

การสร้างลาเบลสังเคราะห์สำหรับการจำแนกภาพถ่ายดาวเทียมผ่านการเรียนรู้แบบมีผู้สอนด้วยการแทนค่าด้วยตัวเอง

Generating synthetic labels for satellite image classification through self-supervised representation learning

ศรัณย์ กุลยานนท์¹, วสิท ลิ้มประเสริฐ¹, ปกป้อง ส่องเมือง² และ รัชฎา คงคะจันทร์¹

Sarun Gulyanon¹, Wasit Limprasert¹, Pokpong Songmuang² and Rachada Kongkachandra¹

Received: 29 November 2023 ; Revised: 15 January 2024 ; Accepted: 9 February 2024

บทคัดย่อ

ในการวิเคราะห์ภาพถ่ายดาวเทียมมักนิยมใช้เทคนิคการเรียนรู้เชิงลึกแบบมีผู้สอน แต่เทคนิคนี้ต้องใช้ชุดข้อมูลที่มีลาเบล (Label) จำนวนมาก และการเก็บข้อมูลลาเบลมีค่าใช้จ่ายสูงทั้งในด้านแรงงานคนและทรัพยากรอื่น ๆ ในขณะที่ปัจจุบันมีฐานข้อมูลภาพถ่ายดาวเทียมที่ไม่มีข้อมูลลาเบลเป็นจำนวนมาก ที่สามารถใช้ได้ทั้งในเชิงพาณิชย์และทางวิชาการ เพื่อแก้ไขปัญหาการขาดแคลนข้อมูลลาเบลอย่างในโจทย์การจำแนกภาพถ่ายดาวเทียม งานวิจัยนี้จึงนำเสนอกระบวนการเพื่อใช้ประโยชน์จากข้อมูลที่ไม่มีลาเบลเหล่านี้ โดยอาศัยเทคนิคการเรียนรู้แทนภาพด้วยตนเอง (Self-Supervised Representation Learning) เพื่อสร้างข้อมูลลาเบลเทียมที่ทำหน้าที่เป็นชุดข้อมูลสอนสำหรับเทรนโมเดลการเรียนรู้แบบมีผู้สอนอีกที ผลการทดลองแสดงให้เห็นว่าโมเดลที่เทรนด้วยข้อมูลลาเบลเทียมมีประสิทธิภาพที่ใกล้เคียงกับโมเดลที่เทรนด้วยข้อมูลลาเบลจริง แต่ใช้ข้อมูลลาเบลน้อยกว่าถึง 9 เท่า โดยมีความแม่นยำ 75% บนชุดข้อมูลไร่ปาล์มน้ำมัน และ 86% บนชุดข้อมูลป่าแอมะซอน นอกจากนี้กระบวนการสร้างข้อมูลลาเบลเทียมยังให้เวกเตอร์คุณลักษณะ (Feature Vector) ที่มีความยืดหยุ่นและสามารถถ่ายทอดความรู้ไปใช้ในการแก้ปัญหาอื่นที่คล้ายกันได้อีกด้วย

คำสำคัญ: การเรียนรู้เชิงลึก, การจำแนกภาพ, การสร้างข้อมูลลาเบล, ภาพถ่ายดาวเทียม, การเรียนรู้ด้วยตนเอง, เวกเตอร์คุณลักษณะ

Abstract

Supervised deep learning techniques are state-of-the-art methods in satellite image analysis; however, this technique requires a large, labeled dataset. The acquisition of such datasets is expensive in terms of both manpower and resources. Conversely, there is an abundance of raw satellite images, available for both commercial and academic purposes. To address the scarcity of labeled data in satellite image classification, this study presents a novel method to utilize these unlabeled data. It uses self-supervised learning technique to create synthetic labels that act as a training dataset for supervised learning models. Experimental results show that models trained with synthetic labels perform comparably to those trained with real labels, using 9 times fewer labeled data, achieving 75% accuracy on the palm oil plantation dataset and 86% accuracy on the Amazon rainforest dataset. Additionally, the process of generating synthetic labels yields versatile and knowledge-transferable visual representation vectors.

Keywords: Deep learning, image classification, label synthesis, satellite imaging, self-supervised learning, feature vector

¹ วิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์, ปทุมธานี 12120

² คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์, ปทุมธานี 12120

¹ College of Interdisciplinary Studies, Thammasat University, Pathum Thani 12120

² Faculty of Science and Technology, Thammasat University, Pathum Thani 12120

* Corresponding author: Email sarung@tu.ac.th

บทนำ

ในยุคของข้อมูลใหญ่ มีข้อมูลอยู่มากมหาศาลโดยเฉพาะข้อมูลจากการสำรวจระยะไกล ที่มีภาพถ่ายดาวเทียมขนาดเพตะไบต์ (เช่น ข้อมูลภาพจาก Sentinel-2) ที่เปิดให้สาธารณะได้ใช้ประโยชน์ในขณะเดียวกันการคิดค้นเทคนิคการเรียนรู้เชิงลึก (deep learning) เมื่อเกือบสิบปีที่ผ่านมา ทำให้เรามีเครื่องมือที่สามารถใช้ประโยชน์จากข้อมูลปริมาณมหาศาลนี้ได้ การเรียนรู้ของเครื่อง (machine learning) โดยเฉพาะการเรียนรู้แบบมีผู้สอน (supervised learning) กลายเป็นเทคนิคที่ทรงพลังที่สุดในหลายด้าน เช่น คอมพิวเตอร์วิทัศน์ (Russakovsky *et al.*, 2015) และการสำรวจระยะไกล (Demir *et al.*, 2018) การเรียนรู้แบบมีผู้สอนสามารถแก้ไขข้อปัญหาใด ๆ ที่สามารถอธิบายได้ด้วยคู่ของข้อมูลนำเข้าและผลลัพธ์ที่คาดหวังจากข้อมูลนำเข้า นั้น เมื่อมีจำนวนข้อมูลตัวอย่างและทรัพยากรคำนวณเพียงพอ

ปัญหาหลักของการใช้เทคนิคการเรียนรู้เชิงลึก คือ ปริมาณชุดข้อมูลลาเบลที่มีไม่เพียงพอ สำหรับโจทย์ทั่วไป เช่น การตรวจหาวัตถุและระบุตำแหน่ง ที่สามารถนำไปประยุกต์ใช้งานได้หลากหลาย นั้นไม่มีปัญหาเรื่องปริมาณข้อมูลลาเบลเนื่องจากมีนักวิจัยจำนวนมากให้ความสนใจโจทย์เหล่านี้ และช่วยกันสร้างข้อมูลลาเบลปริมาณมากได้ ดังที่เห็นได้ในงานวิจัยของ Russakovsky *et al.* (2015) แต่สำหรับงานที่เฉพาะเจาะจง การเก็บข้อมูลลาเบลเป็นเรื่องท้าทาย เพราะมีผู้ให้ความสนใจและกำลังคนในการระบุข้อมูล ลาเบลน้อย อีกทั้งสำหรับบางงาน อาจต้องการผู้เชี่ยวชาญเฉพาะทางในการระบุข้อมูลลาเบลอีกด้วย ทำให้การเก็บข้อมูลลาเบลที่ถูกต้องและมีปริมาณมากพอสำหรับการสร้างโมเดลการเรียนรู้เชิงลึกเป็นเรื่องยาก

