

การพัฒนาโมเดลการจำแนกรูปภาพตามความหมายจากภาษาธรรมชาติ
ด้วยปัญญาประดิษฐ์เชิงสร้างสรรค์
Developing Next-Generation Semantic Image Classification Model
Through Generative Adversarial Networks (GANs)

สุขสวัสดิ์ ณัฏฐวุฒิสิทธิ์
Sooksawaddee Nattawuttisit

หลักสูตรวิทยาศาสตรมหาบัณฑิต และ หลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม
Master of Science Program and Doctor of Philosophy Program
Faculty of Information Technology, Sripatum University

Submitted 07/08/2024 ; Revised 21/09/2024 ; Accepted 29/10/2024

บทคัดย่อ

งานวิจัยนี้มุ่งพัฒนาโมเดลการจำแนกรูปภาพด้วยเจเนอเรทีฟ แอดเวอร์ซารี เน็ตเวิร์ค (จีเอเอ็น) เพื่อปรับปรุงการเรียกค้นรูปภาพและแปลความหมายของภาพ โดยใช้ภาษาธรรมชาติ ซึ่งเทคโนโลยีนี้จะสร้างเนื้อหาใหม่โดยเรียนรู้จากข้อมูลที่มีอยู่และสร้างผลลัพธ์ใหม่ที่คล้ายคลึงกับต้นแบบ กลุ่มตัวอย่างที่ใช้ในงานวิจัยประกอบด้วยข้อมูลจากชุดข้อมูลฟริกเกอร์ 30 เค ทั้งหมด 158,915 รายการ ซึ่งเป็นรูปภาพและคำบรรยายภาษาธรรมชาติ โดยสุ่มตัวอย่างด้วยสมการโคราน ที่ค่าความเชื่อมั่น 95% และค่าความคลาดเคลื่อน 5% ได้กลุ่มตัวอย่าง 384 รายการ ข้อมูลถูกแบ่งสำหรับการฝึกและทดสอบโมเดลในอัตราส่วน 80/20 เพื่อเพิ่มประสิทธิภาพในการแปลความหมายของภาพ การประเมินผลของโมเดลวัดจากความคล้ายคลึงระหว่างผลลัพธ์ที่เอไอพยากรณ์และภาพที่มีคำบรรยาย พร้อมตรวจสอบโดยผู้เชี่ยวชาญด้าน ผลทดสอบพบว่าโมเดลมีค่าความแม่นยำ 82% การเรียกค้น 78% และความถูกต้อง 80% ซึ่งบ่งชี้ถึงประสิทธิภาพของโมเดลในการแปลความหมายภาพตามคำบรรยายภาษาธรรมชาติ ผลการวิจัยนี้สามารถประยุกต์ใช้เชิงพาณิชย์ได้ เช่น การจัดหมวดหมู่รูปภาพอัตโนมัติในสื่อสังคม หรือการค้นหารูปภาพในคลังข้อมูลขนาดใหญ่ การพัฒนาโมเดลในอนาคตควรมุ่งเน้นการปรับปรุงค่าการเรียกค้น เพื่อเพิ่มความสมบูรณ์และความตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น

คำสำคัญ: การจำแนกรูปภาพ การเรียนรู้เชิงลึก ภาษาธรรมชาติ จีเอเอ็น

ผู้ประสานงานหลัก (Corresponding Author)

E-mail: sooksawaddee.na@spu.ac.th

Abstract

This research aims to develop an image classification model using Generative Adversarial Networks (GANs) to improve image retrieval and interpretation through natural language processing. This technology generates new content by learning from existing data and producing outputs similar to the original samples. The study's sample data is drawn from the Flickr 30K dataset, consisting of 158,915 entries of images and natural language descriptions. A sample size of 384 entries was determined using Cochran's formula with a 95% confidence level and a 5% margin of error. The data was split into training and testing sets at a ratio of 80/20 to optimize the model's performance in image interpretation. The model's performance was evaluated based on the similarity between the AI-predicted outcomes and the images with descriptions and validated by AI experts. The test results showed an accuracy of 82%, a recall of 78%, and a precision of 80%, indicating the model's effectiveness in interpreting images based on natural language descriptions. This research has commercial applications, such as automatic image categorization on social media or image retrieval in large-scale databases. Future model development should focus on improving recall to enhance completeness and better meet user needs.

