ละภาพถ่ายระยะไกลโดยหมุนภาพถ่ายไปครั้งละ 45 องศา ดังนั้น 1 ภาพถ่ายระยะไกลจะให้ภาพถ่ายที่มีปริมาณเพิ่มขึ้นอีก 7 ภาพ โดยแสดงการหมุนภาพถ่ายระยะไกลและภาพผลเฉลยไว้ดังรูปที่ 44 และรูปที่ 45

4.2.3. การแปลงชุดข้อมูลภาพถ่ายระยะไกลให้เก็บอยู่ในรูปเวกเตอร์นัมพาย

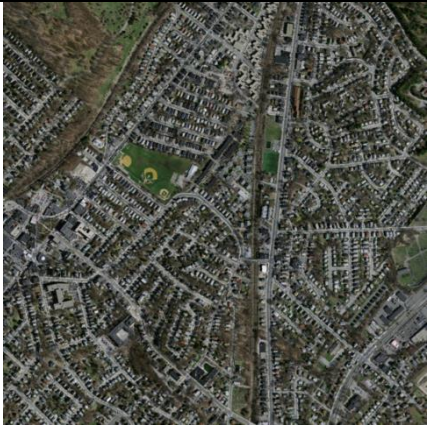

สำหรับการแปลงชุดข้อมูลภาพถ่ายระยะไกลให้อยู่ในรูปแบบเวกเตอร์นัมพาย (นัมพายเป็นไลบรารีของภาษาไพธอนสำหรับใช้คำนวณทางคณิตศาสตร์) มีวัตถุประสงค์เพื่อที่จะลดเวลาในการเปิดไฟล์เวลานำไปเข้าเน็ตเวิร์กสำหรับฝึกสอน เพื่อลดขั้นตอนแบบเดิมซึ่งเป็นการเปิดภาพทีละภาพสำหรับเวลาส่งไปฝึกสอนที่เน็ตเวิร์ก (แสดงรูปที่ 46)

			
ภาพถ่ายระยะไกล	หมุนภาพ 45 องศา	หมุนภาพ 90 องศา	หมุนภาพ 135 องศา
			
หมุนภาพ 180 องศา	หมุนภาพ 225 องศา	หมุนภาพ 270 องศา	หมุนภาพ 315 องศา

รูปที่ 44 การนำภาพถ่ายระยะไกล 1 ภาพมาเพิ่มปริมาณข้อมูลโดยการหมุนภาพไปที่ละ 45 องศา (1 ภาพเพิ่มปริมาณขึ้นอีก 7 ภาพ)

			
ภาพผลเฉลี่ย	หมุนภาพ 45 องศา	หมุนภาพ 90 องศา	หมุนภาพ 135 องศา
			
หมุนภาพ 180 องศา	หมุนภาพ 225 องศา	หมุนภาพ 270 องศา	หมุนภาพ 315 องศา

รูปที่ 45 การนำภาพผลเฉลี่ย 1 ภาพมาเพิ่มปริมาณข้อมูลโดยการหมุนภาพไปที่ละ 45 องศา (1 ภาพเพิ่มปริมาณขึ้นอีก 7 ภาพ)

1		<pre>652c 2027 7368 6170 6527 3a20 2832 3230 2c20 332c 2035 3030 2c20 3530 3029 2c20 7d20 2020 2020 2020 2020 2020 2020 200a 7b67 7667 7b82 8ca6 2762 1e00 0811 030c 1b22 3336 5059 404e 6066 4d56 5565 7fb0 a460 4250 536e 442b 2e1d 2719 2426 0423 122e 361f 1013 1208 0555 756e 746e 727b 7480 3f34 422a 2e31 1a2b 1011 1659 7c50 7476 6d75 6a77 7782 7881 6366 7a8e 1111 0f04 0c06 0b67 6160 6784 867e 8689 858e b893 613b 1112 8c3a 2b3d 3432 4f9a 9690 6d5a 3819 090a 0905 2306 3b39 4816 4b49 4458 576b 4550 5b33 1319 010d 0d06 644b</pre>
2		<pre>0000 0000 0000 0000 0000 0000 0000 0000 0001 0100 0100 0000 0101 0000 0101 0000 0001 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000</pre>
	ภาพถ่ายระยะไกล	ภาพถ่ายถูกเก็บอยู่รูปแบบของเวกเตอร์นัมพาย

รูปที่ 46 การแปลงภาพถ่ายระยะไกล (1) และภาพผลเฉลย (2) ให้อยู่ในรูปแบบของเวกเตอร์นัมพาย

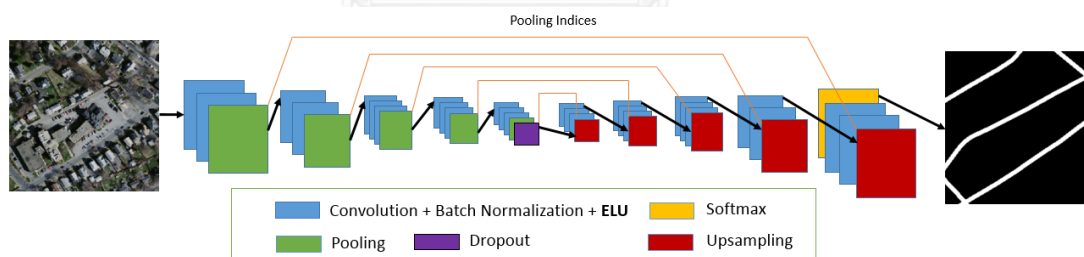
CHULALONGKORN UNIVERSITY

4.3 การจำแนกถนนทางความหมายด้วยการเรียนรู้เชิงลึก

จากงานวิจัยที่ผ่านมา ผู้วิจัยได้เห็นว่า มีเน็ตเวิร์ก (Network) ที่นำมาทำการจำแนกภาพทางความหมายโดยใช้การเรียนรู้เชิงลึกหลายเน็ตเวิร์ก และหนึ่งในเน็ตเวิร์กที่ดีที่สุดโดยประเมินกับชุดข้อมูล PASCAL VOC [24] คือเน็ตเวิร์กชนิดคอนโวลูชันเชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ (Deep Convolutional Encoder-Decoder Networks หรือ DCED) ซึ่งนำมาตอบแบบตอบหลายคำตอบ (Multi Class) สำหรับการจำแนกทางความหมาย และในปี 2016 ที่ผ่านมามีงานวิจัยที่เสนอโดย Kalinovsky, A. Kalinovsky, และ V. Kovalev [26] นำ DCED มาใช้จำแนกทางความหมายกับชุดข้อมูลภาพเอ็กซเรย์หน้าอก (X-Ray Chest Images) ซึ่งตอบผลลัพธ์สุดท้ายแบบสองคำตอบ (Binary Class) คือส่วนที่เป็นปอด (พิกเซลเป็นสีขาว) กับส่วนที่ไม่ใช่ปอด (พิกเซลเป็นสีดำ) ดังนั้น วิทยานิพนธ์จึงเสนอการนำ DCED มาใช้จำแนกถนนทางความหมายออกจากภาพถ่ายระยะไกล เพื่อ

นำมาใช้ตอบแบบตบสองคำตอบ (Binary Class) คือส่วนที่เป็นถนน และส่วนที่ไม่ใช่ถนน สำหรับการปรับปรุงเน็ตเวิร์ก **วิทยานิพนธ์นี้จึงเสนอวิธีการใหม่** ดังต่อไปนี้

1. เพิ่มประสิทธิภาพการจำแนกให้กับเน็ตเวิร์กโดยการเปลี่ยนฟังก์ชันการกระตุ้น (Activation Function) ใหม่ จากเดิม ReLU เป็น ELU โดยเรียกวิธีการที่นำเสนอใหม่นี้ว่า สถาปัตยกรรม **ELU-DCED** (แสดงรูปที่ 47)
2. เพิ่มจำนวนข้อมูลฝึกสอนโดยกำหนดให้แต่ละภาพถ่ายระยะไกลหมุนภาพไปที่ละ 45 องศา (1 ภาพถ่ายระยะไกลเมื่อหมุนครบแล้ว จะได้เพิ่มมาอีก 7 ภาพถ่าย) จากนั้นจึงส่งข้อมูลภาพถ่ายระยะไกลไปทำการเรียนรู้ (Training) ให้กับสถาปัตยกรรม ELU-DCED
3. หลังจากได้แยกถนนจากความหมายออกจากภาพถ่ายระยะไกลโดยทำการทดสอบกับสถาปัตยกรรม ELU-DCED เสร็จสิ้น จากนั้นเป็นขั้นตอนของการทำการประมวลผลหลัง (Post Processing) โดย วิทยานิพนธ์นี้การใช้ตัวชี้วัดภูมิทัศน์ (Landscape Metrics: LMs) [15] โดยจะอธิบายไว้ที่หัวข้อ 4.4 ซึ่งเหตุผลที่วิทยานิพนธ์เสนอตัวชี้วัดภูมิทัศน์เข้ามาทำวิธีการประมวลผลหลัง (Post Processing) เพื่อจะนำมาแก้ปัญหาผลบวกวง (False Positive) โดยเรียกวิธีการใหม่นี้ว่า สถาปัตยกรรม **ELU-DCED-LMs**
4. ขั้นตอนสุดท้ายเป็นการใช้เทคนิค CRFs เข้ามาเพิ่มความคมและความถูกต้องให้กับผลลัพธ์ที่ได้จากขั้นตอนที่ 3 โดยเรียกวิธีการใหม่นี้ว่าสถาปัตยกรรม **ELU-DCED-LMs-CRFs**



