

Feature Extraction using Convolutional Neural Network for Building Image Retrieval

Achirawit Kumphet¹, Wachirawat Malaikhot¹, Kamonrat Somchai¹, Sangdaow Noppitak^{1,*}

ABSTRACT

This research aimed to extract image features using Convolutional Neural Networks (CNNs) for a content-based image retrieval (CBIR) system of buildings. A dataset of 513 photographs of various buildings within Buriram Rajabhat University was utilized. Three CNN models, VGG16, VGG19, and NasNetLarge, were employed for feature extraction, with Cosine similarity. The results indicated that VGG19 yielded the highest accuracy in image retrieval, followed by VGG16 and NasNetLarge, respectively. This suggests that VGG19 is more effective at extracting building image features than VGG16 and NasNetLarge. However, the study revealed that NasNetLarge, a more complex model, had limitations in extracting features from building images, possibly due to its higher structural complexity compared to VGG16 and VGG19, resulting in lower performance in this specific application.

Keywords: Feature Extraction, Convolutional Neural Network, Content-Based Image Retrieval, Cosine Similarity, Deep Learning

Published Online: 17 June 2024

ISSN: 2730-3829

Wachirawit kumphet¹

¹Faculty of Science,
Buriram Rajabhat University
(630112418054@bru.ac.th)

Wachirawat Malaikhot¹

¹Faculty of Science,
Buriram Rajabhat University
(630112418018@bru.ac.th)

Kamonrat Somchai¹

¹Faculty of Science,
Buriram Rajabhat University
(kamonrat.sj@bru.ac.th)

Sangdaow Noppitak^{1,*}

¹Faculty of Science,
Buriram Rajabhat University
(sangdaow.np@bru.ac.th)

*Corresponding Author

Received date: 14 March 2023

Revised date#1: 2 May 2024

Revised date#2: 1 June 2024

Accepted date: 4 June 2024

การสกัดคุณลักษณะโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน สำหรับระบบค้นคืนรูปภาพสิ่งปลูกสร้าง

วชิรวิทย์ กุมเพ็ชร¹, วชิรวัฒน์ มาลัยโคตร¹, กมลรัตน์ สมใจ¹, แสงดาว นพพิทักษ์^{1,*}

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อสกัดคุณลักษณะของภาพโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) เพื่อนำไปประยุกต์ใช้ในระบบค้นคืนรูปภาพตามเนื้อหา (Content-Based Image Retrieval: CBIR) สำหรับการค้นคืนภาพอาคาร โดยใช้ภาพถ่ายอาคารต่างๆ ภายในมหาวิทยาลัยราชภัฏบุรีรัมย์จำนวน 513 ภาพ เป็นชุดข้อมูลทดสอบ โมเดล CNNs ที่ใช้ในการสกัดคุณลักษณะ ได้แก่ VGG16, VGG19 และ NasNetLarge โดยใช้ Cosine similarity ในการวัดความคล้ายคลึงของภาพ ผลการวิจัยพบว่า VGG19 ให้ผลลัพธ์ความแม่นยำในการค้นคืนภาพที่ดีที่สุด รองลงมาคือ VGG16 และ NasNetLarge ตามลำดับ แสดงให้เห็นว่า VGG19 สามารถสกัดคุณลักษณะของภาพอาคารได้อย่างมีประสิทธิภาพมากกว่า VGG16 และ NasNetLarge อย่างไรก็ตาม NasNetLarge ยังคงมีข้อจำกัดในการสกัดคุณลักษณะภาพอาคาร เนื่องจาก NasNetLarge มีความซับซ้อนของโครงสร้างสูงกว่า VGG16 และ VGG19 จึงทำให้ประสิทธิภาพในการสกัดคุณลักษณะภาพอาคารต่ำกว่า

คำสำคัญ: การสกัดคุณลักษณะ, โครงข่ายประสาทแบบคอนโวลูชัน, การค้นคืนรูปภาพ, ค่าความคลึง, การเรียนรู้เชิงลึก

Published Online: 17 มิถุนายน 2566

ISSN: 2730-3829

วชิรวิทย์ กุมเพ็ชร¹

¹คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
(630112418054@bru.ac.th)

วชิรวัฒน์ มาลัยโคตร¹

¹คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
(630112418018@bru.ac.th)

กมลรัตน์ สมใจ¹

¹คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
(kamonrat.sj@bru.ac.th)

แสงดาว นพพิทักษ์^{1,*}

¹คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
(sangdaow.np@bru.ac.th)

^{*}Corresponding Author

Received date: 14 มีนาคม 2566

Revised date#1: 2 พฤษภาคม 2567

Revised date#2: 1 มิถุนายน 2567

Accepted date: 4 มิถุนายน 2567

1. บทนำ

การสกัดคุณลักษณะ (Feature Extraction) เป็นวิธีการแยกคุณสมบัติมีความสำคัญในงานด้านต่าง ๆ เช่น การประมวลผลภาพ และการวินิจฉัยข้อผิดพลาด เป็นต้น งานเหล่านี้เน้นความถูกต้องและการดึงข้อมูลจากข้อมูลตัวอย่างได้อย่างมีประสิทธิภาพ เช่น การประมวลผลภาพสำหรับการตรวจจับหน้ากาอนามัยโดยใช้การสกัดคุณลักษณะ (Usha, Sudeepthi, Mahathi, & Shravya, 2021) อีกทั้งการสกัดคุณลักษณะถือเป็นขั้นตอนที่สำคัญสำหรับงานด้านวินิจฉัยความผิดพลาด ทำให้สามารถแยกลักษณะความผิดพลาดได้อย่างมีประสิทธิภาพ (Shi et al., 2022) วิธีการในการสกัดคุณลักษณะที่หลากหลาย แสดงให้เห็นถึงความสำคัญของการแยกคุณลักษณะในงานด้านต่าง ๆ เพื่อเพิ่มความถูกต้องและทำให้การค้นคืนข้อมูลทำงานได้อย่างมีประสิทธิภาพ ปัจจุบันการสกัดคุณลักษณะโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) เป็นเทคนิคที่มีประสิทธิภาพเทคนิคหนึ่ง CNNs มีความโดดเด่นในการสกัดคุณลักษณะจากข้อมูลที่ซับซ้อน ดังงานวิจัยของ Sharma, Kumar, & Gupta (2023) ได้ทำการสกัดคุณลักษณะของถนนออกจากภาพรับรู้ระยะไกลความละเอียดสูง (High-Resolution Remote Sensing Images: HRRSI) เป็นเรื่องที่ทำนายเนื่องจากโครงสร้างที่ซับซ้อนของถนน เสา และการบดบังที่เกิดจากต้นไม้ รถยนต์ และอาคาร เทคนิคแบบดั้งเดิม (เช่น SVM, ANN ฯลฯ) มีข้อจำกัดในภาพที่มีลักษณะถนนเชื่อมต่อกันหรือภาพของถนนมีลักษณะกระจัดกระจาย แต่ CNNs มีประสิทธิภาพที่เหนือกว่า และงานด้านการแพทย์ต้องการการวินิจฉัยที่แม่นยำจึงถูกนำไปใช้อย่างแพร่หลาย อีกทั้งได้มีการวิจัยเพื่อค้นหาวิธีการสกัดคุณลักษณะของสัญญาณเสียงที่มีประสิทธิภาพมากที่สุด โดยการเปรียบเทียบ CNNs หลายเทคนิค รวมถึง CNNs ถูกนำมาใช้สำหรับการสกัดคุณลักษณะจากภาพเนื้องอกในสมอง เพื่อเพิ่มความแม่นยำในการวินิจฉัยและการรักษาผ่านการวิเคราะห์ภาพที่มีประสิทธิภาพ (Malleswari, G., & Reddy, A. S., 2023; Rashmi, P., & Singh, M. P., 2023)