ในงานวิจัยนี้ เน้นที่ปัญหาข้อมูลลาเบลมีไม่เพียงพอ ในการสร้างหรือเทรนโมเดลการเรียนรู้ของเครื่องให้มีประสิทธิภาพได้ สำหรับโจทย์การจำแนกภาพถ่ายดาวเทียม ในการทดลองจะอิงจากสถานการณ์ปัจจุบัน ที่มีปริมาณข้อมูลภาพถ่ายดาวเทียมพร้อมข้อมูลลาเบลน้อย และมีข้อมูลภาพถ่ายดาวเทียมที่ไม่มีข้อมูลลาเบลจำนวนมาก วิธีการนี้ใช้ประโยชน์จากข้อมูลที่ไม่มีลาเบลด้วยวิธีการเรียนรู้แทนภาพด้วยตนเอง (self-supervised learning) กระบวนการของเราเริ่มต้นด้วยสร้างโมเดลแทนภาพด้วยเวกเตอร์ (embedding vector) ด้วยเทคนิค SimCLR (Chen, Kornblith, Norouzi, & Hinton, 2020) แล้วจึงสร้างโมเดลการเรียนรู้ของเครื่อง เช่น support vector machine (SVM) (Cortes & Vapnik, 1995) ในการจำแนกภาพถ่ายดาวเทียมที่มีข้อมูลลาเบล โดยใช้เวกเตอร์แทนภาพเป็นข้อมูลนำเข้าแทน จากนั้นเราใช้โมเดลที่ได้นี้ทำนายข้อมูลลาเบลให้กับข้อมูลภาพถ่ายดาวเทียมที่ไม่มีข้อมูลลาเบล และเรียกผลลัพธ์ที่ได้นี้ว่า “ข้อมูลลาเบลเทียม” สุดท้ายจึงใช้ชุดข้อมูลนำเข้าพร้อมข้อมูลลาเบลเทียมรวมทั้งข้อมูลนำเข้าพร้อมข้อมูลลาเบลจริงในการเทรนโมเดลการเรียนรู้ของเครื่องที่ต้องการ

งานวิจัยนี้นำเสนอกระบวนการอัตโนมัติในการสร้างข้อมูลลาเบลเทียมและแสดงให้เห็นว่าโมเดลสำหรับโจทย์การจำแนกที่เทรนด้วยป้ายชื่อเทียมให้ผลลัพธ์ที่ดีใกล้เคียงกับโมเดลเทรนด้วยข้อมูลลาเบลจริง จากการทดลองพบว่าวิธีการของเราใช้ข้อมูลที่มีลาเบลน้อยลงถึง 9 เท่าในขณะที่ผลลัพธ์นั้นมีความแม่นยำใกล้เคียงกัน อีกทั้งยังแสดงให้เห็นว่าการคำนวณเวกเตอร์แทนภาพด้วยวิธีการเรียนรู้แทนภาพด้วยตนเองจากข้อมูลที่ไม่มีลาเบลสามารถสร้างเวกเตอร์ที่บ่งบอกคุณสมบัติทั่วไปและมีความยืดหยุ่นสามารถใช้ในการแก้ปัญหาเหมือนกันในโดเมนเดียวกันได้

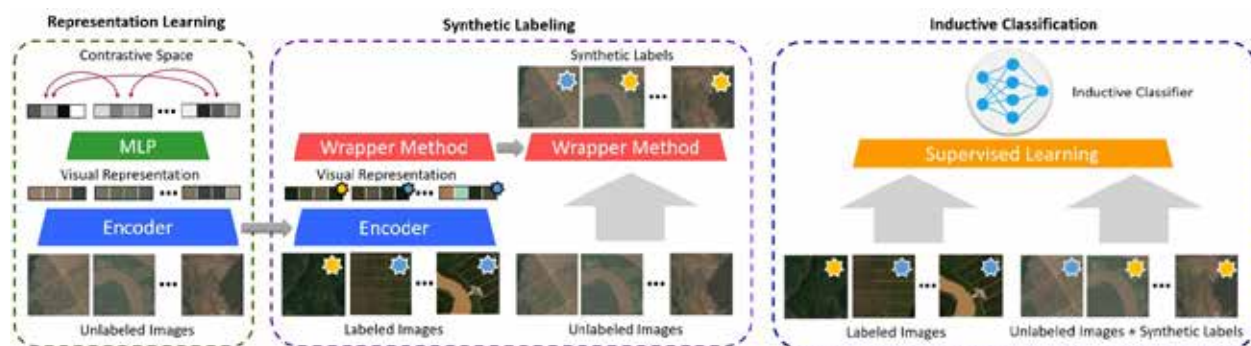


Figure 1 The overview of our method that comprises three steps: representation learning, synthetic-labeling, and inductive classification. (Left) the representation learning step learns the visual representation encoder.

(Middle) the synthetic labeling utilizes the previously learned encoder to generate the synthetic labels for unlabeled data. (Right) the inductive classification trains a supervised model on both the training set and the unlabeled set with their corresponding synthetic labels.

วิธีการศึกษา

กำหนดให้ $\mathcal{D} = \{x_i\}$ สำหรับ $i = 1, \dots, N$ แทนเซตของภาพถ่ายดาวเทียมที่ไม่มีข้อมูลลาเบล (x_i) และ $\mathcal{D} = \{(x_j, y_j)\}$ สำหรับ $j=1, \dots, M$ แทนเซตของภาพถ่ายดาวเทียม (x_j) ที่มีข้อมูลลาเบล (y_j) โดยที่ $N \gg M$

วิธีการของเรามีสามขั้นตอน คือ 1) การเรียนรู้การแทนภาพ (representation learning) 2) การสร้างข้อมูลลาเบลเทียม (synthetic-labeling) และ 3) การจำแนกแบบอุปนัย (inductive classification) (Figure 1)

1. การเรียนรู้การแทนภาพ

วัตถุประสงค์ของขั้นตอนแรก คือ การหาเวกเตอร์แทนภาพ ซึ่งทำได้หลายวิธี แต่วิธีที่ได้รับความนิยม คือ วิธีการเรียนรู้แทนภาพด้วยตนเอง (self-supervised learning) ซึ่งเป็นการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) แต่ทำเสมือนเป็นการเรียนรู้แบบมีผู้สอน ความแตกต่าง คือ ข้อมูลลาเบลจะถูกสร้างโดยอัตโนมัติจากข้อมูลนำเข้า ซึ่งอาจไม่เกี่ยวข้องกับโจทย์ปัญหาที่สนใจได้ วิธีการเรียนรู้แทนภาพด้วยตนเองประกอบด้วยขั้นตอนหลัก คือ pretext tasks และ downstream tasks (Jing & Tian, 2020)