Keywords: image classification, deep learning, natural language, GANs

ตารางที่ 1 ตัวอย่างโมเดลที่มีการใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์สร้างข้อความ รูปภาพ หรือสื่ออื่น ๆ

โมเดล	ประเภท	การใช้งาน	จุดเด่น
DreamFusion	GANs	สร้างภาพ 3 มิติจากข้อความ	สร้างโมเดล 3 มิติ
StyleSDF	GANs	การสร้างภาพสไตล์ต่าง ๆ	สร้างภาพที่มีความหลากหลายสไตล์
YOLOv8	CNNs	การจำแนกรูปภาพ	พัฒนาขึ้นจาก YOLOv5
Stable diffusion	VAEs	สร้างภาพที่มีคุณภาพสูง	การปรับปรุงคุณภาพของภาพ
DALL-E	GANs	สร้างภาพจากข้อความ	สร้างภาพจากข้อความ
StyleGAN	GANs	สร้างภาพบุคคลที่สมจริง	สร้างภาพที่มีความละเอียดสูง

วิธีดำเนินการวิจัย

ขั้นตอนหลักของการวิจัยมีดังนี้

1. การออกแบบโมเดล GANs

- แบ่งออกเป็นสองส่วนหลัก ได้แก่ Generator (G) และ Discriminator (D) ซึ่งทำหน้าที่สร้างและแยกแยะข้อมูล
- Generator สร้างข้อมูลใหม่จากการสุ่มค่าและพยายามให้ข้อมูลมีลักษณะคล้ายคลึงกับข้อมูลจริง
- Discriminator แยกแยะว่าข้อมูลใดเป็นของจริงและข้อมูลใดที่สร้างขึ้น

2. การเตรียมข้อมูล

- ใช้ชุดข้อมูล Flickr30k ที่มีภาพและคำอธิบายภาพประกอบ
- คำนวณขนาดกลุ่มตัวอย่างโดยใช้สมการ Cochran เพื่อกำหนดขนาดกลุ่มตัวอย่างที่เหมาะสม

3. การฝึกโมเดล GANs

- แบ่งข้อมูลเป็นชุดฝึกและชุดทดสอบในอัตราส่วน 80/20
- ฝึก Generator และ Discriminator ผ่านการสร้างและจำแนกข้อมูล เพื่อให้ทั้งสองโมเดลทำงานร่วมกันจนเกิดการปรับปรุงที่มีประสิทธิภาพ

4. การวัดความคล้ายคลึงกันของภาพและคำอธิบายภาพด้วย cosine similarity

- ใช้ cosine similarity ในการคำนวณความคล้ายคลึงระหว่างเวกเตอร์ของข้อมูลที่สร้างขึ้นและข้อมูลจริง
- เปรียบเทียบความเหมือนระหว่างคำบรรยายภาพที่สร้างโดย AI กับข้อมูลจริง เพื่อวัดประสิทธิภาพของโมเดล

5. การทดสอบและการวัดประเมินผลโมเดล

- วัดค่าความแม่นยำ (precision), ค่าการเรียกคืน (recall), และค่าความถูกต้อง (accuracy) ของโมเดล
- ประเมินผลการเรียกคืนภาพและการจับคู่คำบรรยายที่สร้างขึ้นกับคำบรรยายในชุดข้อมูล

รายละเอียดของขั้นตอนการทดสอบและประเมินผลโมเดลมีดังนี้

1. การออกแบบโมเดล GANs

ขั้นตอนการสร้างแบบจำลองด้วย GANs เริ่มจาก Generator ที่พยายามสร้างข้อมูลที่คล้ายคลึงกับข้อมูลจริงจากการสุ่มค่า และ Discriminator ทำหน้าที่แยกแยะว่าข้อมูลใดมาจาก Generator และข้อมูลใดเป็นข้อมูลจริง เป้าหมายของการฝึกฝนทั้งสองโมเดลนี้คือ ให้ Generator สามารถสร้างข้อมูลที่ใกล้เคียงกับข้อมูลจริงให้มากที่สุด และให้ Discriminator สามารถแยกแยะได้อย่างแม่นยำ โดย Objective Functions ของ GANs สามารถเขียนในรูปของสัญลักษณ์ทางคณิตศาสตร์ได้ ดังแสดงในสมการ (1) และ (2)

$$\text{Objective functions} = \min_G \max_D V(D, G) \quad (1)$$

$$V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log (1 - D(G(z)))] \quad (2)$$