จากการสืบค้นงานวิจัยจากหลายแหล่งได้พิจารณาและเปรียบเทียบการใช้ CNNs ในการสกัดและจำแนกคุณลักษณะจากภาพ เช่น Chen, C, (2023) นำเสนอการใช้ CNNs แทนการสกัดคุณลักษณะแบบเดิม โดยใช้เทคนิค Contrastive Language-Image Pretraining (CLIP) เพื่อเพิ่มความแม่นยำในการดึงข้อมูลภาพ Mathews et al., (2022) สำนวจประสิทธิภาพของ CNNs ในการค้นคืนภาพจากฐานข้อมูลเสื้อผ้าและรองเท้า พบว่าโมเดล VGG19, InceptionV3, และ Xception แสดงประสิทธิภาพสูงสุด Wei, D., (2022) ได้ศึกษาวิจัยการรู้จำการแสดงออกบนใบหน้าโดยใช้ VGG19 และ Resnet18 ในการสกัดคุณลักษณะของภาพ ซึ่งให้ผลลัพธ์ที่ดีกว่าวิธีการอื่น ๆ และ Aung, S. P. P., Pa, W. P., & Nwe, T. L., (2023) พัฒนาระบบสร้างคำบรรยายภาพภาษาเมียนมาร์โดยใช้ NASNetLarge และ Bi-LSTM ซึ่งได้คะแนน BLEU-4 สูงสุด 40.05% และ Gayathri et al. (2023) ใช้ VGG-16 เพื่อตรวจจับเนื้องอกในสมอง โดยให้ความแม่นยำสูงสุดที่ 94% หลังการปรับพารามิเตอร์ จากรายละเอียดข้างต้นแสดงให้เห็นว่า VGG16 และ VGG19 เป็นสถาปัตยกรรมที่ได้รับความนิยม เนื่องจากความเรียบง่ายและประสิทธิภาพในการสกัดคุณลักษณะภาพ และได้รับการฝึกฝนกับชุดข้อมูลขนาดใหญ่ (ImageNet) ส่วน NASNetLarge เป็นสถาปัตยกรรมขั้นสูง ซึ่งได้รับการออกแบบโดยใช้เทคนิค Neural Architecture Search (NAS) หรือการค้นหาโครงสร้างเครือข่ายประสาทอัตโนมัติ เป็นที่รู้จักในด้านความแม่นยำและประสิทธิภาพสูงในเมื่อเทียบกับที่อื่นสถาปัตยกรรมอื่น (Himel, G. M. S., & Islam, M. M., 2024)

ดังนั้นงานวิจัยนี้ ใช้ภาพถ่ายของอาคารต่าง ๆ ในมหาวิทยาลัยราชภัฏบุรีรัมย์เป็นข้อมูลในการทดสอบความสามารถของ CNNs สามสถาปัตยกรรม ได้แก่ VGG16, VGG19 และ NasNetLarge ในการสกัดคุณลักษณะจากภาพ การวัดผลจะใช้ค่า Cosine similarity เพื่อเปรียบเทียบความคล้ายคลึงของภาพที่สกัดคุณลักษณะแล้วผลการวิจัยจะช่วยให้เข้าใจถึงประสิทธิภาพและข้อจำกัดของแต่ละโมเดล CNNs ในการค้นคืนภาพตามเนื้อหา ซึ่งสามารถนำไปใช้ปรับปรุงและพัฒนาเทคโนโลยีการค้นคืนภาพในอนาคตได้

วัตถุประสงค์

เพื่อสกัดคุณลักษณะโดยใช้โครงข่ายประสาทแบบคอนโวลูชันสำหรับระบบค้นคืนรูปภาพสิ่งปลูกสร้าง

2. วิธีการวิจัย

งานวิจัยนี้ ใช้ภาพถ่ายของอาคารต่าง ๆ ในมหาวิทยาลัยราชภัฏบุรีรัมย์เป็นข้อมูลในการทดสอบความสามารถของโมเดล CNNs ทั้งสามแบบ การวัดผลจะใช้ค่า Cosine similarity เพื่อเปรียบเทียบความคล้ายคลึงของภาพที่สกัดคุณลักษณะแล้ว ผลการวิจัยจะช่วยให้เข้าใจถึงประสิทธิภาพและข้อจำกัดของแต่ละโมเดล CNNs ในการค้นคืนภาพเฟรมเวิร์คของการวิจัยดัง Figure 1

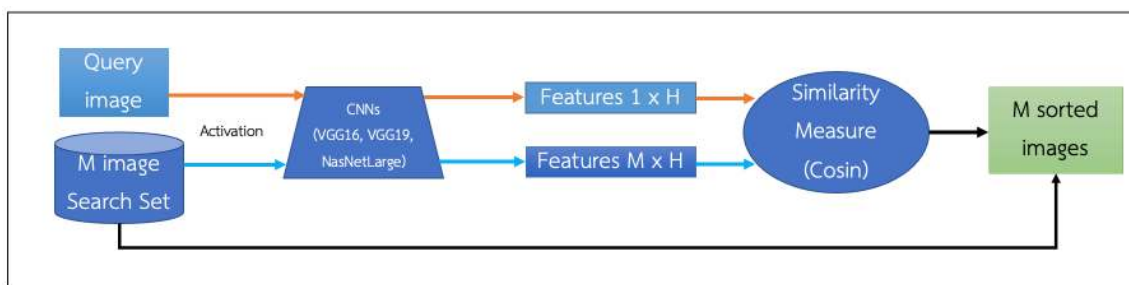


Figure 1 Research framework

จาก Figure 1 แสดงการค้นคืนรูปภาพตามเนื้อหา (Content-Based Image Retrieval, CBIR) โดยใช้ Convolutional Neural Networks (CNNs) ได้แก่ VGG16, VGG19, และ NasNetLarge เพื่อสกัดคุณลักษณะจากภาพคำถาม (Query Image) และภาพในชุดค้นหา (Image Search Set) โดย CNNs จะสร้างเวกเตอร์คุณลักษณะขนาด $1 \times H$ สำหรับภาพคำถาม และ $M \times H$ สำหรับภาพในชุดค้นหา จากนั้นใช้การวัดความคล้ายคลึงกันด้วย Cosine เพื่อเปรียบเทียบเวกเตอร์คุณลักษณะของภาพคำถามกับภาพในชุดค้นหา คำนวณคะแนนความคล้ายคลึงกันระหว่างภาพคำถามและภาพแต่ละภาพในชุดค้นหา และจัดเรียงภาพตามคะแนนความคล้ายคลึงจากมากไปน้อย ผลลัพธ์คือการแสดงภาพที่คล้ายที่สุดกับภาพคำถามตามคะแนนความคล้ายคลึงที่คำนวณได้ ซึ่งช่วยให้สามารถค้นหาและจัดเรียงภาพที่คล้ายกับภาพคำถามได้อย่างมีประสิทธิภาพ

2.1 ทฤษฎีที่เกี่ยวข้อง

1) การสกัดคุณลักษณะ

การสกัดคุณลักษณะ (Feature Extraction) คือ กระบวนการที่ใช้ในการดึงข้อมูลที่สำคัญและเป็นตัวแทนจากชุดข้อมูลดิบ เพื่อนำไปใช้ในการสร้างแบบจำลองหรือการวิเคราะห์ต่าง ๆ เป็นกระบวนการที่สำคัญในหลายสาขา เช่น การวิเคราะห์ข้อมูลเชิงพื้นที่ การประมวลผลสัญญาณ การวินิจฉัยข้อผิดพลาด การจำแนกประเภท และการเรียนรู้ของเครื่อง (Machine Learning) เป็นต้น เทคนิคที่ถูกนำมาใช้ในการสกัดคุณลักษณะ เช่น Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP), Gaussian Models, ลักษณะทางสถิติ และการเรียนรู้เชิงลึก (Deep Learning) เป็นต้น การสกัดคุณลักษณะมีบทบาทสำคัญในการเพิ่มประสิทธิภาพในการคำนวณ ลดมิติ และปรับปรุงความแม่นยำของแบบจำลอง (Liu, Q., Zhang, J., Liu, J., & Yang, Z., 2022; Suhaidi, M., Abdul Kadir, R., & Tiun, S., 2021; Xu, X., & Zheng, L., 2022)

การสกัดคุณลักษณะโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) เป็นเทคนิคที่มีประสิทธิภาพสูง มีความโดดเด่นในการสกัดคุณลักษณะจากข้อมูลที่ซับซ้อน เช่น ภาพความละเอียดสูงจากการตรวจจักษะระยะไกล (Sharma, P., Kumar, R., & Gupta, M., 2023) สัญญาณเสียงของตัวอักษรภาษาอังกฤษ (Zhang, Luo, Cao, Shen, & Wang, 2023), และภาพทางการแพทย์สำหรับการวินิจฉัยเนื้องอก (Malleswari, G., & Reddy, A. S., 2023; Rashmi, P., & Singh, M. P., 2023) เป็นต้น ซึ่ง CNNs สามารถเรียนรู้และสกัดคุณลักษณะได้