ขั้นตอน pretext tasks หมายถึง โจทย์การเรียนรู้แบบมีผู้สอนที่ถูกออกแบบเพื่อช่วยในการสร้างเวกเตอร์แทนภาพที่มีความยืดหยุ่นและสามารถถ่ายทอดความรู้ไปใช้ในการแก้ปัญหาอื่นที่คล้ายกันได้ ด้วยการทำนายข้อมูลลาเบลที่สร้างจากข้อมูลนำเข้า ในขณะที่ downstream tasks หมายถึง โจทย์ที่ใช้ประโยชน์จากเวกเตอร์แทนภาพที่ได้จากขั้นตอนแรก โดยการสร้างโมเดลที่ใช้เวกเตอร์แทนภาพเป็นข้อมูลนำเข้าบนชุดข้อมูลที่มีข้อมูลลาเบล โดยเราวัดคุณภาพของเวกเตอร์แทนภาพจากประสิทธิภาพของโมเดลในโจทย์ downstream tasks

เทคนิควิธีการเรียนรู้แทนภาพด้วยตนเองหนึ่งที่น่าสนใจ คือ การเรียนรู้แบบเปรียบเทียบ (contrastive learning) นำเสนอโดย Hadsell *et al.* (2006) ซึ่งเรียนรู้เวกเตอร์แทนภาพโดยการเปรียบเทียบตัวอย่างกับข้อมูลที่มีลาเบลเหมือนกันและต่างกัน วิธีการประเภทนี้ได้รับความนิยม ชื่อ SimCLR (Chen, Kornblith, Norouzi, & Hinton, 2020; Chen, Kornblith, Swersky, *et al.*, 2020) ซึ่งนำเสนอเฟรมเวิร์คง่าย ๆ สำหรับการเรียนรู้แบบเปรียบเทียบ ด้วยการหาโมเดลที่ให้เวกเตอร์แทนภาพที่สอดคล้องกันของข้อมูลนำเข้าเดียวกัน แต่ถูกดัดแปลงต่างกัน

เราได้นำวิธีการ SimCLR มาใช้สร้างเวกเตอร์แทนภาพจากทั้งข้อมูลที่ไม่มีลาเบล $\mathcal{D}$ และข้อมูลที่มีลาเบล $\mathcal{D}$ เวกเตอร์แทนภาพ h_i ถูกสร้างโดยใช้โมเดล encoder f กับข้อมูลนำเข้าที่ถูกดัดแปลง $t(x_i)$:

$$h_i = f(t(x_i), \theta)$$

โดย θ คือพารามิเตอร์ของ f และ t คือการดัดแปลงข้อมูลแบบสุ่ม

แล้วเวกเตอร์แทนภาพจึงถูกแปลงเป็นเวกเตอร์ z_i สำหรับการเรียนรู้แบบเปรียบเทียบ ด้วยโมเดล multi-layered perceptron (MLP) g ที่มีพารามิเตอร์ W :

$$z_i = g(h_i, W)$$

โดยฟังก์ชันเป้าหมาย ใช้ NT-Xent loss (Chen, Kornblith, Norouzi, & Hinton, 2020), ซึ่งจะหาโมเดลที่ให้เวกเตอร์ z_i ของข้อมูลนำเข้าเดียวกันแต่ถูกดัดแปลงต่างกันให้สอดคล้องกันมากที่สุด โดยใช้สูตรต่อไปนี้

$$\theta^* = \arg \min_{\theta, W} \mathcal{L}$$

$$\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [l(2k-1, 2k) + l(2k, 2k-1)]$$

$$l(i, j) = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}$$

โดย $\text{sim}(\mathbf{z}_i, \mathbf{z}_j)$ คือ ค่า cosine similarity ระหว่างเวกเตอร์ $\mathbf{z}_i$ และ $\mathbf{z}_j$ ส่วน $\mathbf{1}_{[k \neq i]} \in \{0, 1\}$ เป็นฟังก์ชันที่มีค่าเป็น 1 ถ้า $k \neq i$ มิเช่นนั้นเป็น 0 ฟังก์ชันเป้าหมายถูกคำนวณเฉพาะคู่ของเวกเตอร์แทนภาพของข้อมูลในชุดข้อมูลที่มีลาเบลเหมือนกันเท่านั้นทั้ง $(\mathbf{z}_i, \mathbf{z}_j)$ และ $(\mathbf{z}_j, \mathbf{z}_i)$ (Chen, Kornblith, Norouzi, & Hinton, 2020)

2. การสร้างข้อมูลลาเบลเทียม

วัตถุประสงค์ของขั้นตอนนี้คือการสร้างข้อมูลลาเบลเทียมของข้อมูลที่ไม่มีลาเบล $\mathcal{D}$ ต่อไป โดยเราได้เลือกใช้เทคนิค Wrapper ซึ่งเป็นหนึ่งในวิธีการประเภทการเรียนรู้แบบกึ่งมีผู้สอน (semi supervised learning) การเรียนรู้ประเภทนี้ใช้ประโยชน์จากข้อมูลที่ไม่มีลาเบลซึ่งมักเก็บข้อมูลได้ง่ายและมีจำนวนมาก ร่วมกับข้อมูลที่มีลาเบลที่มักต้องใช้ทรัพยากรมากในการเก็บข้อมูลและมีจำนวนน้อยเพื่อสร้างโมเดลที่ดีขึ้น (Zhu, 2005; Chapelle *et al.*, 2006; van Engelen & Hoos, 2020)

เทคนิคการเรียนรู้แบบกึ่งมีผู้สอนสามารถแบ่งได้เป็นสองประเภทตามวัตถุประสงค์ คือ วิธีแบบ Inductive กับ วิธีแบบ Transductive โดยวิธีแบบ Inductive จะสร้างฟังก์ชันที่สามารถทำนายลาเบลของข้อมูลทั้งหมดที่เป็นไปได้ ในขณะที่วิธีแบบ Transductive จะคำนวณลาเบลของข้อมูลที่ไม่มีลาเบลโดยไม่ใช้ฟังก์ชันทำนาย (van Engelen & Hoos, 2020)

โดยเราเลือกใช้เทคนิค Wrapper ซึ่งเป็นเทคนิคหนึ่งในวิธีแบบ Inductive ที่ง่ายและเป็นที่ยอมรับ ซึ่งประกอบด้วย