โดยที่

- $\min_G$: คือ การหาค่าต่ำสุดของฟังก์ชันสำหรับ Generator
- $\max_D$: คือ การหาค่ามากที่สุดของฟังก์ชันสำหรับ Discriminator
- $V(D, G)$: คือ ฟังก์ชันที่ใช้ในการฝึก GANs
- $G(z)$: คือ ข้อมูลที่สร้างขึ้นจาก Generator โดยที่ z เป็นข้อมูลสุ่ม
- $D(x)$: คือ ความน่าจะเป็นที่ Discriminator บอกว่าข้อมูล x เป็นข้อมูลจริง
- $E_{x \sim P_{data}(x)}$: คือ ค่าเฉลี่ยของฟังก์ชันที่คำนวณจากการแจกแจงของข้อมูลจริง x
- $E_{x \sim P_z(z)}$: คือ ค่าเฉลี่ยของฟังก์ชันที่คำนวณจากการแจกแจงของข้อมูลสุ่ม z
- $P_{data}(x)$: คือ การแจกแจงความน่าจะเป็นของข้อมูลจริง
- $P_z(z)$: คือ การแจกแจงความน่าจะเป็นของข้อมูลสุ่ม

2. การเตรียมข้อมูล

ผู้วิจัยได้ใช้ข้อมูลจากชุดข้อมูลมาตรฐาน Flickr30k [6,7] ซึ่งเป็นที่นิยมในการวิจัยด้านการสร้างคำอธิบายภาพ โดยในชุดข้อมูลนี้จะมีคำอธิบายภาพแบบขยาย (augmented image captions) และโซ่อ้างอิง (coreference chains) ที่เชื่อมโยงการกล่าวถึงเอนทิตีเดียวกันในคำอธิบายที่แตกต่างกัน สำหรับภาพเดียวกัน ซึ่งช่วยให้สามารถระบุตำแหน่งการกล่าวถึงข้อความในภาพได้อย่างชัดเจน งานวิจัยนี้ใช้จำนวนภาพทั้งหมด 158,915 ภาพ เพื่อใช้ในการพัฒนาโมเดลการสร้างคำอธิบายภาพ

ในการคำนวณขนาดกลุ่มตัวอย่าง ผู้วิจัยได้นำทฤษฎีของ Cochran (1963) มาใช้ ซึ่งทฤษฎีนี้เหมาะสำหรับการคำนวณขนาดตัวอย่างในกรณีที่ประชากรมีขนาดใหญ่และไม่ทราบสัดส่วนความแปรปรวนของประชากร (p) ที่แน่นอน โดยผู้วิจัยได้คำนวณและเลือกขนาดกลุ่มตัวอย่างทั้งหมด 384 ภาพ แบ่งเป็นชุดการฝึกฝน 80% และชุดทดสอบ 20% เพื่อใช้ในการประเมินประสิทธิภาพของโมเดล GANs ดังสมการ (3) และ (4)

$$n_0 = \frac{z^2 \cdot p \cdot (1-p)}{e^2} \quad (3)$$

โดยที่

n_0 = ขนาดกลุ่มตัวอย่างเริ่มต้น

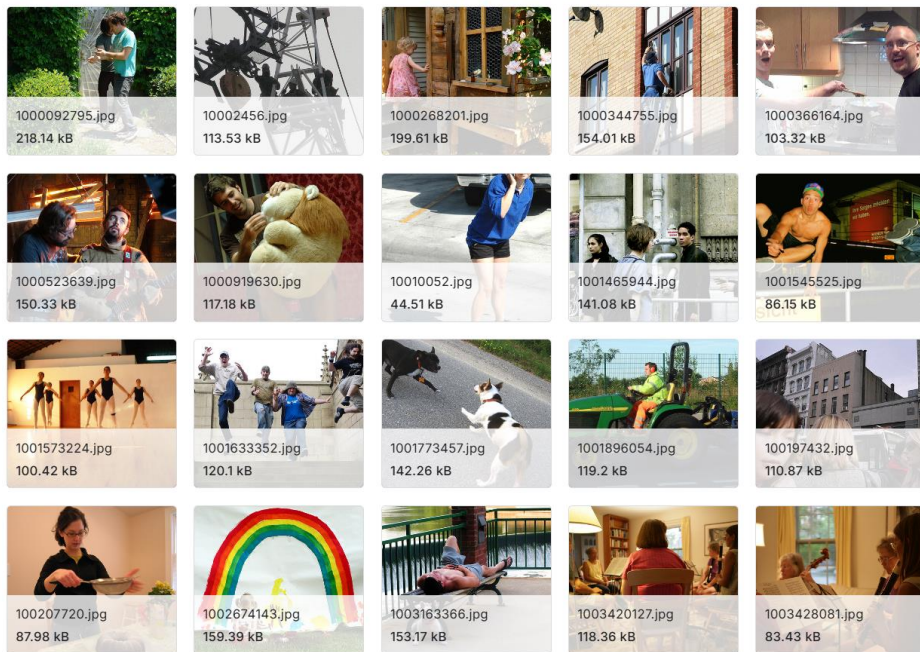
Z = ค่ามาตรฐาน z (เช่น 1.96 สำหรับระดับความเชื่อมั่น 95%)