โดยอัตโนมัติในระดับและความถี่ที่แตกต่างกัน ช่วยเพิ่มความสามารถในการปรับตัวให้เข้ากับชุดข้อมูลที่หลากหลายได้มากขึ้น ในด้านการแพทย์มีการเปรียบเทียบการสกัดคุณลักษณะภาพเนื้องอกโดยใช้ CNNs สองสถาปัตยกรรม ได้แก่ DenseNet และ Xception ผลการวิจัย พบว่า CNNs สามารถสกัดคุณลักษณะได้อย่างแม่นยำ โดยการใช้ความสามารถในการเรียนรู้รูปแบบและคุณลักษณะที่ซับซ้อนจากข้อมูลประเภทต่างๆ (Malleswari, G., & Reddy, A. S., 2023)

2) โครงข่ายประสาทแบบคอนโวลูชัน

โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) เป็นของโครงข่ายประสาทเทียมประเภทหนึ่งที่ถูกออกแบบมาเพื่อใช้ในงานด้านการประมวลผลภาพเป็นหลัก (Celeghin, A. et al., 2023; Jiang, C., & Goldsztein, G., 2023) CNNs มีความสามารถในการประมวลผลและวิเคราะห์ข้อมูลภาพ โดยการเรียนรู้คุณลักษณะเชิงลำดับชั้นโดยอัตโนมัติผ่านชั้นคอนโวลูชัน (Convolutional Layers) ชั้นการรวมข้อมูล (Pooling layer) และชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) (Chen, X., 2023) โครงข่ายเหล่านี้มีความสามารถในงานด้านต่างๆ เช่น การจำแนกภาพ การตรวจจับวัตถุ และการรู้จำ โดยสามารถดึงและจำแนกคุณลักษณะจากภาพได้อย่างมีประสิทธิภาพในระหว่างการเรียนรู้ (Taye, 2023) ในขณะเดียวกัน CNNs ยังเพิ่มความทนทานและความสามารถในการรองรับข้อผิดพลาด (Karrach, L., & Pivarčiová, E., 2023)

VGG16 และ VGG19 เป็น CNNs ที่ได้รับการพัฒนาโดย Visual Geometry Group (VGG) จาก University of Oxford โมเดลเหล่านี้ได้รับความนิยมอย่างแพร่หลายสำหรับการจำแนกภาพ โดยมีการออกแบบที่เน้นการเพิ่มจำนวนเลเยอร์ที่มีความลึกมากขึ้น เพื่อให้สามารถดึงคุณลักษณะจากภาพได้อย่างละเอียด โดย VGG16 ประกอบด้วย 16 เลเยอร์ที่มีการเรียนรู้ (13 Convolutional layers และ 3 Fully connected layers) ในขณะที่ VGG19 มี 19 เลเยอร์ (16 Convolutional layers และ 3 Fully connected layers) ทั้งสองโมเดลนี้ใช้ขนาดของฟิลเตอร์คอนโวลูชันที่เล็ก (3x3) และมีการใช้ Max-Pooling layers เพื่อทำการลดขนาดภาพ (Simonyan, K., & Zisserman, A., 2014)

NasNetLarge เป็น CNNs ที่ถูกออกแบบโดยใช้เทคนิค Neural Architecture Search (NAS) ซึ่งพัฒนาโดย Google AI โมเดลนี้ถูกออกแบบมาให้มีประสิทธิภาพสูงในการจำแนกภาพ โดยใช้โครงสร้างที่ถูกค้นพบโดยอัลกอริธึมการค้นหาอัตโนมัติ ทำให้สามารถค้นหาโครงสร้างที่มีประสิทธิภาพสูงสุดในงานจำแนกภาพต่าง ๆ NasNetLarge ประกอบด้วยหลายโมดูลที่ถูกออกแบบให้สามารถเรียนรู้คุณลักษณะจากภาพได้อย่างมีประสิทธิภาพ โครงสร้างของสถาปัตยกรรมนี้ เน้นการใช้การเชื่อมโยงแบบ Skip connections และการใช้ Batch Normalization เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ (Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V., 2018)

3) การค้นคืนภาพจากเนื้อหาของรูปภาพ

การค้นคืนภาพจากเนื้อหา (Content-Based Image Retrieval: CBIR) เป็นวิธีการที่ใช้เทคนิคการมองเห็นของคอมพิวเตอร์ในการค้นหาภาพดิจิทัลภายในฐานข้อมูลขนาดใหญ่ (Rani & Yuhandri, 2023) ระบบ CBIR เน้นการสกัดคุณลักษณะจากภาพเพื่อจำแนกและเรียกคืนภาพที่คล้ายคลึงกันตามคำค้นหา (Malik, C. K. M., 2022) ระบบเหล่านี้สามารถใช้วิธีการต่าง ๆ เช่น การค้นหาตามสีโดยใช้เมทริกซ์ความคล้ายคลึง HSV, การค้นหาตามพื้นผิวด้วยการแยกย่อยคลื่น และการวัดความคล้ายคลึงที่คำนวณผ่านตัวจำแนกระยะทางยูคลิด (Lazim Lafta & I. Abdulsada, 2022) นอกจากนี้ ความก้าวหน้าต่าง ๆ เช่น สถาปัตยกรรม Residual Network (ResNet-18) และ CNNs ถูกนำมาใช้เพื่อกำหนดคะแนนความคล้ายคลึงของภาพได้อย่างแม่นยำในแอปพลิเคชัน CBIR (Abdullah, M. N., & Sahari, M. A., 2022)

4) การวัดความคล้ายคลึง

Cosine similarity เป็นวิธีการวัดความคล้ายคลึงระหว่างเวกเตอร์สองตัวในมิติหลายมิติ โดยคำนวณจากมุมระหว่างเวกเตอร์เหล่านั้น แทนที่จะคำนวณจากระยะห่างระหว่างเวกเตอร์ วิธีนี้มีประโยชน์ในด้านการวิเคราะห์ข้อมูล เช่น การทำเหมืองข้อความ การจัดกลุ่มข้อมูล และการเรียกคืนข้อมูล Cosine similarity จะให้ค่าระหว่าง -1 ถึง 1 ซึ่ง

ค่า 1 หมายถึง เวกเตอร์สองตัวนั้นมิติศทางเดียวกัน ค่า 0 หมายถึง เวกเตอร์สองตัวนั้นตั้งฉากกัน และค่า -1 หมายถึง เวกเตอร์สองตัวนั้นมิติศทางตรงกันข้าม (Park, Hong, & Kim, 2020) ดังสมการ (1)

$$\text{Cosine similarity} = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \times \|\vec{B}\|} \quad (1)$$

โดยที่ $\vec{A} \cdot \vec{B}$ คือ ผลคูณจุดของเวกเตอร์ $\vec{A}$ และ $\vec{B}$
 $\|\vec{A}\|$ คือ ขนาดของเวกเตอร์ $\vec{A}$
 $\|\vec{B}\|$ คือ ขนาดของเวกเตอร์ $\vec{B}$

2.2 งานวิจัยที่เกี่ยวข้อง

Chen, C., (2023) ใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) ในการสกัดคุณลักษณะจากภาพแทนการสกัดคุณลักษณะด้วยตนเอง CNNs มีประสิทธิภาพในการเรียนรู้และการแสดงคุณสมบัติของภาพแบบลำดับขั้น ทำให้สามารถแยกคุณลักษณะที่แม่นยำยิ่งขึ้น จากนั้นใช้เทคนิค Contrastive Language-Image Pretraining (CLIP) เพื่อดึงเวกเตอร์คุณลักษณะมิติสูงจากชุดข้อมูลภาพก่อนทำการเรียนรู้ ซึ่งเทคนิค CLIP ได้รับการออกแบบมาเพื่อจับคู่ภาพกับข้อความ ทำให้เพิ่มความเข้าใจเกี่ยวกับคุณสมบัติของภาพและแก้ไข “ช่องว่างความหมาย” ระหว่างภาพและข้อความได้ ทำให้การดึงข้อมูลด้วยอัลกอริทึม Local Sensitive Hash (LSH) มีความรวดเร็ว สามารถทำการดึงข้อมูลได้อย่างมีประสิทธิภาพตามเนื้อหาภาพและข้อความ อีกทั้งยังช่วยเพิ่มกระบวนการดึงข้อมูลได้โดยรวม

จากงานวิจัยของ Mathews, A., N, S., & R, V. K. (2022) ได้ทำการสำรวจประสิทธิภาพของ CNNs ในการสกัดคุณลักษณะสำหรับการค้นคืนภาพกับชุดข้อมูล Footwear Dataset และ Women Cothes Dataset งานวิจัยนี้เปรียบเทียบประสิทธิภาพระหว่างอัลกอริทึมแบบดั้งเดิมกับ CNNs ได้แก่ VGG16, VGG19, InceptionV3, Xception และ ResNet50 ในการดึงคุณลักษณะจากชุดข้อมูลรองเท้าผู้ชายและเสื้อผ้าผู้หญิง โดยใช้คุณลักษณะที่ดึงมาในการจัดประเภทด้วย SVM, Random Forest และ K-Nearest Neighbors โมเดล VGG19, InceptionV3 และ Xception แสดงประสิทธิภาพที่เหนือกว่าในด้านการดึงคุณลักษณะและการจัดประเภทภาพ โดยมีความแม่นยำถึง 97.50% งานวิจัยยังตรวจสอบประสิทธิภาพการค้นคืนภาพโดยใช้คุณลักษณะเหล่านี้และการวัดความคล้ายคลึงกัน