2 ขั้นตอน คือ การเทรนบนข้อมูลที่มีลาเบล และการสร้างลาเบลเทียมให้ข้อมูลที่ไม่มีลาเบล ขั้นแรกโมเดล Wrapper F ถูกเทรนบนเวกเตอร์แทนภาพ $\{(h_j, y_j)\}$ ของข้อมูลที่มีลาเบล D จากนั้นจึงใช้โมเดล Wrapper F เพื่อสร้างข้อมูลลาเบลเทียม y_i ให้ข้อมูลที่ไม่มีลาเบล D

$$y_i = F(x_i)$$

3. การจำแนกแบบอุปนัย

ขั้นตอนสุดท้ายนี้มีวัตถุประสงค์เพื่อสร้างโมเดลจำแนกทั้งจากข้อมูลที่มีลาเบล D และข้อมูลที่ไม่มีลาเบลกับข้อมูลลาเบลเทียม $\{(x_i, y_i)\}$ เป็นชุดข้อมูลสอน โดยโมเดลจำแนกนี้สามารถใช้เทคนิคใดก็ได้ เพราะ ชุดข้อมูลนี้เสมือนเป็นชุดข้อมูลที่มีลาเบลชุดหนึ่ง และไม่จำเป็นต้องใช้วิธีเดียวกับโมเดล Wrapper นี้คือคุณสมบัติที่มีประโยชน์ตามที่อธิบาย เรื่อง การกลั่นกรองความรู้ (Knowledge Distillation) (Hinton *et al.*, 2015) ที่เราสามารถสร้างโมเดลหนึ่งโดยใช้ความรู้จากอีกโมเดลได้

การใช้วิธีการเรียนรู้แทนภาพด้วยตนเองอย่าง SimCLR ในการคำนวณเวกเตอร์แทนภาพ ช่วยให้การสร้างโมเดล Wrapper ทำได้มีประสิทธิภาพมากขึ้น เนื่องจากเดิมจะต้องสร้างโมเดล Wrapper ซ้ำหลายรอบบนชุดข้อมูลที่มีลาเบลจริง และชุดข้อมูลที่มีลาเบลเทียม จนกว่าข้อมูลลาเบลเทียมที่ได้จะไม่มีการเปลี่ยนแปลง แต่พอสร้างโมเดล Wrapper บนเวกเตอร์แทนภาพก็ไม่ต้องมีความจำเป็นที่จะต้องทำซ้ำ อีกทั้งผลลัพธ์ที่ได้ก็ดีกว่าโมเดลที่ใช้ข้อมูลนำเข้าโดยตรง

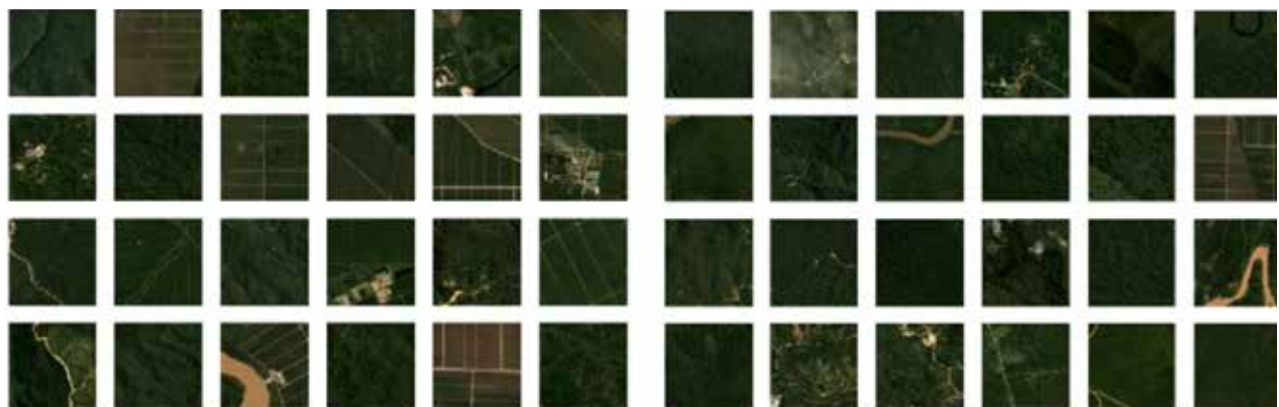


Figure 2 Palm oil plantation dataset. (Left) the images that contain palm oil plantation(s). (Right) the images that don't contain palm oil plantation.

ชุดข้อมูลป่าแอมะซอนเป็นภาพถ่ายดาวเทียมแบบ 4 แบนด์ที่ความละเอียด 3 เมตรของกลุ่มน้ำแอมะซอน ข้อมูลเหล่านี้ถูกเก็บรวบรวมโดยดาวเทียม Flock 2 ของ Planet ระหว่างวันที่ 1 มกราคม 2016 ถึงวันที่ 1 กุมภาพันธ์ 2017 ชุดข้อมูลนี้ประกอบด้วยภาพมากกว่า 40,000 ภาพพร้อมกับข้อมูลลาเบลที่ระบุสภาพอากาศและประเภทการใช้ที่ดิน ใน

การทดลอง

1. ข้อมูลที่ใช้ในการทดลอง

เราทำการทดสอบวิธีการของเราโดยใช้ชุดข้อมูลภาพถ่ายดาวเทียมสองชุด คือ ชุดข้อมูลภาพป่าแอมะซอน และชุดข้อมูลภาพไร่ปาล์มน้ำมัน ทั้งสองชุดข้อมูลนี้ประกอบด้วยรูปประเภท RGB ของภาพถ่ายดาวเทียมขนาด 256x256 พิกเซล แม้ว่าในความเป็นจริงข้อมูลในชุดข้อมูลเหล่านี้จะมีลาเบลทั้งหมด แต่เราจำลองสถานการณ์ที่ข้อมูลที่มีลาเบลมีน้อยเหมือนกับโจทย์ในงานวิจัยของ Kussul *et al.* (2017) กับ Heupel *et al.* (2018) โดยการสุ่มบางส่วนของชุดข้อมูลและลบข้อมูลลาเบลออกเพื่อสร้างชุดข้อมูลที่ไม่มีลาเบล การสุ่มลบข้อมูลลาเบลถูกทำเพื่อหลีกเลี่ยงปัญหาความซ้ำซ้อนในการเก็บข้อมูลและเตรียมข้อมูล ส่วนรายละเอียดของแต่ละชุดข้อมูลมีดังนี้

ชุดข้อมูลไร่ปาล์มน้ำมันเป็นรูปภาพ RGB ความละเอียด 3 เมตรที่รวบรวมโดยดาวเทียมของ Planet ข้อมูลลาเบลบ่งบอกว่ามีไร่ปาล์มน้ำมันปรากฏในภาพหรือไม่ ข้อมูลลาเบลนี้มาจากการระบุลาเบลโดยฝูงชน (crowdsourced annotation) โดยชุดข้อมูลนี้ประกอบด้วยภาพมากกว่า 15,000 ภาพ แต่เราสุ่มเพียง 1,885 รูปภาพเป็นข้อมูลที่มีลาเบล โดยมี 918 ภาพที่มีไร่ปาล์มน้ำมันปรากฏ (Liu *et al.*, 2019) (Figure 2)

งานนี้เราสุ่มเพียง 21,118 รูปภาพเป็นข้อมูลที่มีลาเบล และใช้ภาพ RGB แทนที่จะใช้ภาพแบบ 4 แบนด์ อีกทั้งลาเบลที่ใช้จะระบุแค่ว่ามีการเกษตรปรากฏในภาพหรือไม่ เพื่อให้ได้ปัญหาการจำแนกแบบทวิภาค โดยมีข้อมูลจำนวน 9,852 ภาพที่มีลาเบลการเกษตรปรากฏ (Goldenberg *et al.*, 2017) (Figure 3)