p = สัดส่วนประชากรที่คาดว่าจะเกิด (เช่น 0.5 หากไม่ทราบค่าแน่นอน)

e = ความคลาดเคลื่อนที่ยอมรับได้ (เช่น 0.05)

จากนั้นหาขนาดกลุ่มตัวอย่างสำหรับประชากรที่มีขนาด 158,915 รายการ และใช้ค่าความเชื่อมั่น 95% ($Z=1.96$) ค่าความคลาดเคลื่อนที่ยอมรับได้ 5% ($e=0.05$) และไม่ทราบสัดส่วนประชากรที่แน่นอนจึงใช้ $p=0.5$ ดังนั้นจากการคำนวณด้วยสมการ Cochran จะได้จำนวนกลุ่มตัวอย่างทั้งหมดเท่ากับ 384 ภาพ ผลลัพธ์จากการคำนวณแสดงในสมการ (4)

$$n_0 = \frac{(1.96)^2 \cdot 0.5 \cdot (1-0.5)}{0.05^2} = \frac{3.8416 \cdot 0.25}{0.0025} = 384.16 \quad (4)$$



ภาพที่ 1 ตัวอย่างข้อมูลกลุ่มตัวอย่าง (dataset) ที่หลากหลายจากชุดข้อมูล Flickr30k

3. การฝึกโมเดล GANs

ใช้ข้อมูลจาก Kaggle เข้าสู่ Google Colab ด้วยฟังก์ชัน `load_data` เพื่อทดสอบโมเดล โดยสร้างโมเดล GANs ซึ่งประกอบด้วย generator ที่สร้างภาพและ discriminator ที่จำแนกภาพจริงและภาพปลอมด้วยฟังก์ชัน `build_generator` และ `build_discriminator` และฝึกโมเดลด้วยฟังก์ชัน `train_gan` (การฝึกโมเดลจะใช้ข้อมูลฝึกและทดสอบโมเดลด้วยข้อมูลทดสอบ ซึ่งผู้วิจัยได้แยกข้อมูลเป็นชุดฝึก (train) และทดสอบ (test) ในสัดส่วน 80/20) พร้อมแสดงผลการฝึกในแต่ละ epoch [19,20] บันทึกภาพที่สร้างด้วยฟังก์ชัน `save_imgs`

ผลลัพธ์จากการฝึก GANs ทำให้เกิดการหักล้างและอัปเดตค่า (adversarial) ระหว่าง Generator และ Discriminator ที่ทำให้ทั้งสองส่วนมีการปรับปรุงและเรียนรู้ไปพร้อมกันจนกว่า Generator จะสามารถสร้างข้อมูลที่มีคุณภาพใกล้เคียงกับข้อมูลจริงและ Discriminator จะสามารถแยกแยะข้อมูลจริงและข้อมูลปลอมได้ดีขึ้น

4. การวัดความคล้ายคลึงกันของภาพและคำอธิบายภาพด้วย cosine similarity:

การคำนวณค่า cosine similarity ในรูปแบบเวกเตอร์ TF-IDF (term frequency-inverse document frequency) [9] ซึ่งผลลัพธ์จะถูกเรียงตามค่าคะแนน (similarity score) จากสูงไปต่ำ ที่แสดงระดับความคล้ายคลึงระหว่างคำบรรยายจาก AI และคำบรรยายในชุดข้อมูลจากมากไปน้อย ตามลำดับค่าคะแนนที่ได้ ซึ่งจะช่วยให้สามารถประเมินประสิทธิภาพของโมเดลได้อย่างถูกต้องแม่นยำ การวัดความคล้ายคลึงของข้อความด้วย cosine similarity แสดงในสมการ (5)

$$\text{Cosine similarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (5)$$

โดยที่

A และ B เป็นเวกเตอร์ TF-IDF ของสองข้อความ

$A \cdot B$ เป็น dot product ระหว่างสองเวกเตอร์

$\|A\|$ และ $\|B\|$ เป็นขนาดของเวกเตอร์แต่ละตัว (norm)

การพิจารณาว่าผลลัพธ์คำอธิบายที่สร้างขึ้นกับคำอธิบายจริงนั้นมีความคล้ายคลึงหรือไม่ จะใช้เกณฑ์การกำหนดค่า Threshold ของ similarity ที่ค่าเท่ากับ 0.5 ดังนี้

1. หากความคล้ายคลึง (similarity) มากกว่าหรือเท่ากับค่าเกณฑ์ที่กำหนด 0.5 (threshold ≥ 0.5) ให้ถือว่าเป็นบวก (Positive) ซึ่งจะถือว่าคำอธิบายที่สร้างขึ้นกับคำอธิบายจริงมีความคล้ายคลึง
2. หากความคล้ายคลึง (similarity) น้อยกว่าค่าเกณฑ์ที่กำหนด 0.5 (threshold < 0.5) ให้ถือว่าเป็นลบ (negative) ซึ่งจะถือว่าคำอธิบายที่สร้างขึ้นกับคำอธิบายจริงไม่มีความคล้ายคลึง

5. การทดสอบและการประเมินผล

งานวิจัยนี้จะใช้การคำนวณค่าความแม่นยำ ค่าการเรียกคืน และค่าความถูกต้อง [6, 8] ซึ่งผลลัพธ์ที่ได้แสดงให้เห็นถึงความสามารถของโมเดลในการสร้างคำอธิบายจากสมการ (6) – (8)

ค่าความแม่นยำ (precision) คือ สัดส่วนของตัวอย่างที่ทำนายว่าเป็นบวกที่ถูกต้องเมื่อเทียบกับจำนวนตัวอย่างที่ทำนายว่าเป็นบวกทั้งหมด

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

ค่าการเรียกคืน (recall) คือ สัดส่วนของตัวอย่างที่เป็นบวกที่ถูกต้องเมื่อเทียบกับจำนวนตัวอย่างบวกทั้งหมด

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

ค่าความถูกต้อง (accuracy) คือ สัดส่วนของตัวอย่างที่ทำนายถูกต้องเมื่อเทียบกับจำนวนตัวอย่างทั้งหมด

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

โดยที่:

TP = True Positives (จำนวนตัวอย่างที่โมเดลทำนายว่าถูกต้อง (เป็นบวก) และถูกต้องตามข้อมูลทดสอบจริง)

TN = True Negatives (จำนวนตัวอย่างที่โมเดลทำนายว่าไม่ใช่ (เป็นลบ) และถูกต้องตามข้อมูลทดสอบจริง)

FP = False Positives (จำนวนตัวอย่างที่โมเดลทำนายว่าถูกต้อง (เป็นบวก) แต่ผิดไม่ตรงตามข้อมูลทดสอบ)

FN = False Negatives (จำนวนตัวอย่างที่โมเดลทำนายว่าไม่ใช่ (เป็นลบ) แต่ผิดไม่ตรงตามข้อมูลทดสอบ)

ผลการวิจัย

จากผลการวิจัยที่ใช้เทคนิค GANs ในการสร้างคำอธิบายภาพจากชุดข้อมูล Flickr30k พบว่า โมเดลสามารถสร้างคำอธิบายภาพได้อย่างมีประสิทธิภาพ โดยใช้กระบวนการฝึกฝนโมเดลด้วยการแบ่งชุดข้อมูลออกเป็นชุดฝึกฝน 80% (307 ภาพ) และชุดทดสอบ 20% (77 ภาพ) เพื่อให้โมเดลเรียนรู้การเชื่อมโยงระหว่างภาพและคำอธิบายอย่างแม่นยำ การประเมินประสิทธิภาพของโมเดลหลังการฝึกฝนแสดงให้เห็นค่าความแม่นยำอยู่ที่ 82% และค่าเรียกคืนอยู่ที่ 78% ซึ่งชี้ให้เห็นว่าโมเดลสามารถทำนายคำอธิบายภาพที่ถูกต้องได้อย่างมีนัยสำคัญ นอกจากนี้ผลการวิเคราะห์แสดงถึงค่าความผิดพลาดในจำนวนตัวอย่างที่ทำนายผิดพลาด (false positive, FP) และตัวอย่างลบที่ทำนายผิดพลาด (false negative, FN) ซึ่งชี้ถึงโอกาสในการปรับปรุงโมเดลเพื่อเพิ่มความแม่นยำยิ่งขึ้น