Wei, D. (2022) จากการพัฒนาเทคโนโลยีอย่างรวดเร็วและการเพิ่มขึ้นของสื่อสังคมออนไลน์ ข้อมูลภาพได้เติบโตอย่างรวดเร็ว ดังนั้น ภาพจึงมักถูกใช้เป็นการสื่อสาร เมื่อพูดถึงการรู้จำภาพ การดึงคุณลักษณะของภาพเป็นขั้นตอนที่สำคัญ และผลของการดึงคุณลักษณะของภาพมีผลกระทบโดยตรงต่อประสิทธิภาพของการรู้จำภาพ ยิ่งไปกว่านั้น การดึงคุณลักษณะเป็นปัจจัยสำคัญที่ต้องพิจารณาซึ่งส่งผลต่อความแม่นยำในการรู้จำภาพ อย่างไรก็ตาม เนื่องจากผลกระทบจากความแตกต่างของบุคคลและการจัดแสง ทำให้องค์ประกอบบางอย่างที่เชื่อมโยงอย่างมีนัยสำคัญกับการเปลี่ยนแปลงของภาพยากที่จะดึงออกมา ดังนั้น คุณลักษณะที่แสดงถึงการเปลี่ยนแปลงของภาพได้อย่างเหมาะสมจึงเป็นที่ต้องการอย่างเร่งด่วน เพื่อจุดประสงค์นี้ งานวิจัยนี้เสนอวิธีการระบุการแสดงออกบนใบหน้าที่ใช้ CNNs สำหรับงานการรู้จำการแสดงออกบนใบหน้าภายใต้การดึงคุณลักษณะของภาพออนไลน์ โดยใช้ VGG19 และ Resnet18 ในการรู้จำและจำแนกการแสดงออกบนใบหน้า

Aung, S. P. P., Pa, W. P., & Nwe, T. L., (2023) การพัฒนาการสร้างคำบรรยายภาพภาษาเมียนมาร์อัตโนมัติโดยเรียนรู้เนื้อหาของภาพโดยใช้สถาปัตยกรรม NASNetLarge และ Bi-LSTM การอธิบายเนื้อหาของภาพเป็นงานที่ซับซ้อนสำหรับเครื่องจักรโดยไม่มีกรแทรกแซงจากมนุษย์ คอมพิวเตอร์วิทัศน์ (Computer vision) และการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ถูกใช้อย่างแพร่หลายเพื่อแก้ไขปัญหานี้ งานวิจัยนี้ได้นำเสนอระบบการสร้างคำบรรยายภาพภาษาเมียนมาร์ที่ใช้การเรียนรู้เชิงลึก ซึ่งใช้โมเดลการดึงคุณลักษณะ

NASNetLarge ของ CNNs เป็นตัวเข้ารหัสและใช้เครือข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) กับหน่วยความจำระยะสั้นสองทิศทาง (Bi-LSTM) เป็นตัวถอดรหัส สำหรับการสร้างคลังข้อมูล ได้สร้างและใส่คำบรรยายภาพภาษาเมียนมาร์ (ประกอบด้วยประโยคภาษาเมียนมาร์กว่า 40,000 ประโยค) ซึ่งอ้างอิงจากชุดข้อมูล Flickr8k ผลการทดลอง พบว่า โมเดล NASNetLarge กับ Bi-directional LSTM ที่ใช้การแบ่งพยางค์สามารถทำคะแนน BLEU-4 ได้สูงสุด 40.05%

Gayathri, Dhavileswarapu, Ibrahim, Paul, & Gupta, (2023) ได้ศึกษาศักยภาพสถาปัตยกรรม VGG-16 เพื่อตรวจจับเนื้องอกในสมอง โดยใช้ชุดข้อมูลที่ประกอบด้วยภาพ MRI ของสมองที่มีเนื้องอก 1655 ภาพ และไม่มีเนื้องอก 1598 ภาพ ผลการทดลอง พบว่า มีความแม่นยำ 91% และเพิ่มขึ้นเป็น 94% หลังการปรับพารามิเตอร์ประสิทธิภาพของโมเดล VGG-16 ถูกเปรียบเทียบกับเทคนิคอื่น ๆ ในการตรวจจับเนื้องอกในสมอง เช่น EasyDL, GoogLeNet, GrayNet, ImageNet, CNN และโมเดล Multivariable Regression and Neural Network แม้ว่าจะไม่ได้มีความแม่นยำสูงสุด แต่ก็ทำได้ดีกว่า GoogLeNet และมีความแม่นยำใกล้เคียงกับ GrayNet และโมเดล Multivariable Regression and Neural Network

3. ผลการวิจัย

3.1 ชุดข้อมูล

งานวิจัยนี้สร้างชุดข้อมูลภาพอาคารภายในมหาวิทยาลัยราชภัฏบุรีรัมย์ เพื่อใช้เป็นภาพคำถามและภาพในชุดค้นหา ทุกภาพถ่ายจากกล้องโทรศัพท์มือถือ ประเภทไฟล์เป็น .jpg และมีขนาดภาพ 1080 x 1920 พิกเซล บันทึกภาพในหลายช่วงเวลาแต่อยู่ในช่วงกลางวันทั้งหมด มีภาพทั้งหมดจาก 5 อาคาร ได้แก่ อาคาร 17, 22, 24, 25, และ 26 รวมทั้งสิ้น 518 ภาพ ดัง Figure 2

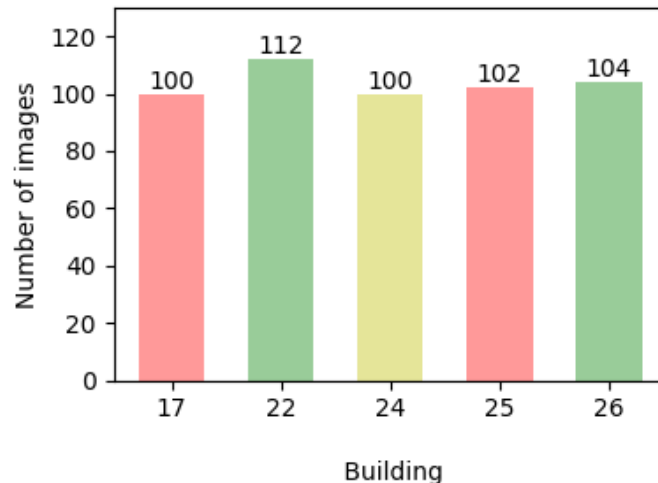


Figure 2 Number of building images of the building dataset

จาก Figure 2 กราฟแสดงจำนวนภาพที่มีอยู่ในอาคารต่างๆ ได้แก่ อาคาร 17, 22, 24, 25, และ 26 โดยอาคาร 17 และอาคาร 24 มี 100 ภาพเท่ากัน อาคาร 22 มีภาพมากที่สุดคือ 112 ภาพ อาคาร 25 และ 26 มีภาพ 102 และ 104 ตามลำดับ มีภาพรวมทั้งสิ้น 518 ภาพ



Figure 3 Example of an image of a building in the Building dataset

จาก Figure 3 แสดงตัวอย่างภาพตามอาคาร 17, 22, 24, 25, และ 26 ของชุดข้อมูลอาคาร แต่ละภาพแสดงมุมมองหรือส่วนต่าง ๆ ของอาคาร เช่น ด้านหน้า ด้านข้าง หรือส่วนของโครงสร้างอาคาร ความยากในการหาค่าความเหมือนของภาพอาคารในชุดข้อมูลนี้พบความท้าทายหลายประการ ได้แก่