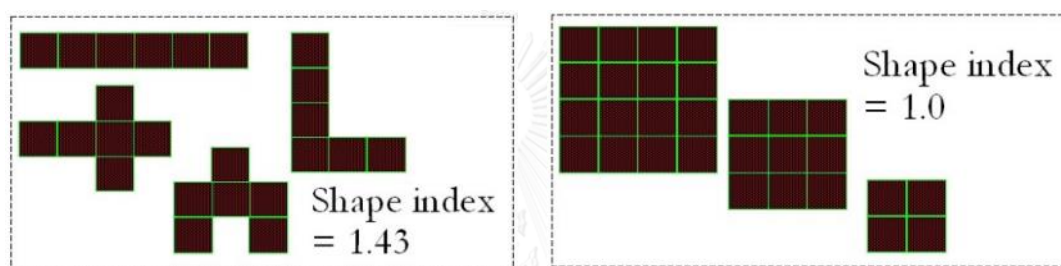
รูปที่ 47 ภาพสถาปัตยกรรมการเรียนรู้เชิงลึกที่งานวิทยานิพนธ์นี้นำเสนอ (ELU-DCED)

4.4 การประมวลผลหลัง (Post Processing)

วิธีการนี้เป็นการทำหลังจากกระบวนการทำงานของการเรียนรู้เชิงลึกเสร็จสิ้น โดยเลือกตัวชี้วัดภูมิทัศน์ [15] ประเภทตัวชี้วัดรูปร่าง (Shape Metrics) เพื่อนำมาคำนวณหาดัชนีความซับซ้อน (Shape Index) โดยจะพิจารณาวัตถุที่ได้ทั้งหมดที่ตอบออกมาทั้งหมด แล้วเลือกวัตถุที่อยู่ติดกันทั้งหมดโดยใช้เทคนิคการกำหนดให้วัตถุที่อยู่ติดกันเชื่อมกัน (Connected-Component Labeling)

[28] ร่วมกันกับการเพิ่มประสิทธิภาพการจำแนกให้กับวัตถุที่เชื่อมกันแล้วโดยการเพิ่มความราบเรียบโดยใช้ตัวกรองเกาส์เซียน (Smoothing with Gaussian Filter) จากนั้นจึงนำไปพิจารณาที่ละวัตถุที่อยู่ติดกันโดยการดูดัชนีความซับซ้อน (Shape Index) หากดัชนีความซับซ้อนต่ำวัตถุนั้นจะถือว่าเป็นเงื่อนไข สำหรับตัวอย่างวัตถุที่มีดัชนีความซับซ้อนในแบบต่างๆ จะแสดงไว้รูปที่ 48 โดยสูตรการหาดัชนีความซับซ้อน (Shape Index) จะแสดงไว้ที่สมการ 32

$$\text{Shape Index} = \frac{\text{perimeter}}{4x\sqrt{\text{area}}} \quad (32)$$

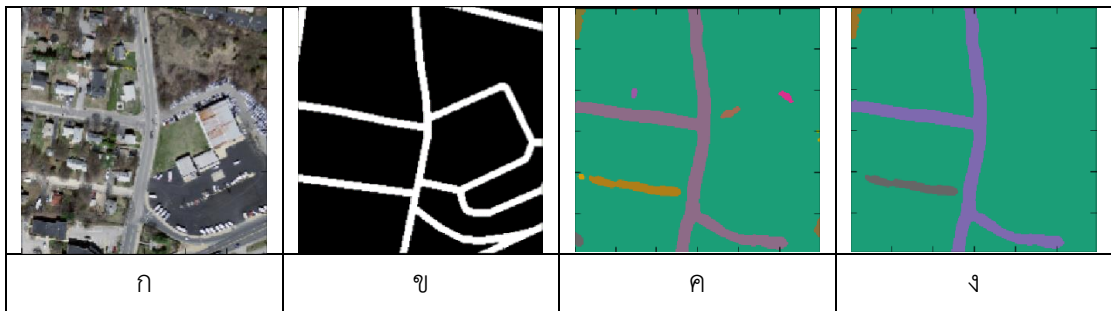


รูปที่ 48 ตัวอย่างวัตถุที่มีดัชนีความซับซ้อนรูปแบบต่างๆ

วัตถุที่มีดัชนีความซับซ้อนสูง (รูปซ้าย) และวัตถุที่มีดัชนีความซับซ้อนต่ำ (รูปขวา)³⁹

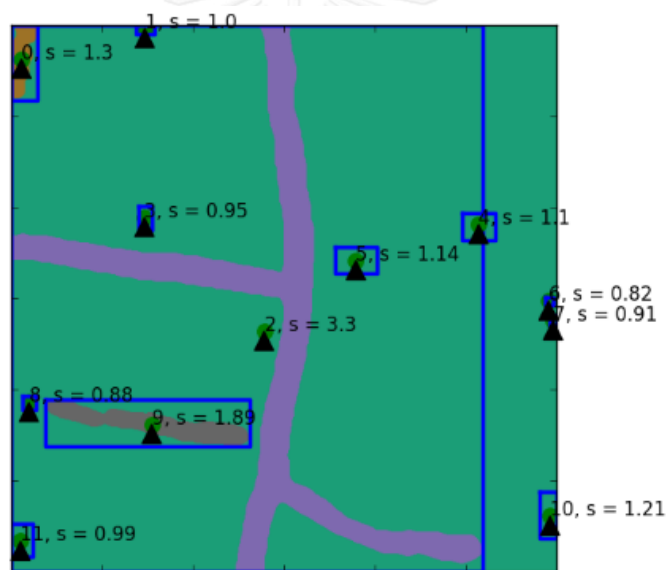
จากรูปที่ 48 ทางด้านซ้าย จะพบว่า เส้นรอบวงมีค่า 14 หน่วย พื้นที่มีค่า 6 หน่วย เมื่อนำไปเข้าสู่สูตรเพื่อหาดัชนีความซับซ้อนสูงจะพบว่ามีค่า 1.43 และในทำนองเดียวกันกับภาพด้านขวาพบว่า เส้นรอบวงมีค่า 16 หน่วย พื้นที่มีค่า 16 หน่วย เมื่อนำไปเข้าสู่สูตรเพื่อหาดัชนีความซับซ้อนสูงจะพบว่ามีค่า 1 ซึ่งแนวคิดนี้ถูกนำมาใช้เพื่อหาวัตถุถนน เนื่องจากวัตถุถนนจริงๆ จะมีค่าความซับซ้อนที่สูงคล้ายกับตัวอย่างที่ได้แสดงดังรูปที่ 48 ด้านซ้าย โดยวิทยานิพนธ์นี้ได้ใช้เทคนิคของ Connected-Component Labeling และ Gaussian Filter เข้ามาช่วยในการจับวัตถุถนนในแต่ละก้อนก่อนนำไปพิจารณาหาค่าความซับซ้อนจากเทคนิคของตัวชี้วัดภูมิทัศน์โดยจะแสดงผลลัพธ์ที่ใช้ทั้งสองเทคนิคข้างต้นไว้รูปที่ 49 และแสดงค่าความซับซ้อนของวัตถุแต่ละก้อนไว้รูปที่ 50

³⁹ อ้างอิงจาก Fig. 9.16 ใน [15]



รูปที่ 49 ผลลัพธ์จากการใช้เทคนิค Connected-Component Labeling และ Gaussian Filter เพื่อเข้ามาช่วยจับวัตถุถนนแต่ละก้อนก่อนนำไปคำนวณหาค่าความซับซ้อนจากเทคนิคตัวชี้วัดภูมิทัศน์ โดยแสดงดังรูป ค และผลลัพธ์ที่ได้หลังจากผ่านเทคนิคตัวชี้วัดภูมิทัศน์จะแสดงไว้ดังรูป ง โดยที่รูป ก

คือ ภาพถ่ายต้นฉบับ และรูป ข คือ ภาพผลเฉลย ตามลำดับ



รูปที่ 50 ผลลัพธ์การคำนวณดัชนีความซับซ้อนแต่ละวัตถุที่ได้จากหลังจากการเรียนรู้เชิงลึก เพื่อนำมาพิจารณาหาวัตถุที่มีค่าความซับซ้อนน้อยจะถือว่าวัตถุนั้นไม่เป็นวัตถุนน

หลังจากเสร็จสิ้นวิธีการตัวชี้วัดภูมิทัศน์ ขั้นตอนสุดท้ายจะเป็นการใช้เทคนิค CRFs เข้ามาเพิ่มความคมและความถูกต้องให้กับผลลัพธ์ที่ได้ โดยวิทยานพจน์นี้ได้กำหนดรอบของ CRFs (CRFs Iteration) เท่ากับ 5 เพื่อให้เอาต์พุตที่ได้มีค่าความใกล้เคียงผลเฉลยมากยิ่งขึ้น

4.5 การสอน (Training)