ตารางที่ 2 ผลการทดสอบและการวัดประเมินประสิทธิภาพของโมเดล

จำนวนภาพทดสอบทั้งหมด	TP	FP	FN	TN
77	35.45	7.77	10	26.15
ค่าความแม่นยำ (precision)	82%			
ค่าเรียกคืน (recall)	78%			
ค่าความถูกต้อง (accuracy)	80%			

การประเมินความถูกต้องของคำอธิบายที่สร้างขึ้นโดยใช้ค่าความคล้ายคลึงระหว่างคำอธิบายภาพที่สร้างโดยโมเดลและคำอธิบายภาพที่แท้จริง ผลลัพธ์ที่ได้บ่งชี้ถึงความคล้ายคลึงในระดับสูง โดยมีค่า similarity score มากกว่า 0.90 ในหลายภาพ เช่น ภาพของการเดินบนบันไดและภาพของรถแทรกเตอร์ ซึ่งแสดงให้เห็นว่าโมเดลสามารถจับคู่คำอธิบายภาพได้อย่างมีประสิทธิภาพ การทดสอบนี้ยืนยันถึงศักยภาพของ GANs ในการสร้างคำอธิบายภาพที่มีความสอดคล้องกับข้อมูลภาพจริงอย่างใกล้เคียง ซึ่งผลการทดสอบแสดงผลไว้ในตารางที่ 3

ตารางที่ 3 ผลการทดสอบโมเดลการจำแนกรูปภาพตามความหมายและคำอธิบายภาพ

ภาพที่ใช้ทดสอบ	คำอธิบายภาพจากเอไอ	Similarity score
	a man driving a green tractor while wearing a safety vest	0.91672711730003357
	two dogs interacting with each other on a paved road.	0.92926058292388916
	a colorful drawing of a rainbow with red, yellow, green, and blue	0.97672711730003357
	a woman is shown holding a kitchen utensil	0.90208024561405182
	a group of ballet dancers performing in a studio	0.9162032449245453
	three people, two wearing casual clothes and one in a cap, jumping	0.9264212667942047

อภิปรายผลการวิจัย

งานวิจัยนี้รองรับ SDG หมวดที่ 9 ด้านการส่งเสริมและพัฒนาอุตสาหกรรมที่ยั่งยืนด้วยนวัตกรรมเอไอ โดยผู้วิจัยได้ทดสอบให้เห็นความสามารถของ GANs ในการจับคู่ระหว่างภาพและคำอธิบาย ซึ่งสอดคล้องกับงานของ จักรินทร์ สันติรัตนภักดี และ ศุภกฤษฎ์ นิวัฒนากุล [7] ที่ใช้การจัดตำแหน่งระหว่างภาพและภาษาธรรมชาติที่มีประสิทธิภาพ งานวิจัยนี้มีความแตกต่างจากการใช้ variational autoencoder (VAE) model ของ Nattawuttisit และ Chantron [10] ที่เน้นการแปลงข้อมูลภาพด้วยการเข้ารหัส ซึ่งจะมีความสามารถในการสร้างภาพที่มีมิติที่ซับซ้อนได้ อย่างไรก็ตาม การใช้โมเดล GANs ในงานวิจัยนี้ยังมีข้อจำกัดบางประการ เช่น ข้อมูลที่ใช้ยังขาดความหลากหลายจะทำให้เกิดความเสี่ยงของ mode collapse เมื่อ Generator สร้างข้อมูลที่ซ้ำกันบ่อย ๆ หรือสร้างข้อมูลในรูปแบบเดียวกัน ซึ่งทำให้ข้อมูลที่สร้างขาดความหลากหลายและไม่สามารถครอบคลุมรูปแบบของข้อมูลทั้งหมดได้ นอกจากนี้ GANs ยังไม่สามารถเทียบเท่ากับโมเดลที่ใช้ transformer-based อย่าง BERT หรือ GPT ในการสร้างคำอธิบายที่เป็นลำดับได้อย่าง