- 1) มุมมองภาพที่ต่างกัน เนื่องแต่ละภาพถ่ายจากมุมที่ต่างกัน ทำให้การจับคู่หรือเปรียบเทียบอาคารจากภาพที่ต่างกันยากขึ้น
- 2) แสงและเงา ทำให้สีภาพแสงที่ต่างกันในแต่ละภาพอาจทำให้สีและความชัดเจนของภาพแตกต่างกันไป ซึ่งอาจทำให้การจับคู่ภาพยากขึ้น
- 3) สิ่งแวดล้อมรอบข้าง ทำให้องค์ประกอบรอบ ๆ อาคาร เช่น ต้นไม้ รถยนต์ หรือผู้คน อาจแตกต่างกันในแต่ละภาพ ทำให้การเปรียบเทียบรายละเอียดเฉพาะตัวของอาคารยากขึ้น
- 4) คุณภาพและความละเอียดของภาพ ภาพที่มีความละเอียดต่ำหรือเบลอ อาจทำให้การจับคู่และเปรียบเทียบรายละเอียดของอาคารทำได้ยากขึ้น
- 5) ลักษณะทางสถาปัตยกรรมที่คล้ายกัน หากอาคารหลายหลังมีลักษณะทางสถาปัตยกรรมที่คล้ายกัน เช่น รูปทรง สี หรือวัสดุ เป็นต้น การแยกแยะอาคารแต่ละหลังอาจยากขึ้น
- 6) ความหลากหลายขององค์ประกอบในภาพ องค์ประกอบหลากหลายในภาพ เช่น รูปแบบของหน้าต่าง ประตู หรือรายละเอียดเฉพาะอื่น ๆ อาจทำให้การหาความเหมือนซับซ้อนขึ้น

3.2 การสกัดคุณลักษณะ

งานวิจัยนี้ สกัดคุณลักษณะจากภาพคำถามและภาพในชุดค้นหาโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Networks: CNNs) ได้แก่ VGG16, VGG19, และ NasNetLarge ดัง Figure 4 แสดง

แผนภาพของโมเดล VGG16 Figure 5 แสดงแผนภาพของโมเดล VGG19 และ Figure 6 แสดงแผนภาพของโมเดล NasNetLarge

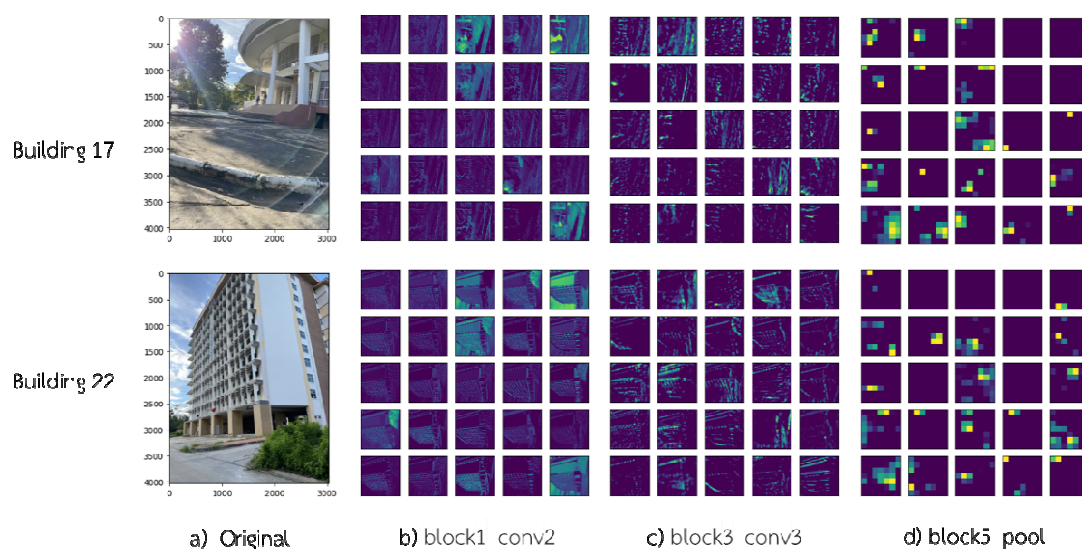


Figure 4 Feature map visualization of convolutional layers (VGG16) from buildings

จาก Figure 4 แสดงการทำงานของโมเดล VGG16 ในการสกัดคุณลักษณะเฉพาะจากภาพ โดยแถวด้านบนเป็นภาพของอาคาร 17 และแถวด้านล่างเป็นภาพของอาคาร 22 โดยมีการแสดงภาพต้นฉบับ (a) และ Feature maps จากเลเยอร์ต่าง ๆ ของโมเดล ได้แก่ เลเยอร์ที่ 2 (block1_conv2) (b), เลเยอร์ที่ 9 (block3_conv3) (c), และเลเยอร์ที่ 18 (block5_pool) (d) ซึ่งเป็นชั้นคอนโวลูตท้ายของ VGG16 โดยในภาพรวมแสดงถึงกระบวนการประมวลผลและการเน้นลักษณะเฉพาะของภาพในแต่ละขั้นตอนอย่างชัดเจน

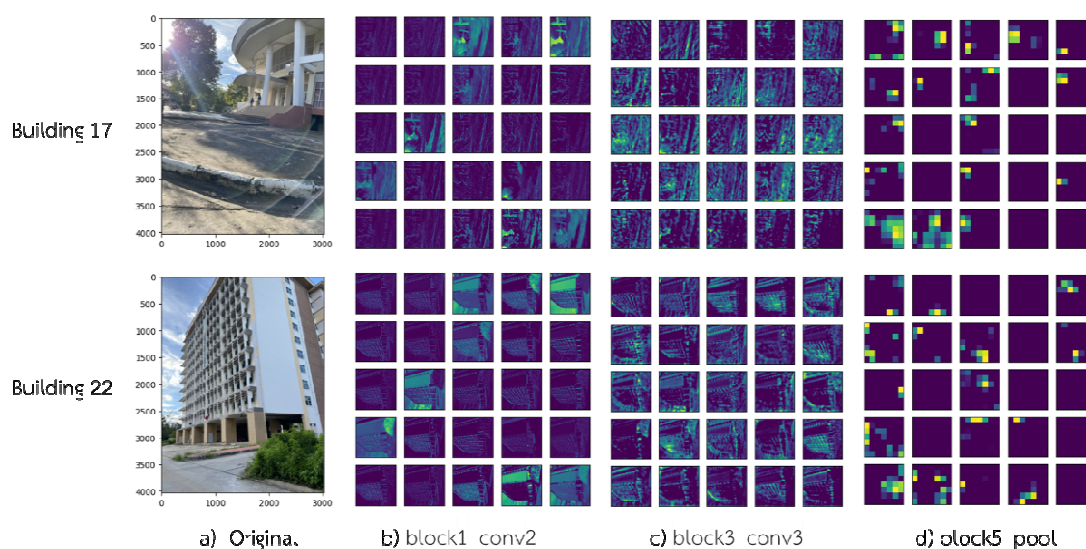


Figure 5 Feature map visualization of convolutional layers (VGG19) from buildings

จาก Figure 5 แสดงการทำงานของโมเดล VGG19 ในการสกัดคุณลักษณะเฉพาะจากภาพ โดยแถวด้านบนเป็นภาพของอาคาร 17 และแถวด้านล่างเป็นภาพของอาคาร 22 โดยมีการแสดงภาพต้นฉบับ (a) และ Feature maps จากเลเยอร์ต่างๆ ของโมเดล ได้แก่ เลเยอร์ที่ 2 (block1_conv2) (b), เลเยอร์ที่ 9 (block3_conv3) (c),

และเลเยอร์ที่ 21 (block5_pool) (d) ซึ่งเป็นชั้นคอนโวลูตท้ายของ VGG19 โดยในภาพรวมแสดงถึงกระบวนการประมวลผลและการเน้นลักษณะเฉพาะของภาพในแต่ละขั้นตอนอย่างชัดเจน

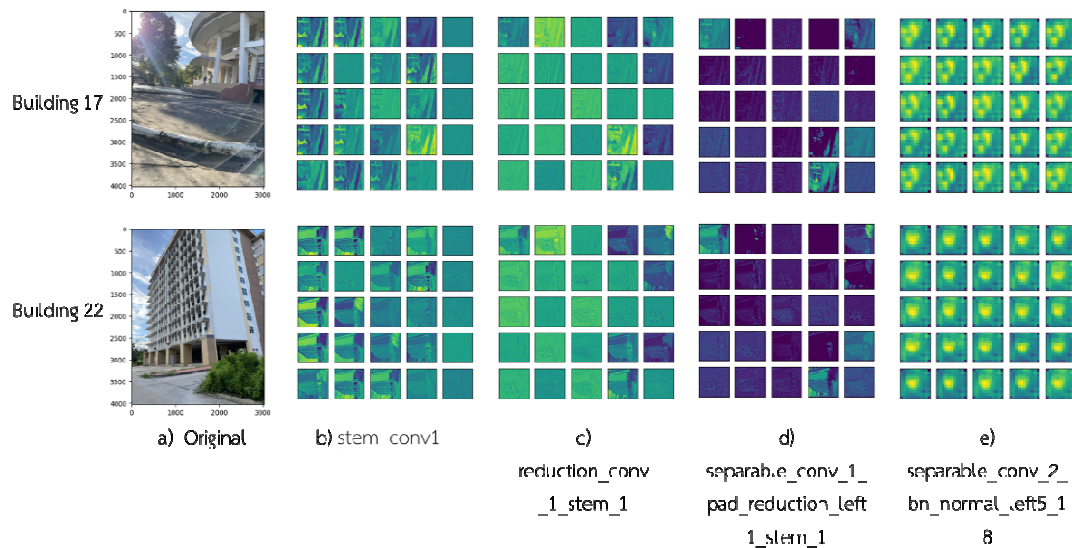


Figure 6 Feature map visualization of convolutional layers (NasNetLarge) from buildings