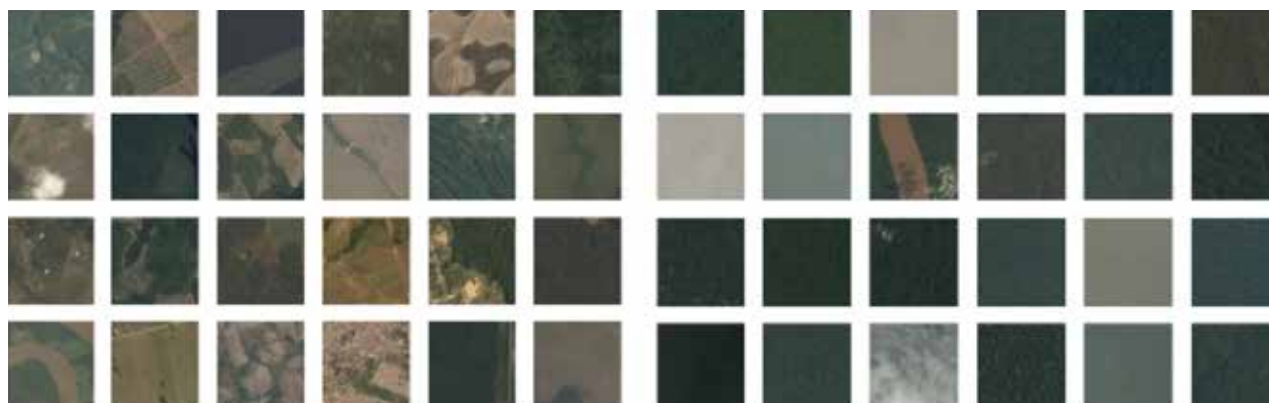


Figure 3 Amazon rainforest dataset. (Left) the images with agriculture label. (Right) the images without agriculture label.

2. วิธีการทดลอง

ทุกการทดลองประมวลผลบนคอมพิวเตอร์ที่ใช้ซีพียู AMD Ryzen 7 2700X 3.7GHz การ์ดจอ GeForce Nvidia RTX 2070 และ RAM 16 GB การเทรนโมเดลทั้งสามขั้นตอนใช้เวลา 2-3 ชั่วโมง

การทดลองของเรามีเป้าหมายสองอย่าง คือ 1) ประเมินประสิทธิภาพของโมเดลการจำแนกที่เทรนบนข้อมูลลาเบลเทียม และ 2) ประเมินคุณภาพและความยืดหยุ่นในการใช้ประโยชน์ของเวกเตอร์แทนภาพ

สำหรับการประเมินประสิทธิภาพของโมเดลการจำแนกที่เทรนบนข้อมูลลาเบลเทียม ส่วนที่หนึ่ง เราแสดงให้เห็นว่าข้อมูลลาเบลเทียมสามารถใช้เทรนโมเดลแทนข้อมูลลาเบลจริงได้ โดยเฉพาะในชุดข้อมูลขนาดเล็ก เราตรวจสอบโดยการเทรนโมเดล Convolutional Neural Network (CNN) ที่เป็น pretrained model อย่าง VGG16 (Simonyan & Zisserman, 2014) นำมาเทรนบนทั้งข้อมูลลาเบลจริงและลาเบลเทียมเพื่อเปรียบเทียบผลลัพธ์กัน โดยโมเดลอื่นที่ซับซ้อนมากกว่านี้ก็สามารถใช้ได้เช่นกัน ในการทดลองนี้ลาเบลเทียมถูกสร้างขึ้นโดยใช้เทคนิค SVM เทรนบนเวกเตอร์แทนภาพที่คำนวณโดยเทคนิค SimCLR ดังนั้นเราจึงเปรียบเทียบกับเทคนิค SVM นี้ด้วยบนทั้งข้อมูลลาเบลจริงและลาเบลเทียม เทียบผลลัพธ์กับวิธีการของเรา ในส่วนที่สอง เราเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง 3 เทคนิค เพื่อหาว่าเทคนิคใดใช้เป็นโมเดล Wrapper สร้างข้อมูลลาเบลเทียมได้ดีที่สุด โดยพิจารณาจากประสิทธิภาพของโมเดล VGG16 ที่เทรนบนข้อมูลลาเบลเทียม

สำหรับการประเมินคุณภาพและความยืดหยุ่นในการใช้ประโยชน์ของเวกเตอร์แทนภาพ ในส่วนแรก เราประเมินคุณภาพของเวกเตอร์แทนภาพด้วยการแสดงผลเวกเตอร์แทนภาพโดยใช้เทคนิคอย่าง t-SNE (van der Maaten & Hinton, 2008) ซึ่งแสดงให้เห็นว่า เวกเตอร์แทนภาพที่คำนวณด้วยเทคนิค SimCLR ช่วยให้การแยกแยะระหว่างภาพต่างคลาสนั้น

มีความชัดเจนมากขึ้น ในขณะที่ส่วนที่สอง เราแสดงให้เห็นว่าเวกเตอร์แทนภาพเหล่านี้มีความยืดหยุ่นในการใช้ประโยชน์ และสามารถถ่ายทอดความรู้ไปใช้แก้ปัญหาอื่นที่คล้ายกันได้ โดยการนำโมเดล encoder ที่เทรนบนชุดข้อมูลป่าแอมะซอนมาใช้กับภาพในชุดข้อมูลไร่ปาล์มน้ำมันในการคำนวณเวกเตอร์แทนภาพ จากนั้นจึงใช้เวกเตอร์แทนภาพนี้สร้างลาเบลเทียม แล้วจึงนำลาเบลเทียมไปเทรนโมเดลจำแนกด้วยวิธีการของเราต่อไป สุดท้ายจึงนำโมเดลนี้เทียบกับโมเดลอย่าง VGG16 ที่เทรนบนข้อมูลลาเบลจริงเท่านั้น

ผลการทดลองและอภิปรายผล

1. ประสิทธิภาพในการจำแนกด้วยลาเบลเทียม

เราเปรียบเทียบวิธีการของเรา (Our method) กับ 1) โมเดล VGG16 (VGG16) และ 2) โมเดล SVM ที่เทรนบนเวกเตอร์แทนภาพคำนวณด้วยเทคนิค SimCLR (SimCLR+SVM) โดยโมเดลทั้งสองนี้ใช้เฉพาะข้อมูลลาเบลจริงเท่านั้น ส่วนโมเดล VGG16 ใช้เทคนิคการถ่ายทอดความรู้ (transfer learning) (Giorgiani do Nascimento & Viana, 2020) ด้วยการใช้ pretrained model ที่ได้จากการเทรนบนชุดข้อมูลขนาดใหญ่ อย่าง ImageNet (Russakovsky *et al.*, 2015) ส่วนโมเดล SVM ใช้ kernel เป็น radial basis function (RBF) ส่วนโมเดล SimCLR ใช้โมเดล ResNet18 (He *et al.*, 2016) เป็นโมเดล encoder