มีประสิทธิภาพ ดังนั้นงานวิจัยในอนาคตควรมุ่งเน้นการเปรียบเทียบโมเดลเหล่านี้เพื่อนำมาใช้ปรับปรุงขยายผลต่อผลงานวิจัยนี้ให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น

สรุปผลการวิจัย

จากการพัฒนาแบบจำลองครั้งนี้พบว่า เป็นไปตามวัตถุประสงค์ที่วางไว้ ผู้วิจัยได้ใช้เกณฑ์การวัดความคล้ายคลึงกันของภาพและคำอธิบายภาพจากการคำนวณค่า cosine similarity ในรูปแบบเวกเตอร์ TF-IDF ซึ่งผลลัพธ์จะถูกเรียงตามค่าคะแนน (similarity score) จากสูงไปต่ำที่แสดงระดับความคล้ายคลึงระหว่างคำบรรยายจาก AI และคำบรรยายในชุดข้อมูลจากมากไปน้อยตามลำดับค่าคะแนนที่ได้ ผลการประเมินประสิทธิภาพของโมเดล GAN ด้วยชุดข้อมูลทดสอบ โดยการตรวจสอบได้ทำร่วมกับผู้เชี่ยวชาญด้านปัญญาประดิษฐ์ จากผลการวัดและประเมินได้แสดงให้เห็นว่าโมเดล GAN สามารถสร้างคำอธิบายภาพที่มีความหมายใกล้เคียงกับต้นฉบับที่ใช้ทดสอบ โดยให้ค่าความแม่นยำและการเรียกคืนในระดับความถูกต้องได้ในระดับดี ซึ่งผลการวิจัยนี้สามารถประยุกต์ใช้เชิงพาณิชย์ได้ เช่น การจัดหมวดหมู่รูปภาพอัตโนมัติในสื่อสังคม หรือการค้นหารูปภาพในคลังข้อมูลขนาดใหญ่ หรือ สามารถนำโมเดลนี้ไปประยุกต์ใช้ร่วมกับโมเดล transformer-based อย่าง BERT หรือ GPT สำหรับการสร้างคำอธิบายที่เป็นลำดับเพื่อให้โมเดลมีประสิทธิภาพมากขึ้น

เอกสารอ้างอิง

- [1] Goodfellow, I., Bengio, Y., & Courville, A. (2023). *Deep Learning*. MIT Press.
- [2] Alif, M. D. N., & Fahrudin, N. F. (2024). Performance Analysis of Oversampling and Undersampling on Telco Churn Data Using Naive Bayes, SVM And Random Forest Methods (pp 1–13). In *E3S Web of Conferences*, 484, 02004.
- [3] Zhang, Y., Wang, S., & Li, H. (2023). Enhancing image classification with convolutional neural networks: a comprehensive review. *Journal of Computer Vision and Image Processing*, 45(2), 123–140.
- [4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- [5] Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes (pp 1–14). In *The International Conference on Learning Representations (ICLR) 2014*.
- [6] Papers with Code. (2024). Flickr30k Dataset. [Online], Retrieved from <https://paperswithcode.com/dataset/flickr30k> (7 August 2024).
- [7] จักรินทร์ สันติรัตนภักดี และ ศุภกฤษฎี นิวัฒนากุล. (2567). แบบจำลองการจำแนกรูปภาพตามความหมาย ได้รับการฝึกฝนสำหรับการเรียกค้นรูปภาพโดยใช้ภาษาธรรมชาติ. *วารสารวิชาการชาชนไทย มรภ.ภูเก็ต*, 8(1), 68–82.

- [8] Vivekananthan, S. (2024). Comparative analysis of generative models: Enhancing image synthesis with VAEs, GANs and stable diffusion. *arXiv*, 2408.08751.
- [9] Hassan, R. T., & Ahmed, N. S. (2023). Evaluating of efficacy semantic similarity methods for comparison of academic thesis and dissertation texts. *Science Journal of University of Zakho*, 11(3), 396–402.
- [10] Nattawuttisit, S., & Chantron, P. (2024). Revolutionizing AI driven innovations in gemstone classification: a synergistic approach integrating visual and semantic NLP techniques. *Nanotechnology Perceptions*, 20(4), 333–345.