จาก Figure 6 แสดงการทำงานของโมเดล NasNetLarge ในการสกัดคุณลักษณะเฉพาะจากภาพ โดยแถวด้านบนเป็นภาพของอาคาร 17 และแถวด้านล่างเป็นภาพของอาคาร 22 โดยมีการแสดงภาพต้นฉบับ (a) และ Feature maps จากเลเยอร์ต่างๆ ของโมเดล ได้แก่ เลเยอร์ที่ 1 (stem_conv1) (b), เลเยอร์ที่ 4 (reduction_conv_1_stem_1) (c), เลเยอร์ที่ 8 (separable_conv_1_pad_reduction_left1_stem_1) (d), และ (e) เลเยอร์ที่ 1031 (separable_conv_2_bn_normal_left5_18) ซึ่งเป็นชั้นคอนโวลูตท้ายของ NasNetLarge โดยในภาพรวมแสดงถึงกระบวนการประมวลผลและการเน้นลักษณะเฉพาะของภาพในแต่ละขั้นตอนอย่างชัดเจน

การเปรียบเทียบภาพจากโมเดล CNNs สามแบบ ได้แก่ VGG16 (Figure 4), VGG19 (Figure 5) และ NasNetLarge (Figure 6) ในการสกัดคุณลักษณะเฉพาะจากภาพอาคาร 17 และ 22 พบว่า VGG16 และ VGG19 มีโครงสร้างคล้ายกัน แต่ VGG19 มีเลเยอร์ที่ลึกกว่า ทำให้การดึงลักษณะเฉพาะละเอียดขึ้น ส่วน NasNetLarge มีโครงสร้างที่แตกต่างออกไปและดึงลักษณะเฉพาะได้หลากหลายกว่า จากตัวอย่างแผนภาพที่สกัดคุณลักษณะจากสามโมเดล แสดงให้เห็นว่าแต่ละโมเดลมีความสามารถในการดึงลักษณะเฉพาะที่ต่างกัน ซึ่งมีผลต่อการนำไปใช้งานในการจำแนกหรือค้นหาความเหมือนของภาพ โดยขึ้นอยู่กับความต้องการด้านความแม่นยำและประสิทธิภาพในการประมวลผล

3.3 การวัดความคล้ายคลึง

ชุดข้อมูลอาคารมีความซับซ้อนและความหลากหลายของภาพในภาพชุดค้นหา (Image search set) ที่ทำให้การจับคู่หรือเปรียบเทียบอาคารจากภาพที่แตกต่างกันยากขึ้น ความแตกต่างของภาพที่เกิดจากหลายปัจจัย ได้แก่ มุมมองการถ่ายภาพที่แตกต่างกัน ปัญหาด้านแสงและเงา สิ่งแวดล้อมรอบข้างหรือองค์ประกอบรอบๆ อาคารแตกต่างกันในแต่ละภาพ คุณภาพและความละเอียดของภาพ ภาพที่มีความละเอียดต่ำหรือเบลอ ลักษณะทางสถาปัตยกรรมที่คล้ายกัน และความหลากหลายขององค์ประกอบในภาพหรือรายละเอียดเฉพาะอื่นๆ อาจทำให้การหาความเหมือนซับซ้อนขึ้น ดังนั้น เพื่อแสดงให้เห็นถึงประสิทธิภาพในการวัดความคล้ายคลึงด้วย Cosine similarity ได้ทำการวัดความคล้ายคลึงจากภาพคำถาม (Query Image) จากการถ่ายด้วยกล้องโทรศัพท์มือถือเพิ่มจำนวน 12 ภาพ ผลการทดลองดัง Figure 7

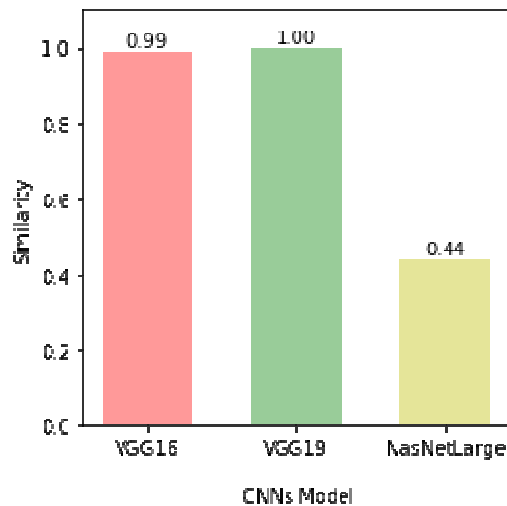


Figure 7 Similarity of VGG16, VGG 19, and NasNetLarge

จาก Figure 7 กราฟแสดงค่าความคล้ายคลึงของโมเดล VGG16, VGG19 และ NasNetLarge โดยโมเดล VGG19 มีค่าความคล้ายคลึงสูงสุดที่ 1.00 รองลงมาคือ VGG16 ที่ 0.99 และ NasNetLarge มีค่าต่ำสุดที่ 0.44 สรุปได้ว่า VGG19 มีความสามารถในการคาดการณ์ที่แม่นยำที่สุดในสามโมเดลนี้ ผลการทดลองแสดงให้เห็นว่า VGG กับ 16 VGG ให้ 19 ค่าความคล้ายคลึงใกล้เคียงกัน แต่ค่าความคล้ายคลึงของ NasNetLarge ต่ำกว่าทั้ง 2 โมเดล โดยค้นคืนรูปภาพผิดจำนวน 2 ภาพ เมื่อพิจารณาภาพคำถามปัจจัยที่ทำให้เกิดการค้นคืนรูปภาพที่ผิดพลาด เช่น พื้นผิวของคอนกรีต รูปทรงและสีของต้นไม้มีความใกล้เคียงกับอาคารอื่น ตัวอย่างภาพคำถามและผลลัพธ์จากการค้นคืนภาพของ VGG16, VGG19 และ NasNetLarge ดัง Figure 8, 9, และ 10 ตามลำดับ

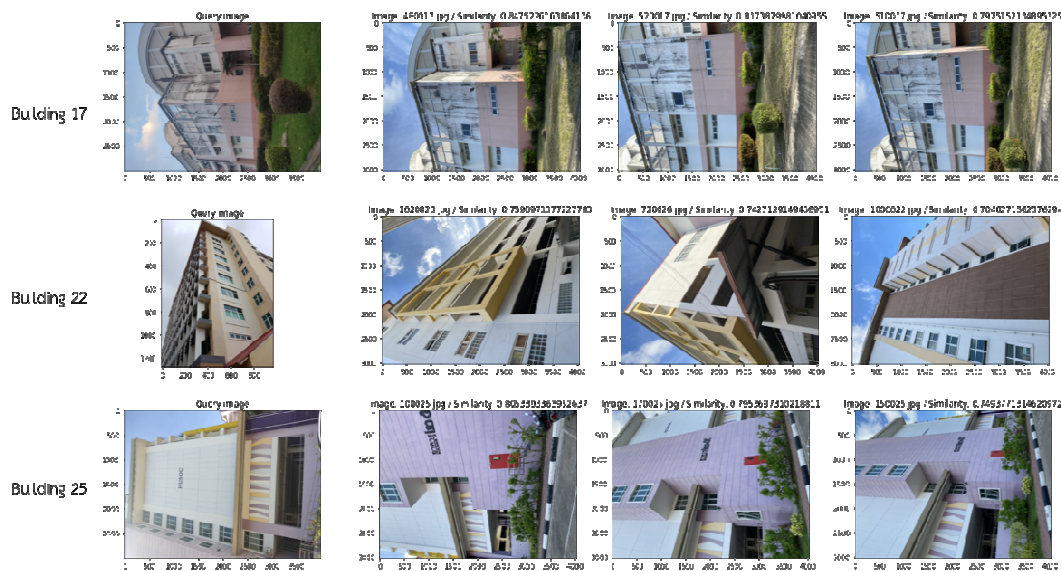


Figure 8 Query images and image retrieval results of VGG16

จาก Figure 8 แสดงตัวอย่างของภาพคำถามและผลการค้นคืนภาพโดยใช้โมเดล VGG16 ของอาคาร 17, 22, และ 25 โดยภาพในคอลัมน์แรกเป็นภาพคำถามและภาพในคอลัมน์ที่ 2 – 4 เป็นผลการค้นคืนภาพตามลำดับ จากค่าความคล้ายคลึงมากไปน้อย ซึ่งแสดงให้เห็นถึงความสามารถของโมเดล VGG16 ในการดึงข้อมูลภาพที่มีลักษณะใกล้เคียงกับภาพคำถามได้อย่างมีประสิทธิภาพ โดยผลการค้นคืนภาพส่วนใหญ่มีความคล้ายคลึงกับภาพคำถามทั้งในด้านรูปทรงและมุมมองของอาคาร