สำหรับการสอนให้กับเน็ตเวิร์กเป็นขั้นตอนหลักที่สำคัญที่สุดในการเรียนรู้ของเครื่อง ประสิทธิภาพการจำแนกของวิธีที่นำเสนอจะขึ้นอยู่กับขั้นตอนกระบวนการสอนนี้ วิทยานิพนธ์นี้ได้ทำการสุ่มภาพที่อยู่ในชุดข้อมูลภาพถ่ายระยะไกลโดยสุ่มตัด (Random Crop) เพื่อส่งภาพเข้าไปในชั้นนำเข้า (Input Layer) ที่ละ 224×224 พิกเซล โดยกำหนดขนาดรุ่นของการผลิต (Batch Size) เท่ากับ 24 และจำนวนรุ่นของการผลิตในแต่ละรอบอีพอค (Number batches per epoch) อยู่ที่ 100 โดยกำหนดให้มีจำนวนรอบการสอนเท่ากับ 30 รอบอีพอค (Epochs) โดยแต่ละ 1 รอบอีพอคใช้เวลาประมาณ 45 นาที สำหรับโครงสร้างเน็ตเวิร์กที่วิทยานิพนธ์นี้นำเสนอใหม่จะประกอบไปด้วยฝั่ง Encoder จำนวน 13 ชั้น Decoder จำนวน 13 ชั้น โดยฟังก์ชันการกระตุ้น (Activation Function) ที่ใช้ในเน็ตเวิร์กนี้จะเป็นตัว ELU นำมาแทนที่ตัว ReLU ซึ่งทั้ง 26 เลเยอร์จะเป็นตัวทำหน้าที่การสกัดคุณลักษณะที่สำคัญ (Feature Extraction) ของวิทยานิพนธ์นี้ และสุดท้ายชั้นการจำแนก (Classification Layer) ใช้ Softmax เข้ามาตอบออกมาเป็นประเภทสองคำตอบ (Binary Class) โดยกระจายออกมาเป็นแบบกระจายไปบนทุกพิกเซลบนภาพ (Pixel-wise Class Prediction Map) โดยจะแสดงเน็ตเวิร์กที่วิทยานิพนธ์นี้เสนอไว้รูปที่ 47 สำหรับการปรับพารามิเตอร์สำหรับการทำให้ดีที่สุด (Optimization) กับตัวโมเดลนั้น จะใช้วิธีการสโตแคสติกเกรเดียนต์เดสเซนท์ (Stochastic Gradient Descent หรือ SGD) โดยปรับค่าอัตราการเรียนรู้ที่ 0.001 และค่าโมเมนตัมอยู่ที่ 0.9 ทั้งนี้สาเหตุที่ปรับให้ค่าอัตราการเรียนรู้ต่ำมีจุดประสงค์เพื่อทำให้การเรียนรู้มีการลู่เข้าที่ดีขึ้นจากการหลีกเลี่ยงการติดอยู่ที่โลคอลลออปติมา (Local Optima)

4.6 การทดสอบ (Testing)

สำหรับการทดสอบจะใช้โมเดลที่ดีที่สุดในช่วงการเรียนรู้ของเครื่องโดยวิธีที่วิทยานิพนธ์นี้นำเสนอ (สถาปัตยกรรม ELU-DCED, สถาปัตยกรรม ELU-DCED-LMs, และสถาปัตยกรรม ELU-DCED-LMs-CRFs) จะพิจารณาโมเดลที่ให้ประสิทธิภาพการจำแนก (ค่า F1) มากที่สุดจากการประเมินบนชุดข้อมูลตรวจสอบ (Validation) เพื่อที่จะนำโมเดลที่ดีที่สุดไปทดสอบกับชุดข้อมูลทดสอบของภาพถ่ายระยะไกลที่เตรียมไว้

บทที่ 5

การทดลองและผลการทดลอง

ในส่วนของการทดลองจะถูกแบ่งออกเป็นทั้งหมด 3 การทดลอง โดยเปรียบเทียบระหว่างสิ่ง
ที่วิทยานิพนธ์นี้นำเสนอทั้ง 3 วิธีเทียบกับวิธีการเรียนรู้เชิงลึกมาตรฐาน ได้แก่ (1) การทดลองในส่วน
ของการเพิ่มประสิทธิภาพการจำแนกของสถาปัตยกรรมการเรียนรู้ที่ใช้ด้วยหน่วยเอกซ์โพเนนเชียลเชิง
เส้น (ELU) เรียกว่า สถาปัตยกรรม ELU-DCED การทดลองถัดไป (2) เป็นการทดลองในส่วนของการ
ประยุกต์ใช้ตัวชี้วัดภูมิทัศน์กับการเรียนรู้เชิงลึกที่นำเสนอ เรียกว่า สถาปัตยกรรม ELU-DCED-LMs
และสุดท้าย (3) เป็นการใช้นิเทศน์ CRFs เข้ามาเพิ่มความคมและความถูกต้องให้กับเอาต์พุตที่ได้
เรียกว่า สถาปัตยกรรม ELU-DCED-LMs-CRFs โดยแต่ละการทดลองจะทำอยู่บน 2 ชุดข้อมูล ได้แก่
1. ชุดข้อมูลภาพถ่ายทางอากาศ (ชุดข้อมูลของถนนของรัฐแมสซาชูเซตส์ (Massachusetts Roads
Data Set)) และ 2. ชุดข้อมูลภาพถ่ายทางดาวเทียม (ชุดข้อมูลของประเทศไทย (Thailand Roads
Data Set)) สำหรับการทดลองทั้งหมดใช้ภาษาไพธอน (Python) ในการเขียนโปรแกรม โดยใช้ไลบรารี
ที่ชื่อว่า ลาซานญา (Lasagne) ซึ่งถูกขยายมาจากไลบรารีเธียนโน (Theano) โดยทำงานบนสเปค
คอมพิวเตอร์ Intel Core i5-4590S Processor (6MCache, up to 3.70 GHz) โดยหน่วยความจำ
ของแรมเท่ากับ 32 GB และใช้ GPU ของการ์ดจอ GeForce GTX 960 (หน่วยความจำ 4 GB) และ
การ์ดจอ GeForce GTX 1080 (หน่วยความจำ 8 GB) จากค่าย Nvidia ในการทดลองทั้งหมด

5.1 ผลการทดลองกับชุดข้อมูลภาพถ่ายทางอากาศ

ในส่วนของผลการทดลองกับชุดข้อมูลภาพถ่ายทางอากาศ ซึ่งคุณสมบัติของภาพจะเป็นภาพ
ที่มีความละเอียดสูงมาก (Very High Resolution) ในงานวิทยานิพนธ์ใช้ชุดข้อมูลถนนรัฐ
แมสซาชูเซตส์ โดยผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอให้ประสิทธิภาพการจำแนกที่ดีกว่ามี
วิธีมาตรฐานทั้ง 4 ตัว ได้แก่ สถาปัตยกรรม Basic-model, สถาปัตยกรรม FCN-no-skip,
สถาปัตยกรรม FCN-8s และสถาปัตยกรรม SegNet โดยรูปที่ 51 ด้านซ้ายเป็นการแสดงกราฟค่า
ความสูญเสียโดยมีลักษณะของการลู่เข้าเมื่อทำการสอนในแต่ละรอบอีพอก และรูปทางด้านขวาแสดง
ค่าการประเมิน Precision, Recall และ F1 ในแต่ละรอบอีพอกในชุดข้อมูลตรวจสอบ เพื่อทำการ
เลือกโมเดลที่ดีที่สุดมาทำการทดสอบในชุดข้อมูลทดสอบโดยสำหรับชุดข้อมูลภาพถ่ายทางอากาศ
เลือกโมเดลที่ดีที่สุดในรอบอีพอกเท่ากับ 29

ตารางที่ 5 แสดงผลลัพธ์จากการทดลองทั้งหมดว่าวิธีการที่เรานำเสนอให้ผลลัพธ์การสกัดถนนออกจากภาพถ่ายระยะไกลชนะวิธีการมาตรฐานทุกการประเมิน โดยหัวข้อถัดไปจะกล่าวถึงผลการทดลองที่เกิดขึ้นของวิธีการที่วิทยานิพนธ์นี้แนะนำเสนอทั้งหมดโดยทำการทดลอง 3 การทดลอง ดังนี้

5.1.1 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ (ELU-DCED)

ผลการทดลองนี้แสดงให้เห็นถึงวิธีการที่เรานำเสนอวิธีแรก คือ การเปลี่ยนฟังก์ชันเรคตีไฟด์เชิงเส้นเติมในสถาปัตยกรรมแบบเอนโคเดอร์-ดีโคเดอร์ มาเป็นฟังก์ชันกระตุ้นตัวใหม่ที่เรียกว่าหน่วยเอกซ์โพเนนเชียลเชิงเส้น โดยจากผลการทดลองพบว่า วิธีการที่เรานำเสนอวิธีแรกนั้นให้ผลที่ดีกว่าวิธีการมาตรฐานตัวที่ดีที่สุดอย่าง SegNet โดยค่า F1 สูงกว่า 2.6 เปอร์เซ็นต์ (ELU-DCED ได้ค่าประเมิน F1 อยู่ที่ 0.788 และ SegNet ได้ค่า F1 อยู่ที่ 0.768 ตามลำดับ)

5.1.2 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์ (ELU-DCED-LMs)