ในการประเมินจะใช้ค่าความแม่นยำ (Accuracy; Acc) ค่า precision (P) ค่า recall (R) และค่า F1-score (F1) เป็นเกณฑ์ ซึ่งมีสูตรดังนี้

$$Acc = (tp + tn) / (tp + tn + fp + fn)$$

$$P = tp / (tp + fp)$$

$$R = tp / (tp + fn)$$

$$F1 = 2PR / (P + R) = 2tp / (2tp + fp + fn)$$

โดย
 tp คือ จำนวนผลบวกจริง tn คือ จำนวนผลลบจริง
 fp คือ จำนวนผลบวกปลอม fn คือ จำนวนผลลบปลอม

ขั้นตอนการเตรียมภาพที่ใช้ในการทดลองมีเพียงการลดขนาดภาพเป็น 128×128 พิกเซล แม้ว่ากระบวนการก่อนการประมวลผลภาพจะมีผลกระทบมากต่อประสิทธิภาพ (D.R. *et al.*, 2017) แต่เรื่องนี้ไม่ใช่สิ่งที่เราต้องการศึกษา

เราจึงเลือกใช้วิธีที่ง่ายที่สุด เพื่อหลีกเลี่ยงปัญหาข้อมูลที่ไม่สมดุล เราจึงสุ่มและแบ่งชุดข้อมูลเพื่อให้ทั้งชุดข้อมูลสอนและทดสอบมีจำนวนลาเบลในแต่ละคลาสที่ใกล้เคียงกัน โดยในชุดข้อมูลไร่ปาล์มน้ำมันมีภาพทดสอบ 193 ภาพ โดยมี 93 ภาพที่มีไร่ปาล์มน้ำมันปรากฏ ส่วนในชุดข้อมูลป่าแอมะซอนมีทั้งหมด 1,056 ภาพ โดยมี 493 ภาพมีลาเบลการเกษตรปรากฏ

Table 1 Comparison between the methods when 187 labeled data from palm oil plantation dataset are available.

Method	Acc.	P	R	F1
Our method	74.9	.798	.643	.712
VGG16	65.2	.708	.647	.636
SimCLR+SVM	74.1	.772	.656	.709

Table 2 Comparison between the methods when 1,055 labeled data from Amazon rainforest dataset are available.

Method	Acc.	P	R	F1
Our method	85.9	.897	.831	.863
VGG16	84.6	.851	.862	.856
SimCLR+SVM	84.4	.819	.854	.836

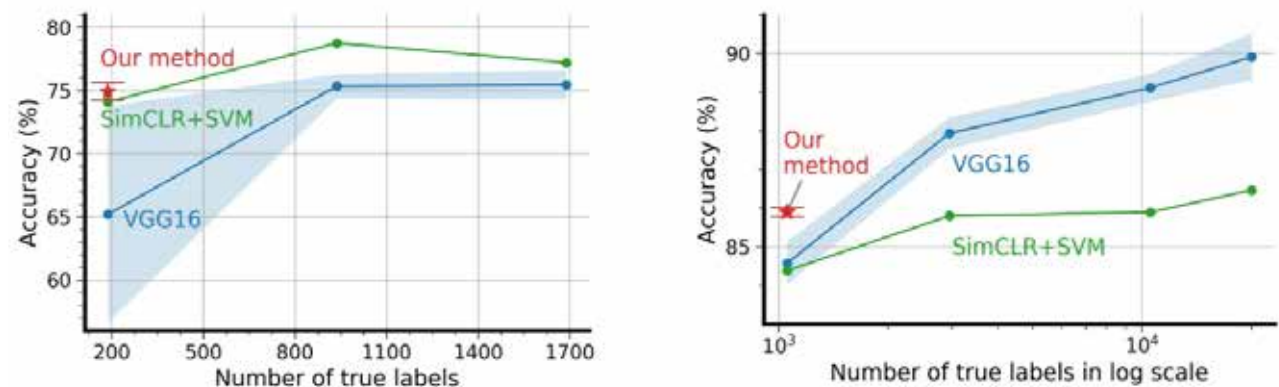


Figure 4 Accuracy comparison of different methods: our method (red), pre-trained VGG16 model (blue), and SVM over SimCLR features (green). The error bar and highlighted area show the range of one standard deviation of the mean from running the algorithm 5 times. The X-axis shows the number labeled data available to the algorithm. (Left) the result from Palm oil plantation dataset (Right) the result from Amazon rainforest dataset.

Table 1 และ 2 แสดงให้เห็นว่าวิธีการของเราให้ผลลัพธ์ที่มีความแม่นยำสูงกว่าเมื่อเปรียบเทียบกับวิธีอื่น ๆ ในกรณีจำนวนข้อมูลลาเบลมีน้อย วิธีการของเรามีความแม่นยำสูงกว่าเกือบ 10% เมื่อเทียบกับโมเดล VGG16 ในชุดข้อมูลไร่ปาล์มน้ำมัน Figure 4 (Left) ยังแสดงว่าวิธีการของเราให้ผลลัพธ์ที่มีความแม่นยำใกล้เคียงกับโมเดล VGG16 เมื่อเทรน

บนชุดข้อมูลที่มีข้อมูลลาเบลมากกว่าถึง 9 เท่า เนื่องจากขั้นตอนบางขั้นตอนของอัลกอริทึม CNN มีขั้นตอนการสุ่ม เราจึงรายงานผลเฉลี่ยจากการทดลอง 5 รอบ ปรากฏการณ์นี้มีผลลดลงเมื่อจำนวนข้อมูลลาเบลมีสูงขึ้น ตามที่แสดงใน Figure 4 (Right) แต่ในโดเมนที่เฉพาะเจาะจง การเก็บข้อมูลลาเบลปริมาณมากเป็นเรื่องท้าทายมาก และกรณีการศึกษา

ก็เป็นกรณีที่ข้อมูลลาเบลมีปริมาณน้อย มิเช่นนั้นเราก็สามารถสร้างโมเดลจำแนกด้วยวิธีการทั่วไปได้ ถ้าเรามีปริมาณข้อมูลมาก

ในส่วนที่สอง เราเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง 3 เทคนิค คือ 1) SVM 2) k-nearest neighbor (kNN) และ 3) logistic regression (Logistic) เพื่อหาว่าเทคนิคใดเมื่อใช้เป็นโมเดล Wrapper สร้างข้อมูลลาเบลเทียมแล้ว จะช่วยสร้างโมเดลจำแนกที่มีประสิทธิภาพสูงสุด โดยใช้โมเดล

VGG16 ที่เทรนบนข้อมูลลาเบลเทียมเป็นตัวเปรียบเทียบ อีกทั้งยังเทียบกับการใช้โมเดล Wrapper จำแนกผลลัพธ์โดยตรงเลยด้วย ผลลัพธ์ใน Table 3 และ 4 แสดงให้เห็นว่า VGG16 ที่เทรนบนข้อมูลลาเบลเทียมให้ความแม่นยำสูงกว่า นอกจากนี้ยังสังเกตเห็นว่าโมเดลไม่เชิงเส้น เช่น SVM ที่มี kernel เป็น RBF และ kNN มักให้ผลลัพธ์ที่ดีกว่าโมเดลเชิงเส้น เช่น logistic regression

Table 3 Comparison between different wrapper methods and the VGG16 model trained on synthetic labels. 187 labeled data from palm oil plantation dataset are available.