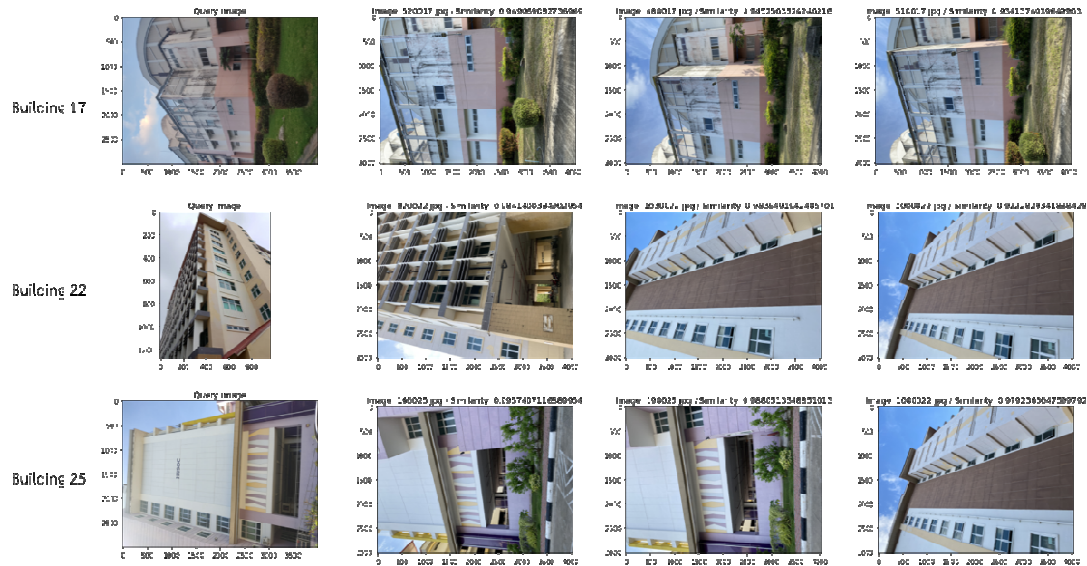


Figure 9 Query images and image retrieval results of VGG19

จาก Figure 9 แสดงตัวอย่างของภาพคำถามและผลการค้นคืนภาพโดยใช้โมเดล VGG19 ของอาคาร 17, 22, และ 25 โดยภาพในคอลัมน์แรกเป็นภาพคำถามและภาพในคอลัมน์ที่ 2 – 4 เป็นผลการค้นคืนภาพตามลำดับจากค่าความคล้ายคลึงมากไปน้อย โดยผลลัพธ์แสดงให้เห็นว่าโมเดล VGG19 สามารถดึงข้อมูลภาพที่มีความคล้ายคลึงกับภาพคำถามได้อย่างแม่นยำ ภาพที่ถูกค้นคืนมีลักษณะและมุมมองที่ใกล้เคียงกับภาพคำถาม แสดงถึงความสามารถของโมเดลในการประมวลผลและการจับคู่ภาพได้อย่างมีประสิทธิภาพ

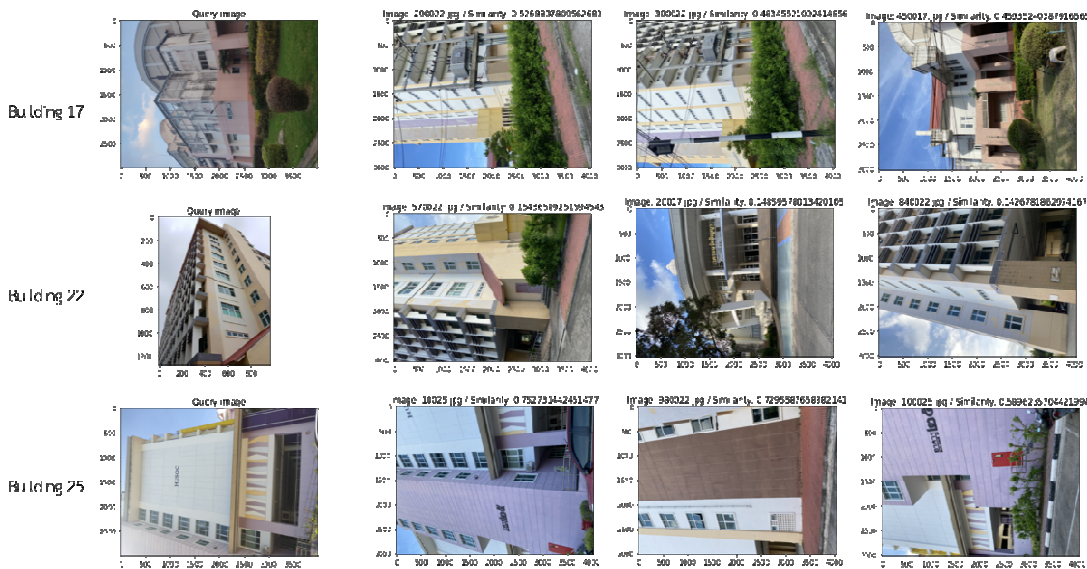


Figure 10 Query images and image retrieval results of NasNetLarge

จาก Figure 10 แสดงตัวอย่างของภาพคำถามและผลการค้นคืนภาพโดยใช้โมเดล VGG19 ของอาคาร 17, 22, และ 25 โดยภาพในคอลัมน์แรกเป็นภาพคำถามและภาพในคอลัมน์ที่ 2 – 4 เป็นผลการค้นคืนภาพตามลำดับจากค่าความคล้ายคลึงมากไปน้อย ผลลัพธ์แสดงให้เห็นว่าโมเดล NasNetLarge สามารถค้นคืนภาพที่มีความคล้ายคลึงกับภาพคำถามได้ในระดับหนึ่ง แต่ผลการค้นคืนบางภาพมีลักษณะและมุมมองที่ไม่ตรงกับภาพคำถาม (Query Image) แสดงถึงข้อจำกัดของโมเดลในด้านการประมวลผลและการจับคู่ภาพเมื่อเทียบกับ VGG16 และ VGG19

จากการเปรียบเทียบผลการค้นคืนภาพของอาคาร 17, 22, และ 25 ด้วยโมเดล VGG16, VGG19 และ NasNetLarge พบว่าโมเดล VGG19 ให้ผลลัพธ์ที่มีความคล้ายคลึงกับภาพคำถามมากที่สุด ตามมาด้วย VGG16 ที่สามารถดึงภาพที่มีลักษณะใกล้เคียงได้อย่างมีประสิทธิภาพ ส่วนโมเดล NasNetLarge ให้ผลลัพธ์ที่มีความคล้ายคลึงในระดับปานกลางถึงต่ำ โดยมีบางภาพที่มึ่มมองและลักษณะไม่ตรงกับภาพคำถาม แสดงให้เห็นว่า VGG19 และ VGG16 มีความสามารถในการค้นคืนภาพที่ดีกว่า NasNetLarge