จากผลการทดลองนี้แสดงให้เห็นถึงวิธีการที่เรานำเสนอวิธีสอง คือ การนำวิธีที่เราเสนอวิธีแรกร่วมกับการใช้ตัวชี้วัดภูมิทัศน์ เพื่อกำจัดวัตถุที่ไม่ใช่ถนนออกไปโดยอาศัยการดูจากค่าความซับซ้อนซึ่งเป็นหนึ่งในตัววัดของตัวชี้วัดภูมิทัศน์ โดยจากผลการทดลองพบว่า วิธีที่การที่เราเสนอนี้ให้ผลการประเมิน F1 สูงกว่าวิธีการที่เรานำเสนอวิธีแรกและวิธีการมาตรฐานที่ดีที่สุดอย่าง SegNet เท่ากับ 6.9 เปอร์เซ็นต์และ 8.9 เปอร์เซ็นต์ตามลำดับ (ELU-DCED-LMs ได้ค่าประเมิน F1 อยู่ที่ 0.857 โดยมากกว่า ELU-DCED ที่ได้ค่าประเมิน F1 อยู่ที่ 0.788 และ SegNet ได้ค่าประเมิน F1 อยู่ที่ 0.768)

5.1.3 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส (ELU-DCED-LMs-CRFs)

การทดลองสุดท้ายเป็นการรวมทุกวิธีที่เรานำเสนอเข้าด้วยกัน คือ วิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับการใช้ตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส โดยจากผลการทดลองพบว่า วิธีการที่เรานำเสนอนี้ได้รับการประเมินประสิทธิภาพการจำแนกมากกว่าการประเมินในทุกการทดลองที่ผ่านมาทั้งหมด โดยคะแนน F1 ชนะวิธีการมาตรฐานที่ดีที่สุดอยู่ที่ 10.8 เปอร์เซ็นต์ และผลการทดลองแสดงให้เห็นว่าคะแนน

Precision อยู่ที่ 0.858 และคะแนน Recall อยู่ที่ 0.894 ซึ่งสูงกว่าทุกวิธีที่นำเสนอและทุกวิธีการมาตรฐาน โดยรูปที่ 52 เป็นการแสดงผลการทดลองทั้งหมดของชุดข้อมูลภาพถ่ายทางอากาศ จากผลการทดลองแสดงให้เห็นว่าผลลัพธ์จากวิธี ELU-DCED-LMs-CRFs รูปขวาสุดให้ผลลัพธ์ได้ใกล้เคียงภาพผลเฉลยมากที่สุด โดยแสดงผลจากวิธีที่นำเสนอด้วยวิธี ELU-DCED, ELU-DCED-LMs และ ELU-DCED-LMs-CRFs จากภาพกลางไปจนถึงภาพขวาตามลำดับ

ตารางที่ 3 จำนวนชุดข้อมูลในข้อมูลการสอน, ข้อมูลตรวจสอบ และข้อมูลทดสอบ

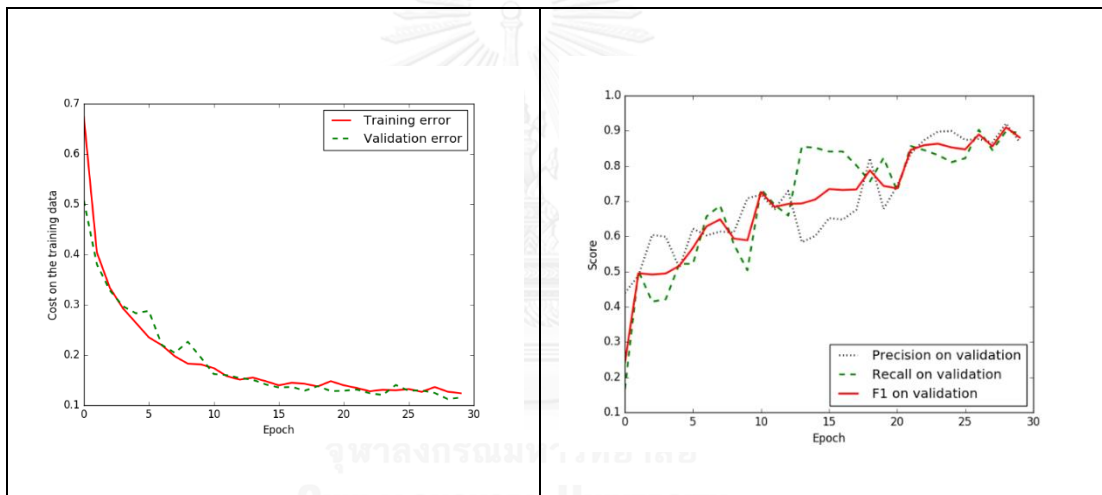
	ชุดข้อมูลฝึกสอน	ชุดข้อมูลตรวจสอบ	ชุดข้อมูลทดสอบ
แมสซาชูเซตส์	1,108	14	49
นครปฐม	200	14	49
ชลบุรี	100	14	49
สงขลา	100	14	49
สุรินทร์	70	14	49
อุบลราชธานี	70	14	49

ตารางที่ 4 จำนวนชุดข้อมูลในข้อมูลการสอนหลังจากการหมุนภาพ, ข้อมูลตรวจสอบและข้อมูลทดสอบ

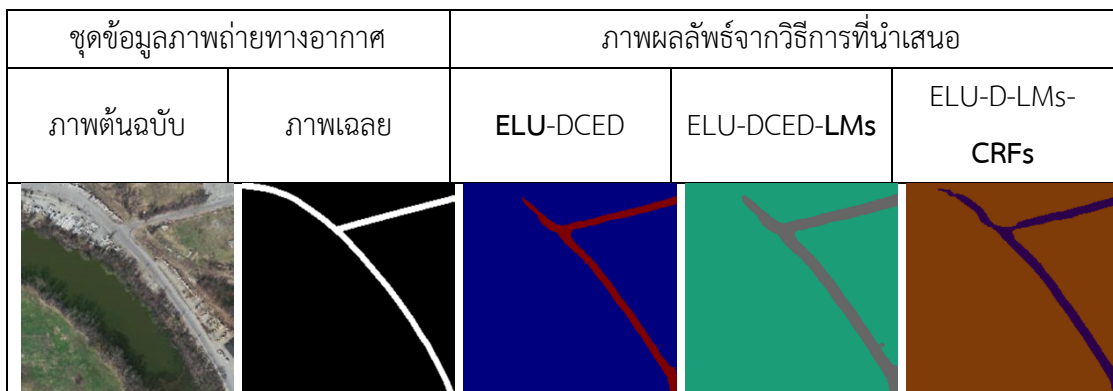
	ชุดข้อมูลฝึกสอน	ชุดข้อมูลตรวจสอบ	ชุดข้อมูลทดสอบ
แมสซาชูเซตส์	8,864	14	49
นครปฐม	1,600	14	49
ชลบุรี	800	14	49
สงขลา	800	14	49
สุรินทร์	560	14	49
อุบลราชธานี	560	14	49



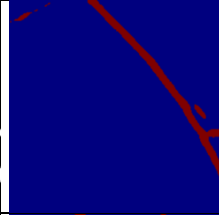
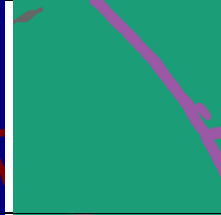








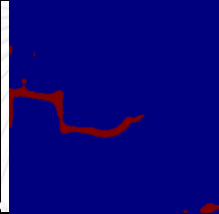
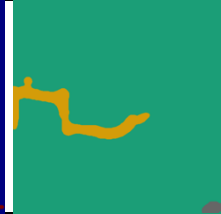


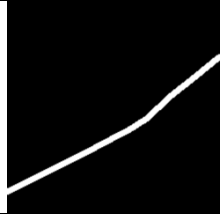
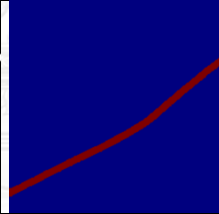
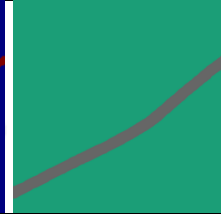
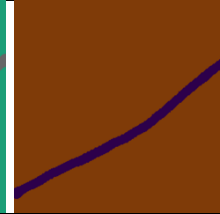


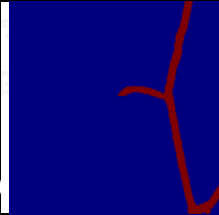


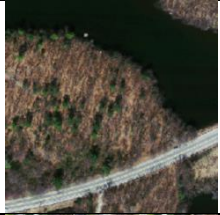
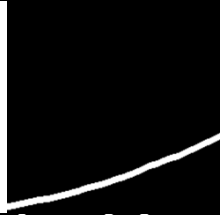
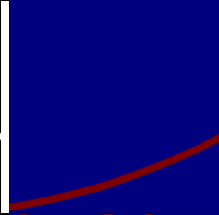
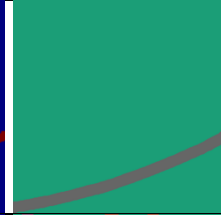
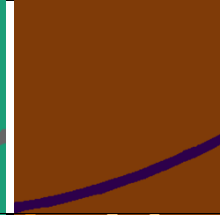


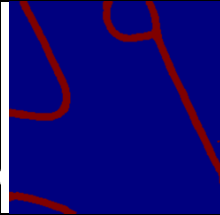
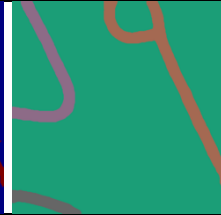
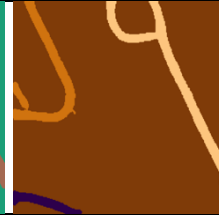
ตารางที่ 5 ประสิทธิภาพการจำแนกโดยทดสอบกับชุดข้อมูลภาพถ่ายทางอากาศ