Method	Acc.	VGG Acc.
SVM	74.1	74.9 $\pm$ 0.7
kNN	70.5	76.2 $\pm$ 0.9
Logistic	57.5	71.2 $\pm$ 2.0

Table 4 Comparison between different wrapper methods and the VGG16 model trained on synthetic labels. 1,055 labeled data from Amazon rainforest dataset are available.

Method	Acc.	VGG Acc.
SVM	84.4	85.2 $\pm$ 0.2
kNN	83.0	84.9 $\pm$ 0.5
Logistic	79.2	83.7 $\pm$ 0.5

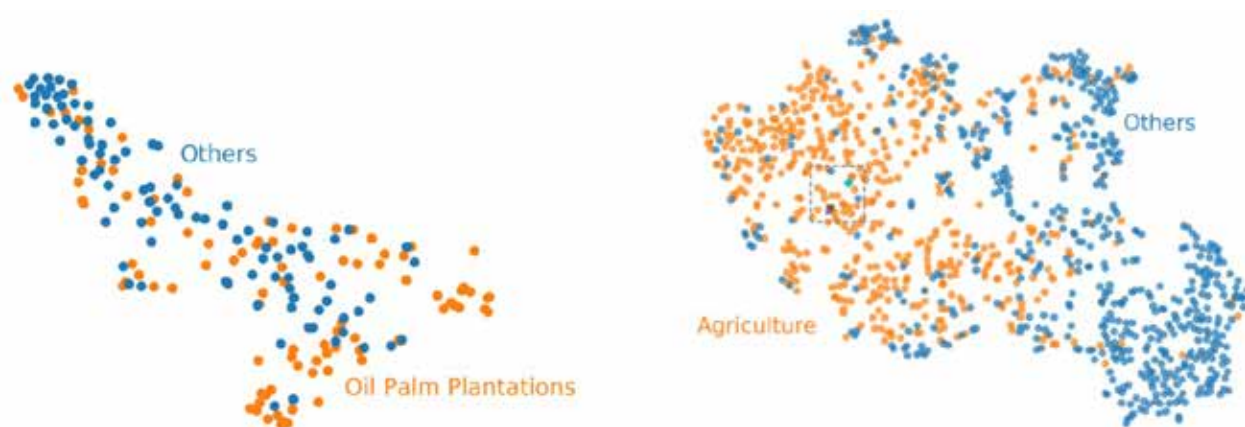


Figure 5 T-SNE visualization of visual representation vectors. The encoder function is optimized via self-supervised learning. (Left) Oil palm plantation training set. (Right) Amazon rainforest training set.

2. คุณภาพและความยืดหยุ่นของเวกเตอร์แทนภาพ

ในส่วนแรก เราประเมินคุณภาพของเวกเตอร์แทนภาพด้วยการแสดงผลเวกเตอร์แทนภาพโดยใช้เทคนิค t-SNE (van der Maaten & Hinton, 2008) ใน Figure 5 ในภาพทางซ้าย ซึ่งแสดงชุดข้อมูลไร่ปาล์มน้ำมัน เนื่องจากภาพในชุดข้อมูลนี้มีความแตกต่างกันเล็กน้อยระหว่างสองคลาสดังใน Figure 2

กราฟ t-SNE จึงแยกแยะจุดข้อมูลของภาพที่อยู่ต่างคลาสได้ยาก แต่ในขณะเดียวกัน ในภาพทางขวาที่แสดงชุดข้อมูลป่าแอมะซอน ภาพในชุดข้อมูลนี้มีความแตกต่างที่สังเกตเห็นได้ระหว่างสองคลาสดังใน Figure 3 กราฟ t-SNE จึงแยกแยะจุดข้อมูลของภาพที่อยู่ต่างคลาสได้ง่ายขึ้นนี้เนื่องจากการคำนวณเวกเตอร์แทนภาพและสร้างกราฟ t-SNE ไม่ได้

ใช้ข้อมูลลาเบล จึงเห็นได้ว่าเวกเตอร์แทนภาพนี้ช่วยให้การแยกแยะระหว่างภาพต่างคลาสนั้นทำได้ง่ายขึ้น

ในขณะที่ส่วนที่สอง เราแสดงให้เห็นว่าเวกเตอร์แทนภาพเหล่านี้มีความยืดหยุ่นและสามารถถ่ายทอดความรู้ไปใช้กับปัญหาอื่นที่คล้ายกันได้ โดยการนำโมเดล encoder ที่เทรนบนชุดข้อมูลป่าแอมะซอนมาใช้กับภาพในชุดข้อมูลไร่ปาล์มน้ำมันในการคำนวณเวกเตอร์แทนภาพ จากนั้นจึงเทรนโมเดล SVM ที่ใช้เวกเตอร์แทนภาพนี้สร้างลาเบลเทียมให้ข้อมูลไร่ปาล์มน้ำมันที่ไม่มีลาเบล สุดท้ายจึงนำลาเบลเทียมไปเทรนโมเดล VGG16 ตามวิธีการของเรา

ใน Table 5 เราได้เปรียบเทียบโมเดล VGG16 ที่เทรนบนข้อมูลลาเบลจริง (VGG16) กับโมเดล Wrapper ที่เป็นเทคนิค SVM เทรนบนเวกเตอร์แทนภาพที่คำนวณด้วยโมเดล encoder ของชุดข้อมูลป่าแอมะซอน (Transferred

SVM) และโมเดล VGG16 ที่เทรนบนข้อมูลลาเบลเทียมที่ทำนายด้วยโมเดล Transferred SVM (Transferred VGG16) ผลลัพธ์แสดงให้เห็นว่าการใช้ข้อมูลเวกเตอร์แทนภาพทั้งในรูปแบบโมเดล Wrapper (Transferred SVM) หรือใช้สร้างโมเดลจำแนกจากข้อมูลลาเบลเทียม (Transferred VGG16) ให้โมเดลที่มีประสิทธิภาพดีกว่าโมเดลที่ถูกเทรนบนลาเบลจริง ผลการทดลองนี้แสดงให้เห็นว่าเวกเตอร์แทนภาพที่คำนวณโดยใช้เทคนิคการเรียนรู้แทนภาพด้วยตนเองมีคุณสมบัติหลายอย่างที่น่าประหลาดใจ ตั้งแต่การเข้ารหัสลักษณะของวัตถุในภาพซึ่งสามารถถ่ายทอดองค์ความรู้ไปใช้ในโจทย์ที่คล้ายกันได้ วิธีของเราใช้ประโยชน์จากหลักการนี้ร่วมกับวิธีการเรียนรู้แบบกึ่งมีผู้สอน เพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพสูงบนชุดข้อมูลลาเบลขนาดเล็กได้

Table 5 Comparison between the models with transfer learning when 187 labeled data from palm oil plantation dataset are available.