4. สรุปผลการวิจัย

งานวิจัยนี้มุ่งเน้นการสกัดคุณลักษณะของภาพโดยใช้ CNNs สามโมเดล ได้แก่ VGG16, VGG19 และ NasNetLarge เพื่อนำไปประยุกต์ใช้ในระบบ Content-Based Image Retrieval (CBIR) สำหรับการค้นคืนภาพอาคาร โดยใช้ภาพถ่ายอาคารต่าง ๆ ภายในมหาวิทยาลัยราชภัฏบุรีรัมย์เป็นชุดข้อมูล การทดลองได้ทำการสกัดคุณลักษณะจากภาพถ่าย และประเมินความคล้ายคลึงกันของภาพโดยใช้ Cosine similarity พบว่า VGG19 ให้ผลลัพธ์ที่ดีที่สุด รองลงมา คือ VGG16 และ NasNetLarge ตามลำดับ VGG19 มีประสิทธิภาพสูงสุดในการสกัดคุณลักษณะภาพอาคารเนื่องจาก VGG19 มีโครงสร้างที่ลึกกว่า VGG16 ทำให้สามารถดึงคุณลักษณะภาพได้ละเอียดมากขึ้น อย่างไรก็ตาม NasNetLarge ซึ่งเป็นโมเดลที่มีความซับซ้อนสูง กลับมีประสิทธิภาพต่ำกว่า VGG16 และ VGG19 ซึ่งอาจเกิดจากการที่ NasNetLarge ต้องการข้อมูลจำนวนมากในการฝึกฝน หรือโครงสร้างของโมเดลอาจไม่เหมาะสมกับลักษณะของภาพอาคาร การวิจัยนี้มีประโยชน์ในการทำความเข้าใจถึงประสิทธิภาพของ CNNs ในการสกัดคุณลักษณะภาพ และสามารถนำไปประยุกต์ใช้ในการพัฒนาระบบ CBIR สำหรับการค้นคืนภาพอาคารได้ อย่างไรก็ตาม การวิจัยนี้ยังมีข้อจำกัดในเรื่องขนาดของชุดข้อมูล ซึ่งควรได้รับการแก้ไขในงานวิจัยในอนาคต

5. กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยราชภัฏบุรีรัมย์ เพื่อนร่วมงานทุกคนที่อำนวยความสะดวกและช่วยเหลือในการทำวิจัยครั้งนี้

ผลประโยชน์ทับซ้อน

ผู้เขียนขอยืนยันว่างานวิจัยนี้ไม่มีความขัดแย้งทางผลประโยชน์ที่เกี่ยวข้องกับบทความนี้กับบทความอื่นๆ

REFERENCES

- Abdullah, M. N., & Sahari, M. A. (2022, November). Digital image clustering and colour model selection in content-based image retrieval (CBIR) approach for biometric security image. In *AIP Conference Proceedings* (Vol. 2617, No. 1). AIP Publishing.
- Aung, S. P. P., Pa, W. P., & Nwe, T. L. (2023). Improving Myanmar Image Caption Generation Using NASNetLarge and Bi-directional LSTM. In *2023 IEEE Conference on Computer Applications (ICCA)* (pp. 1–6). IEEE. doi: 10.1109/ICCA51723.2023.10181326
- Celeghin, A., Borriero, A., Orsenigo, D., Diano, M., Méndez Guerrero, C. A., Perotti, A., ... Tamietto, M. (2023). Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues. *Frontiers in Computational Neuroscience*, 17, 1–13. doi: 10.3389/fncom. 2023.1153572
- Chen, C. (2023, May). Image retrieval based on multimodality neural network and local sensitive hash. In *Eighth International Symposium on Advances in Electrical, Electronics, and Computer Engineering (ISAEECE 2023)* (Vol. 12704, pp. 542–545). SPIE. doi: 10.1117/12.2680159

- Chen, X. (2023). The Study for Convolutional Neural Network and Corresponding Applications. *Theoretical and Natural Science*, 5(1), 182–187. doi: 10.54254/2753-8818/5/ 20230387
- Gayathri, P., Dhavileswarapu, A., Ibrahim, S., Paul, R., & Gupta, R. (2023). Exploring the Potential of VGG-16 Architecture for Accurate Brain Tumor Detection Using Deep Learning. *Journal of Computers, Mechanical and Management*, 2(2), 23056-23056. doi: 10.57159/gadl.jcmm.2.2.23056
- Himel, G. M. S., & Islam, M. M. (2024). Benchmark Analysis of Various Pre-trained Deep Learning Models on ASSIRA Cats and Dogs Dataset. *arXiv e-prints*, arXiv-2401.
- Jiang, C., & Goldsztein, G. (2023). Convolutional Neural Network Approach to Classifying the CIFAR-10 Dataset. *Journal of Student Research*, 12(2), 1-7. doi: 10.47611/jsrhs.v12i2.4388
- Karrach, L., & Pivarčiová, E. (2023). Using a Convolutional Neural Network for Machine Written Character Recognition. *TEM Journal*, 12(3), 1252-1259.
- Lazim Lafta, A., & I. Abdulsada, A. (2022). SMPP-CBIR: shorted and mixed aggregated image features for privacy-preserving content-based image retrieval. *Bulletin of Electrical Engineering and Informatics*, 11(5), 2930–2937. doi: 10.11591/eei.v11i5.4323
- Liu, Q., Zhang, J., Liu, J., & Yang, Z. (2022). Feature extraction and classification algorithm, which one is more essential? An experimental study on a specific task of vibration signal diagnosis. *International Journal of Machine Learning and Cybernetics*, 13(6), 1685–1696.
- Malik, C. K. M. (2022). Content based Image Retrieval Using Clustering Method. *International Academic Journal of Science and Engineering*, 6(2), 06–12. doi: 10.9756/IAJSE/V6I2/1910020
- Malleswari, G., & Reddy, A. S. (2023). Diverse Convolutional Neural Network Models for Feature Extraction from Brain Tumor Images. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 405–410). India: IEEE. doi:10.1109/ICCMC56507.2023.10083973
- Mathews, A., N, S., & R, V. K. (2022). Analysis of Content Based Image Retrieval using Deep Feature Extraction and Similarity Matching. *International Journal of Advanced Computer Science and Applications*, 13(12). 646-655. doi: 10.14569/IJACSA.2022.0131277
- Park, K., Hong, J. S., & Kim, W. (2020). A Methodology Combining Cosine Similarity with Classifier for Text Classification. *Applied Artificial Intelligence*, 34(5), 396–411. doi: 10.1080/08839514.2020.1723868
- Pratibha Rashmi, & Manu Pratap Singh. (2023). Convolution neural networks with hybrid feature extraction methods for classification of voice sound signals. *World Journal of Advanced Engineering Technology and Sciences*, 8(2), 110–125. doi: 10.30574/wjaets.2023.8.2.0083
- Rani, L. N., & Yuhandri, Y. (2023). Similarity Measurement on Logo Image Using CBIR (Content Base Image Retrieval) and CNN ResNet-18 Architecture. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)* (pp. 228–233). Indonesia: IEEE. doi: 10.1109/ICCoSITE57641.2023.10127711

- Rashmi, P., & Singh, M. P. (2023). Convolution neural networks with hybrid feature extraction methods for classification of voice sound signals. *World Journal of Advanced Engineering Technology and Sciences*, 8(2), 110–125. doi: 10.30574/wjaets.2023.8.2.0083
- Sharma, P., Kumar, R., & Gupta, M. (2023). Road Features Extraction Using Convolutional Neural Network. In *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (pp. 881–886). India: IEEE. doi:10.1109/InCACCT57535.2023.10141712
- Shi, J., Su, Z., Qin, H., Shen, C., Huang, W., & Zhu, Z. (2022). Generalized Variable-Step Multiscale Lempel-Ziv Complexity: A Feature Extraction Tool for Bearing Fault Diagnosis. *IEEE Sensors Journal*, 22(15), 15296–15305. doi: 10.1109/JSEN.2022.3187763
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv e-prints*, arXiv:1409.1556.
- Suhaidi, M., Abdul Kadir, R., & Tiun, S. (2021). A REVIEW OF FEATURE EXTRACTION METHODS ON MACHINE LEARNING. *Journal of Information System and Technology Management*, 6(22), 51–59. doi: 10.35631/JISTM.622005
- Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation*, 11(3), 52. doi: 10.3390/computation11030052
- Usha, K., Sudeepthi, B., Mahathi, D., & Shravya, P. (2021). Face Mask Detection Using Feature Extraction. In *Data Engineering and Communication Technology: Proceedings of ICDECT 2020* (pp. 437-447). Springer Singapore.
- Wei, D. (2022). An Online Extraction Algorithm for Image Feature Information Based on Convolutional Neural Network. *Mobile Information Systems*, 2022, 1–11. doi: 10.1155/2022/1606438
- Xu, X., & Zheng, L. (2022, September). Multivariate feature extraction. In *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 1-8). USA: IEEE. doi: 10.1109/Allerton49937.2022.9929401
- Zhang, L., Luo, Y., Cao, X., Shen, H., & Wang, T. (2023). Frequency and Scale Perspectives of Feature Extraction. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1–5). USA: IEEE.
- Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8697-8710). USA: IEEE.