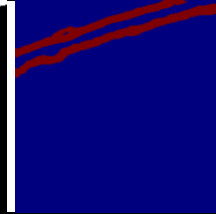
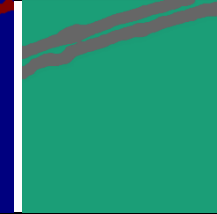
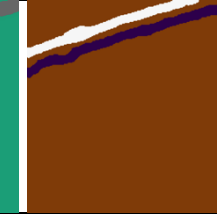


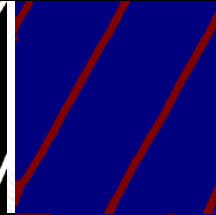
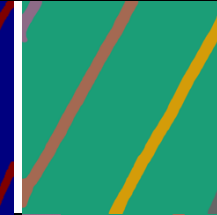
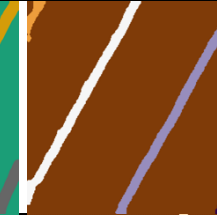




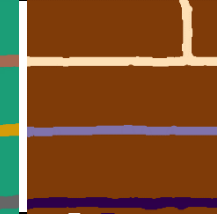


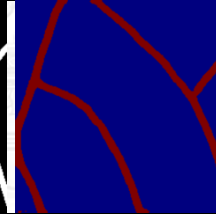
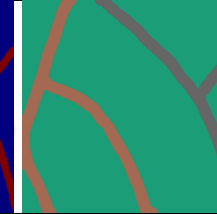



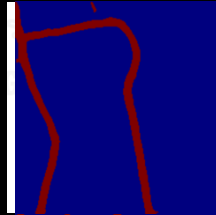
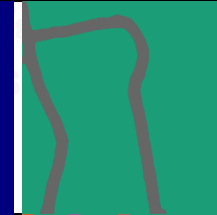


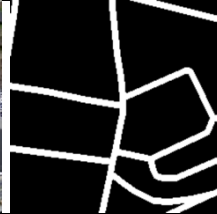
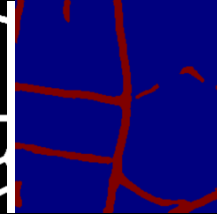
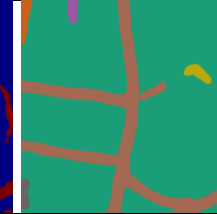

	โมเดล	Precision	Recall	F1
วิธีการมาตรฐาน	Basic-model	0.657	0.657	0.657
	FCN-no-skip	0.742	0.742	0.742
	FCN-8s	0.762	0.762	0.762
	SegNet	0.773	0.765	0.768
วิธีการที่นำเสนอ	ELU-DCED	0.852	0.733	0.788
	ELU-DCED-LMs	0.854	0.861	0.857
	ELU-DCED-LMs-CRFs	0.858	0.894	0.876



รูปที่ 51 กราฟค่าความสูญเสียในแต่ละรอบอีพอค (รูปซ้าย) และกราฟของการประเมินค่า Precision Recall และ F1 บนชุดข้อมูลวาลิตเดชั่นเพื่อเลือกโมเดลที่ดีที่สุด (รูปขวา)



ชุดข้อมูลภาพถ่ายทางอากาศ		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				
				

ชุดข้อมูลภาพถ่ายทางอากาศ		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				

รูปที่ 52 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายทางอากาศ

5.2 ผลการทดลองกับชุดข้อมูลภาพถ่ายดาวเทียม

ในส่วนของผลการทดลองกับชุดข้อมูลภาพถ่ายดาวเทียม ซึ่งคุณสมบัติของภาพจะเป็นภาพที่มีความละเอียดสูง (High Resolution) งานวิทยานิพนธ์ใช้ชุดข้อมูลของประเทศไทยเป็นชุดข้อมูลที่ถูกถ่ายโดยดาวเทียมไทยโชต (Thaichote) หรือดาวเทียมธีออส (THEOS) โดยภายในชุดข้อมูลจะประกอบด้วย 5 สถานที่ คือ พุทธมณฑล (นครปฐม), ศรีราชา (ชลบุรี), สงขลา, สุรินทร์ และ อุบลราชธานี โดยผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอให้ประสิทธิภาพของการจำแนกที่ดีกว่าวิธีมาตรฐานที่ดีที่สุดอย่างสถาปัตยกรรม SegNet โดยรูปที่ 53 ด้านซ้ายเป็นการแสดงกราฟค่าความสูญเสียโดยมีลักษณะของการลู่เข้าเมื่อทำการสอนในแต่ละรอบอีพอค และรูปทางด้านขวาแสดงค่าการประเมิน Precision, Recall และ F1 ในแต่ละรอบอีพอคในชุดข้อมูลตรวจสอบ เพื่อทำการเลือกโมเดลที่ดีที่สุดมาทำการทดสอบในชุดข้อมูลทดสอบ โดยรูปภาพแต่ละแถวอ้างอิงถึงแต่ละชุดข้อมูลของแต่ละจังหวัด โดยสำหรับชุดข้อมูลภาพถ่ายดาวเทียมเลือกโมเดลที่ดีที่สุดในรอบอีพอคของแต่ละชุดข้อมูล (พุทธมณฑล (นครปฐม), ศรีราชา (ชลบุรี), สงขลา, สุรินทร์ และอุบลราชธานี) เท่ากับ 25, 15, 30, 21 และ 20 ตามลำดับ สุดท้ายจากตารางที่ 6, 7 และ 8 แสดงผลลัพธ์จากการทดลองทั้งหมดว่าวิธีการที่เรานำเสนอให้ผลลัพธ์การสกัดถนนออกจากภาพถ่ายระยะไกลชนะวิธีการมาตรฐานทุกการประเมิน โดยรายละเอียดของการทดลองจะอธิบายในหัวข้อถัดไป

5.2.1 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ (ELU-DCED)

ผลการทดลองนี้แสดงให้เห็นถึงวิธีการที่เรานำเสนอวิธีแรก คือ การเปลี่ยนฟังก์ชันเรกติไฟด์เชิงเส้นเดิมในสถาปัตยกรรมแบบเอนโคเดอร์-ดีโคเดอร์ มาเป็นฟังก์ชันกระตุ้นตัวใหม่ที่เรียกว่าหน่วยเอกซ์โพเนนเชียลเชิงเส้น โดยจากผลการทดลองพบว่า วิธีการที่เรานำเสนอวิธีแรกนั้นให้ผลการประเมินเฉลี่ยในทุกชุดข้อมูลภาพถ่ายดาวเทียมที่ดีกว่าวิธีการมาตรฐานตัวที่ดีที่สุดอย่าง SegNet โดยค่า F1 สูงกว่า 9.08 เปอร์เซ็นต์ โดยเฉพาะกับชุดข้อมูลของจังหวัดอุบลราชธานีและจังหวัดชลบุรี ชนะวิธีการมาตรฐานอยู่ที่คะแนน F1 มากกว่า 10 เปอร์เซ็นต์ (10.48% และ 10.68% ตามลำดับ)

5.2.2 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์ (ELU-DCED-LMs)

จากผลการทดลองนี้แสดงให้เห็นถึงวิธีการที่เรานำเสนอวิธีสอง คือ การนำวิธีที่เราเสนอวิธีแรกร่วมกับการใช้ตัวชี้วัดภูมิทัศน์ เพื่อกำจัดวัตถุที่ไม่ใช่ถนนออกไปโดยอาศัยการดูจากค่าความซับซ้อนซึ่งเป็นหนึ่งในตัววัดของตัวชี้วัดภูมิทัศน์ โดยจากผลการประเมินเฉลี่ยในทุกชุดข้อมูลภาพถ่าย

ดาวเทียมพบว่า วิธีที่การที่เราเสนอนี้ให้ผลการประเมิน F1 สูงกว่าวิธีการที่เราแนะนำวิธีแรกและวิธีการมาตรฐานที่ดีที่สุดอย่าง SegNet เท่ากับ 3.2 เปอร์เซ็นต์และ 12 เปอร์เซ็นต์ตามลำดับ

5.2.3 ผลการทดลองกับวิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส (ELU-DCED-LMs-CRFs)