Method	Acc.	P	R	F1
VGG16	65.2	.708	.647	.636
Transferred SVM	72.5	.692	.774	.731
Transferred VGG16	73.2	.716	.733	.725

สรุปผลการทดลองและข้อเสนอแนะ

ในงานวิจัยนี้ เราได้นำเสนอกระบวนการสร้างข้อมูลลาเบลเทียมที่ช่วยให้สามารถสร้างโมเดลการเรียนรู้แบบมีผู้สอนที่มีประสิทธิภาพดี แม้ว่าจะมีปริมาณข้อมูลลาเบลน้อยและไม่เพียงพอต่อการสร้างโมเดลที่มีประสิทธิภาพได้ วิธีของเราใช้เทคนิคการเรียนรู้แทนภาพด้วยตนเองร่วมกับเฟรมเวิร์กการเรียนรู้แบบกึ่งมีผู้สอน เพื่อลดการพึ่งพาข้อมูลลาเบลปริมาณมากได้ การสร้างเวกเตอร์แทนภาพช่วยให้โมเดลสามารถเรียนรู้ลักษณะของวัตถุในภาพโดยไม่ต้องมีข้อมูลลาเบล ในขณะที่การเรียนรู้แบบกึ่งมีผู้สอนอย่างเทคนิค Wrapper จะเลือกใช้ข้อมูลบางส่วนของเวกเตอร์แทนภาพมาคำนวณเพื่อให้ได้โมเดลที่เหมาะสมกับโจทย์เป้าหมายที่สนใจ จึงเป็นการแสดงให้เห็นด้วยว่าวิธีการเรียนรู้แทนภาพด้วยตนเองสามารถสร้างเวกเตอร์ที่แทนภาพที่สามารถถ่ายทอดองค์ความรู้จากโจทย์หนึ่งไปยังโจทย์อื่นที่ใกล้เคียงได้ เราได้ทำการประเมินวิธีการของเราด้วยโจทย์การจำแนกภาพถ่ายดาวเทียมบนชุดข้อมูลสองชุด โดยวิธีการของเราให้ความแม่นยำ 75% บนชุดข้อมูลไร่ปาล์มน้ำมัน และ 86% บนชุดข้อมูลป่าแอมะซอน ซึ่งสูงกว่าวิธีการแบบดั้งเดิมอย่างการเรียนรู้แบบมีผู้สอน อีกทั้งวิธีการของเราใช้ข้อมูลลาเบลน้อยกว่าถึง 9 เท่า สองชุดข้อมูลนี้มีปริมาณข้อมูลที่ต่างกันและมีวัตถุประสงค์ต่างกัน แต่งานวิจัยนี้ก็แสดงให้เห็นว่าวิธีการนี้สามารถให้องค์ความรู้จากโจทย์หนึ่งไปยัง

ในอีกโจทย์ได้ผ่านโมเดล encoder ที่ใช้คำนวณเวกเตอร์คุณลักษณะ

จากการศึกษายังพบว่าความแม่นยำของผลลัพธ์ยังสามารถพัฒนาเพิ่มได้อีก โดยในอนาคตหากใช้เทคนิคการเรียนรู้แทนภาพด้วยตนเองที่ซับซ้อนและทันสมัยกว่านี้ ก็จะสามารถสกัดคุณลักษณะสำคัญของภาพออกมาได้ดีกว่าเดิม และให้ลาเบลเทียมและผลลัพธ์ที่ดีขึ้นต่อไป

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักงานกองทุนสนับสนุนการวิจัย สัญญาเลขที่ RDG61Q0020

เอกสารอ้างอิง

Benjamin Goldenberg, Burak Uzkent, Christian Clough, Dennis Funke, Deven Desai, grischka, Jesus MartinezManso, Kat Scott, Meg Risdal, Mike Ryan, Pete, Rachel Holm, Ramesh Nair, Sean Herron, Tony Stafford, Wendy Kan. (2017). *Planet: Understanding the Amazon from Space*. Kaggle. <https://kaggle.com/competitions/planet-understanding-the-amazon-from-space>

- Chapelle, O., Schoellkopf, B., & Zien, A. (2006). *Semi-supervised learning (adaptive computation and machine learning)*. The MIT Press.
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). *A simple framework for contrastive learning of visual representations*. arXiv preprint arXiv:2002.05709.
- Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. (2020). *Big self-supervised models are strong semi-supervised learners*. arXiv preprint arXiv:2006.10029.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3), 273-297. doi: 10.1007/BF00994018
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., ... Raska, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. In *2018 IEEE/CVF conference on computer vision and pattern recognition workshops (cvprw)* (pp. 172-17209). doi: 10.1109/CVPRW.2018.00031
- D.R., S., Shenoy, P. D., & R., V. K. (2017, Mar). Remote sensing satellite image processing techniques for image classification: A comprehensive survey. *IJCA*, 161(11), 24-37. doi: 10.5120/ijca2017913306
- Emily Liu, Judy Logan, kitushan, Meredith Lee, phoebeliuF8, QMirsharif, Sam Roy, Sherrie Wang, Valerie, Vani M, Varsha Venkatesh, WiDS Datathon. (2019). *WiDS Datathon 2019*. Kaggle. <https://kaggle.com/competitions/widsdatathon2019>
- Giorgiani do Nascimento, R., & Viana, F. (2020). Satellite image classification and segmentation with transfer learning. In *Aiaa scitech 2020 forum* (p. 1864). doi: 10.2514/6.2020-1864
- Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *Cvpr* (Vol. 2, pp. 1735-1742). doi: 10.1109/CVPR.2006.100
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on cvpr* (pp. 770-778).
- Heupel, K., Spengler, D., & Itzerott, S. (2018). A progressive crop-type classification using multitemporal remote sensing data and phenological information. *PFG-J Photogramm Rem*, 86(2), 53-69. doi: 10.1007/s41064-018-0050-7
- Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network*. arXiv preprint arXiv:1503.02531.
- Jing, L., & Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE TPAMI*, 1-1. doi: 10.1109/TPAMI.2020.2992393
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci Remote S*, 14(5), 778-782. doi: 10.1109/LGRS.2017.2681128
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *IJCV*, 115(3), 211-252. doi: 10.1007/s11263-015-0816-y
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.
- van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov), 2579-2605.
- van Engelen, J., & Hoos, H. (2020, feb). A survey on semi-supervised learning. *Mach. Learn.*, 109(2), 373-440. doi: 10.1007/s10994-019-05855-6
- Zhu, X. (2005). *Semi-supervised learning literature survey* (Tech. Rep.). University of Wisconsin-Madison Department of Computer Sciences.