การทดลองสุดท้ายเป็นการรวมทุกวิธีที่เราแนะนำเข้าด้วยกัน คือ วิธีการเพิ่มประสิทธิภาพการจำแนกให้กับสถาปัตยกรรมการเรียนรู้เชิงลึกแบบเอนโคเดอร์-ดีโคเดอร์ร่วมกับการใช้ตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส โดยจากผลการประเมินเฉลี่ยในทุกชุดข้อมูลภาพถ่ายดาวเทียมพบว่า วิธีที่การที่เราแนะนำนี้ได้รับการประเมินประสิทธิภาพการจำแนกมากกว่าการประเมินในทุกการทดลองที่ผ่านมาทั้งหมด โดยคะแนน F1 ชนะวิธีการที่เราแนะนำมาก่อนหน้า (ELU-DCED และ ELU-DCED-LMs) และวิธีการมาตรฐานที่ดีที่สุดอยู่ที่ 6.28 เปอร์เซ็นต์ 9.44 เปอร์เซ็นต์ และ 18.48 เปอร์เซ็นต์ และผลการทดลองแสดงให้เห็นว่าคะแนน Precision เฉลี่ยอยู่ที่ 0.751 และคะแนน Recall เฉลี่ยอยู่ที่ 0.584 ซึ่งสูงกว่าทุกวิธีที่แนะนำและทุกวิธีการมาตรฐาน โดยรูปที่ 54 จนถึงรูปที่ 58 เป็นการแสดงผลการทดลองทั้งหมดของชุดข้อมูลภาพถ่ายดาวเทียมชื่ออส จากผลการทดลองแสดงให้เห็นว่าผลลัพธ์จากวิธี ELU-DCED-LMs-CRFs รูปขวาสุดให้ผลลัพธ์ได้ใกล้เคียงภาพผลเฉลยมากที่สุด โดยแสดงผลการจากวิธีที่แนะนำด้วยวิธี ELU-DCED, ELU-DCED-LMs และ ELU-DCED-LMs-CRFs จากภาพกลางไปจนถึงภาพขวาตามลำดับ

ตารางที่ 6 ประสิทธิภาพการจำแนกโดยวัดจากค่า Precision

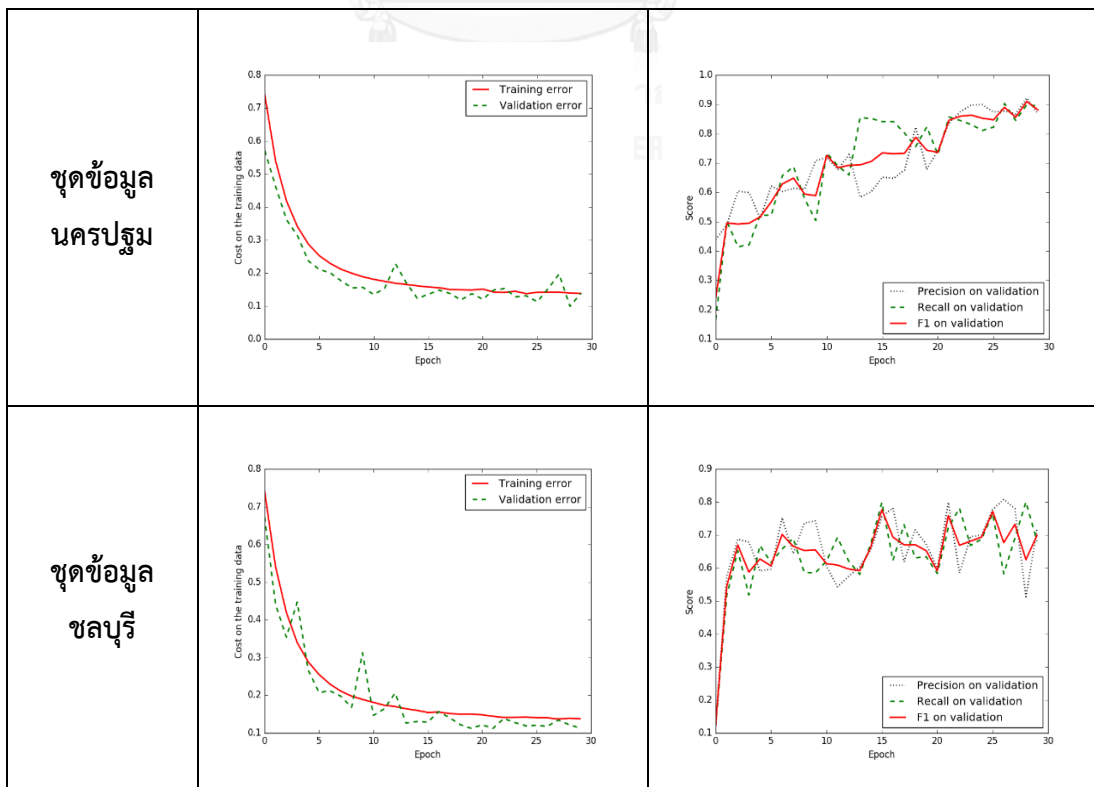
	โมเดล	นครปฐม	ชลบุรี	สงขลา	สุรินทร์	อุบล	เฉลี่ย
วิธีการมาตรฐาน	SegNet	0.435	0.668	0.456	0.598	0.601	0.55
วิธีการที่แนะนำ	ELU-DCED	0.41	0.702	0.478	0.84	0.852	0.66
	ELU-DCED-LMs	0.494	0.852	0.557	0.77	0.867	0.71
	ELU-DCED-LMs-CRFs	0.535	0.909	0.65	0.786	0.871	0.75

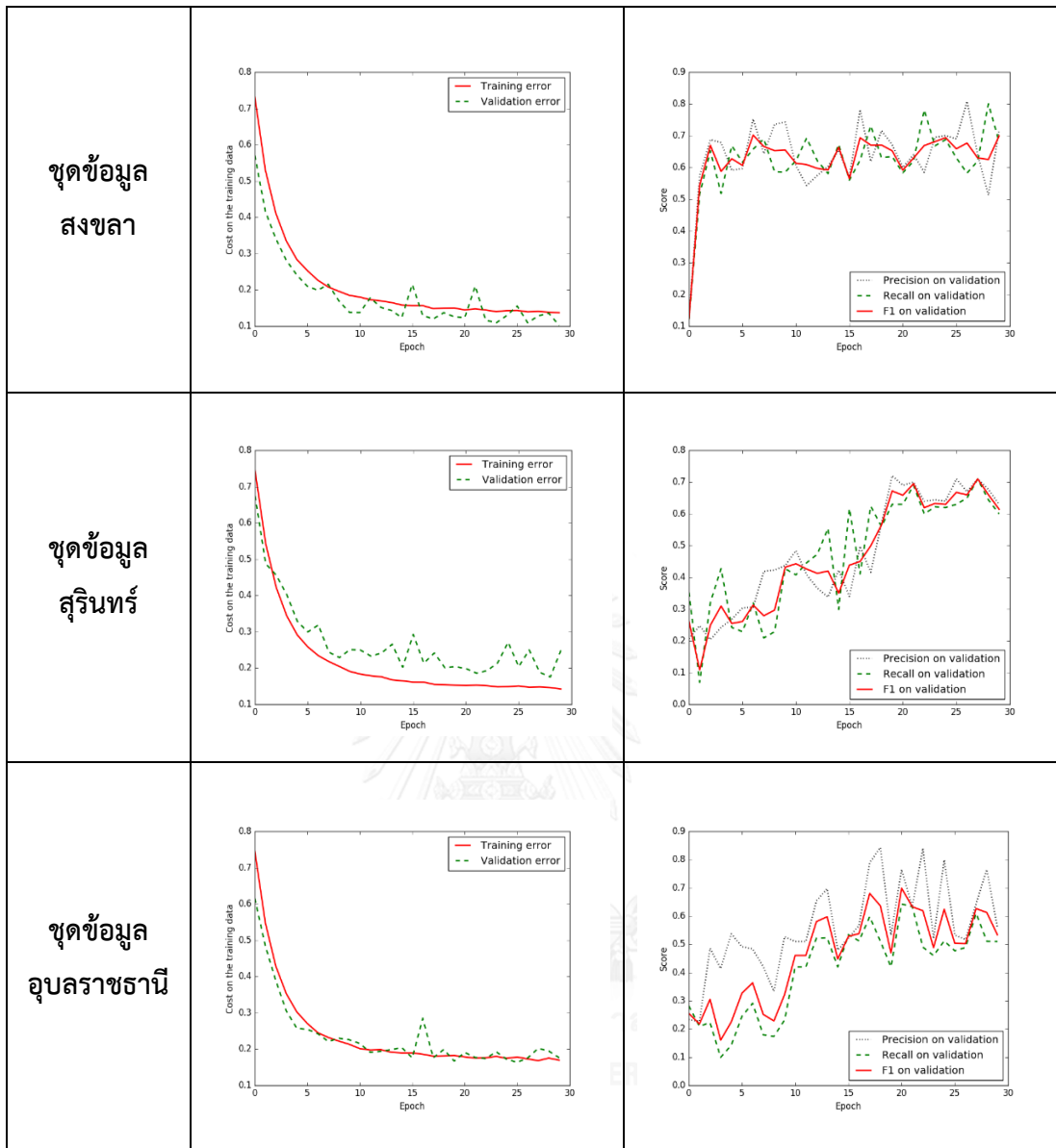
ตารางที่ 7 ประสิทธิภาพการจำแนกโดยวัดจากค่า Recall

	โมเดล	นครปฐม	ชลบุรี	สงขลา	สุรินทร์	อุบล	เฉลี่ย
วิธีการมาตรฐาน	SegNet	0.41	0.499	0.395	0.431	0.306	0.41
วิธีการที่นำเสนอ	ELU-DCED	0.532	0.678	0.517	0.456	0.389	0.52
	ELU-DCED-LMs	0.483	0.642	0.498	0.526	0.416	0.51
	ELU-DCED-LMs-CRFs	0.566	0.676	0.57	0.643	0.467	0.58



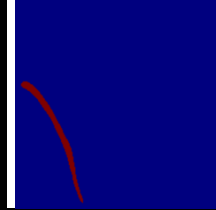
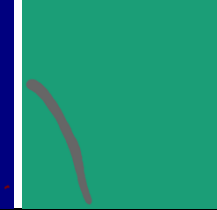
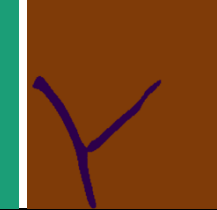


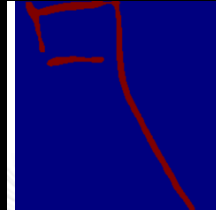
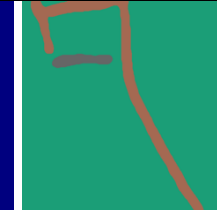
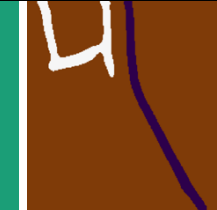


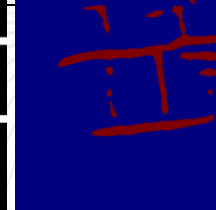

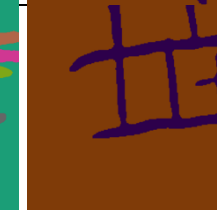
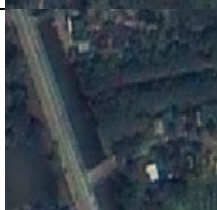

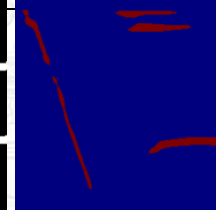
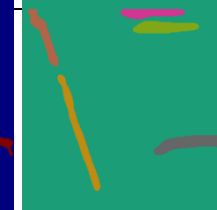
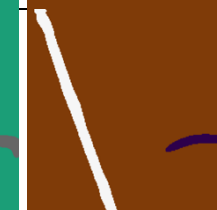


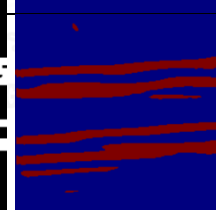
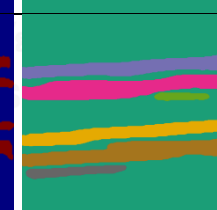



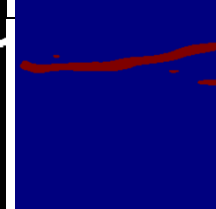
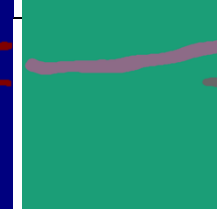
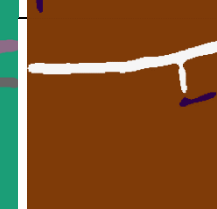
ตารางที่ 8 ประสิทธิภาพการจำแนกโดยวัดจากค่า F1

	โมเดล	นครปฐม	ชลบุรี	สงขลา	สุรินทร์	อุบล	เฉลี่ย
วิธีการมาตรฐาน	SegNet	0.422	0.572	0.424	0.501	0.406	0.47
วิธีการที่นำเสนอ	ELU-DCED	0.463	0.69	0.497	0.591	0.534	0.56
	ELU-DCED-LMs	0.488	0.732	0.526	0.625	0.562	0.59
	ELU-DCED-LMs-CRFs	0.55	0.775	0.607	0.707	0.608	0.65



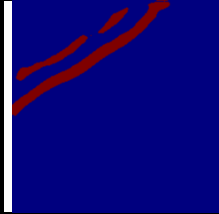
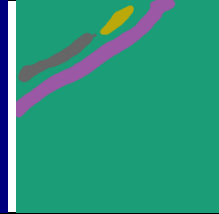
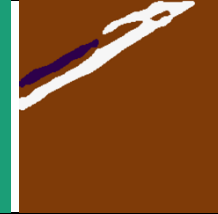


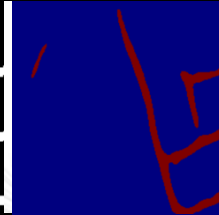
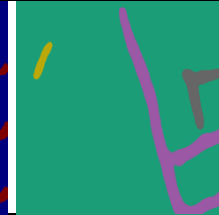



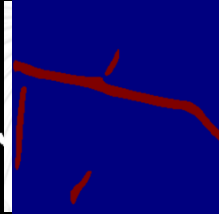
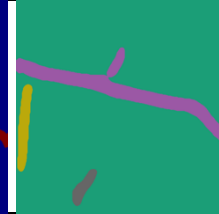
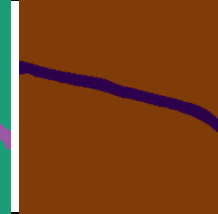


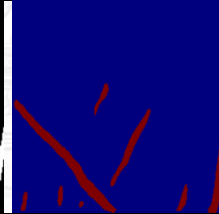

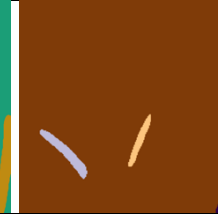


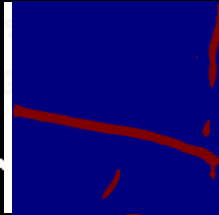
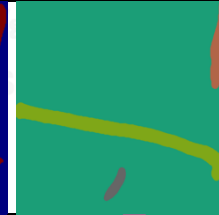
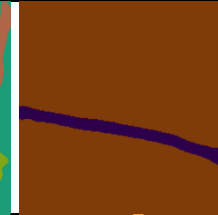









รูปที่ 53 กราฟค่าความสูญเสียในแต่ละรอบอีพอค (รูปซ้าย) และกราฟของการประเมินค่า Precision Recall และ F1 บนชุดข้อมูลวาลิดเดชั่นเพื่อเลือกโมเดลที่ดีที่สุด (รูปขวา)

ชุดข้อมูลภาพถ่ายดาวเทียม		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				



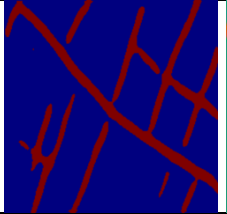




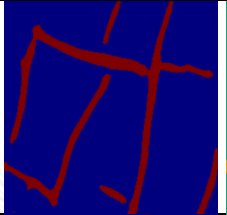






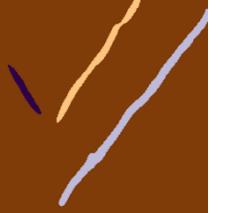









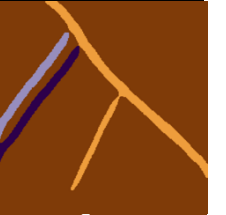




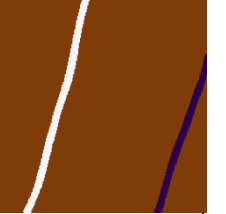
รูปที่ 54 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัด นครปฐม)

ชุดข้อมูลภาพถ่ายดาวเทียม		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				



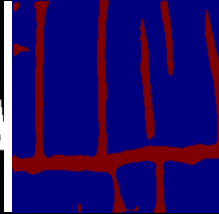




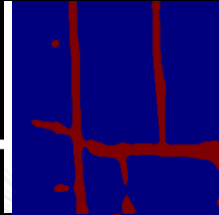
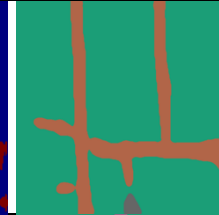
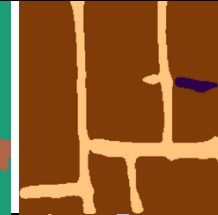


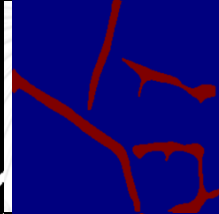




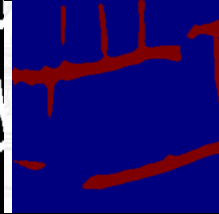
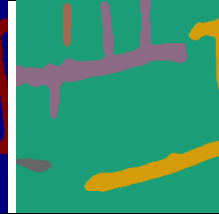



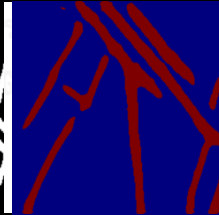
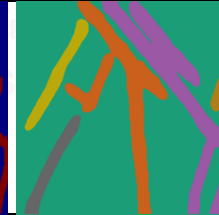



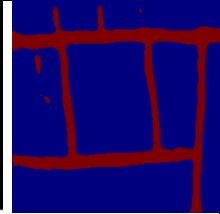

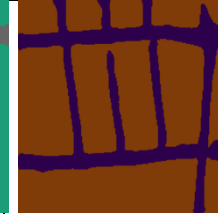
รูปที่ 55 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัดชลบุรี)

ชุดข้อมูลภาพถ่ายดาวเทียม		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs

รูปที่ 56 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัดสงขลา)

ชุดข้อมูลภาพถ่ายดาวเทียม		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				

รูปที่ 57 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัดสุรินทร์)

ชุดข้อมูลภาพถ่ายดาวเทียม		ภาพผลลัพธ์จากวิธีการที่นำเสนอ		
ภาพต้นฉบับ	ภาพเฉลย	ELU-DCED	ELU-DCED-LMs	ELU-D-LMs-CRFs
				
				
				
				
				
				

รูปที่ 58 ตัวอย่างผลลัพธ์จากวิธีการที่นำเสนอโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียม (ชุดข้อมูลจังหวัด อุบลราชธานี)

บทที่ 6

สรุปการวิจัยและแนวทางการวิจัยในชั้นถัดไป

6.1 สรุปการวิจัย

วิทยานิพนธ์นี้นำเสนอวิธีการสกัดถนนออกจากภาพถ่ายระยะไกล (ภาพถ่ายทางอากาศและภาพถ่ายดาวเทียม) ด้วยการใช้การเรียนรู้เชิงลึกแบบนิรอรเน็ตเวิร์กคอนโวลูชันร่วมกับตัวชี้วัดภูมิทัศน์และคอนดิชันนอลแรนดอมฟิลด์ส โดยเลือกใช้ฟังก์ชันกระตุ้นตัวใหม่ที่เรียกว่า หน่วยเอกซ์โพเนนเชียลเชิงเส้นเข้ามาแทนที่ฟังก์ชันเรคตีไฟด์เชิงเส้นเดิมในสถาปัตยกรรมแบบเอนโคเดอร์-ดีโคเดอร์ เพื่อเพิ่มจำนวนวัตถุถนนให้ได้มากยิ่งขึ้น ถัดมาใช้ตัวชี้วัดภูมิทัศน์ในการกำจัดวัตถุที่ไม่ใช่ถนนออกจากภาพถ่ายระยะไกล และสุดท้ายใช้วิธีคอนดิชันนอลแรนดอมฟิลด์สเข้ามาช่วยทำให้วัตถุถนนมีความคมและความถูกต้องมากยิ่งขึ้น จากการทดสอบวิธีการที่นำเสนอกับข้อมูลภาพถ่ายระยะไกลปรากฏว่าวิธีการทั้งหมดที่นำเสนอมีประสิทธิภาพการจำแนกที่ดีกว่าวิธีมาตรฐานในทุกการประเมิน

ประโยชน์ที่ได้จากการใช้นิรอรเน็ตเวิร์กคอนโวลูชันร่วมกับวิธีการที่วิทยานิพนธ์นี้นำเสนอ คือ ช่วยลดขั้นตอนในการผลิตชั้นถนนโดยที่ผ่านมาจำเป็นต้องใช้ทักษะและความรู้ของผู้ที่ชำนาญการในการจำแนกถนนออกจากภาพถ่ายระยะไกล ซึ่งใช้เวลาและทรัพยากรค่อนข้างมาก และเป็นการเพิ่มประสิทธิภาพในการผลิตชั้นข้อมูลถนนให้ดียิ่งขึ้น

6.2 ข้อเสนอแนะและข้อจำกัดในงานวิทยานิพนธ์นี้

สถาปัตยกรรมของการเรียนรู้เชิงลึกที่ใช้ในงานวิทยานิพนธ์นี้ให้ประสิทธิภาพการจำแนกที่ดีกับชุดข้อมูลภาพถ่ายระยะไกลที่มีความละเอียดสูงมากและชนิดที่มีความละเอียดสูงเท่านั้น อีกทั้งค่าพารามิเตอร์ในส่วนของการเรียนรู้เชิงลึกและส่วนของการประมวลผลหลังจะถูกปรับโดยเลือกการตั้งค่าที่ดีที่สุดให้กับแต่ละชุดข้อมูลทั้งภาพถ่ายทางอากาศและภาพถ่ายดาวเทียมที่ใช้ในวิทยานิพนธ์นี้ ดังนั้น การที่จะนำวิธีการที่วิทยานิพนธ์นี้ไปใช้กับชุดข้อมูลภาพถ่ายระยะไกลอื่นๆ อาจจะต้องมีการปรับพารามิเตอร์ใหม่อีกครั้งเพื่อให้ได้ประสิทธิภาพการจำแนกที่ดีกับชุดข้อมูลภาพถ่ายระยะไกลนั้นๆ

6.3 แนวทางวิจัยในชั้นถัดไป

เนื่องจากวิธีการที่นำเสนอใช้นิรอรเน็ตเวิร์กเพียงสองคลาส คือ คลาสที่เป็นถนนและคลาสที่ไม่ใช่ถนน โดยยังมีอีกหลายคลาสในชั้นข้อมูลภูมิสารสนเทศที่ใช้การเรียนรู้เชิงลึกเข้ามาสกัดออกมา

จากภาพถ่ายระยะไกลได้ เช่น คลาสของพื้นที่ป่าไม้ คลาสของพื้นที่เกษตรกรรม คลาสของพื้นที่ชุมชน หรือคลาสของกลุ่มแม่น้ำ โดยนำวิธีการที่นำเสนอมาสกัดคลาสเหล่านี้รวมถึงนำไปสกัดแบบหลายคลาสได้ด้วย ซึ่งแนวทางการพัฒนาต่อทำได้โดยนำวิธีการเรียนรู้เชิงลึกที่วิทยานิพนธ์นี้เสนอไปสกัดทุกคลาสที่กล่าวมาข้างต้นได้ โดยอาจต้องปรับพารามิเตอร์ในส่วนของ การเรียนรู้เชิงลึกเพิ่มเติม ซึ่งจะส่งผลให้การสกัดวัตถุต่างๆ ออกจากภาพถ่ายระยะไกลมีประสิทธิภาพการจำแนกที่ดีขึ้น



รายการอ้างอิง

- [1] J. S. J. Wang, M. Chen, and Z. Yang, "Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine," *International Journal of Remote Sensing*, vol. 36, 2015.
- [2] X. H. a. L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines," *International Journal of Remote Sensing*, vol. 30, pp. 1977-1987, 2009.
- [3] T. Y. a. Y. A. S. Saito, "Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks," *Journal of Imaging Science and Technology*, vol. 60, pp. 1-9, 2016.
- [4] A. K. V. Badrinarayanan, R. Cipolla, and S. Member, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *arXiv*, 2016.
- [5] C. Poullis, "Tensor-Cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 95, pp. 93-108, 2014.
- [6] P. V. a. K. J. T. Panboonyuen, "Image Vectorization of Road Satellite Data Sets," *Journal of Remote Sensing and GIS Association of Thailand*, vol. 17, pp. 310-321, 2016.
- [7] P. R. G. R. Maurya, and A. S. Shukla, "Road Extraction Using K-Means Clustering and Morphological Operations," presented at the International Conference on Intelligent Information Processing, 2011.
- [8] D. Z. a. X. L. G. Xu, "Road Extraction in High Resolution Images from Google Earth," presented at the The International Conference on Information and Communication Systems, 2009.
- [9] X. L. Z. Liu, P. Luo, C. C. Loy, and X. Tang, "Deep Learning Markov Random Field for Semantic Segmentation," *arxiv*, 2016.

- [10] E. S. J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," presented at the Conference on Computer Vision and Pattern Recognition, 2015.
- [11] A. H. V. Badrinarayanan, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling," *arXiv*, 2015.
- [12] A. K. V. Badrinarayanan, R. Cipolla, and S. Member, "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding," *arxiv*, 2016.
- [13] S. Muruganandham, "Semantic Segmentation of Satellite Images using Deep Learning," Master, Cybernetics, Czech Technical University in Prague & Luleå University of Technology, 2016.
- [14] S. H. H. Noh, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," presented at the International Conference on Computer Vision, 2015.
- [15] K. Mcgarigal, *Landscape metrics for categorical map patterns*. 2001.
- [16] G. P. L. Chen, S. Member, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *arxiv*, 2016.
- [17] P. Krähenbühl, and Vladlen Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *Advances in neural information processing systems*, 2011.
- [18] L.-C. Chen, et al, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv*, 2014.
- [19] V. Mnih, "Machine Learning for Aerial Image Labeling," University of Toronto, University of Toronto, 2013.
- [20] T. Koomsubha, "Text Categorization for Thai Corpus Using Character-Level Convolutional Neural Network," Master, Department of Computer Engineering, Chulalongkorn University, 2016.
- [21] F.-F. L. A. K. J. Johnson, Deep Learning Lecture, 2016. [Online]. Available.

- [22] I. Wichakam, "Breast Cancer Detection in Tomosynthesis Imaging Using Deep Convolutional Network," Department of Computer Engineering, Chulalongkorn University, 2017.
- [23] D. P. Kingma, "Adam: A Method for Stochastic Optimization," presented at the International Conference on Learning Representations, 2015.
- [24] S. A. E. M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, pp. 98-136, 2015.
- [25] K. a. Z. Simonyan, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," presented at the International Conference on Learning Representations, 2015.
- [26] A. K. A. Kalinovsky, and V. Kovalev, "Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks," presented at the International Conference on Pattern Recognition and Information Processing, 2016.
- [27] C. e. al, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," presented at the International Conference on Learning Representations, 2016.
- [28] C. S. a. D. Selvathi, "Connected component-based technique for automatic extraction of road centerline in high resolution satellite images," *EURASIP Journal on Image and Video Processing*, 2015.

ภาคผนวก



ประวัติผู้เขียนวิทยานิพนธ์

นายธีรพงศ์ ปานบุญยืน เกิดวันที่ 5 ธันวาคม 2535 ที่จังหวัดราชบุรี สำเร็จการศึกษาปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ จากภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือในปีการศึกษา 2557 และเข้าศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2558 โดยได้รับทุนอุดหนุนการศึกษาระดับบัณฑิตศึกษาจากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัวทรงเจริญพระชนมายุครบ 72 พรรษา

