การศึกษาเปรียบเทียบระหว่างวิธีการหาคุณลักษณะเฉพาะพื้นที่และวิธีการเรียนรู้เชิงลึกสำหรับการค้นคืนรูปภาพลายผ้าไหม

Comparative Study Between Local Descriptors and Deep Learning for Silk Pattern Image Retrieval

นัธทวัฒน์ รักสะอาด[1], โอพาริก สุรินต๊ะ[2]

Nattawat Raksaard[1], Olarik Surinta[2]

Received : 18 April 2018 ; Accepted : 11 June 2018

บทคัดย่อ

งานวิจัยฉบับนี้มีจุดประสงค์เพื่อการศึกษาเปรียบเทียบระหว่างวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ และโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) สำหรับการค้นคืนรูปภาพลายผ้าไหมไทย วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ถูกนำมาเพื่อเปรียบเทียบในการสร้างข้อมูลลักษณะพิเศษ ประกอบด้วย วิธี Histogram of Oriented Gradients และวิธี Scale-Invariant Feature Transform ดังนั้น ข้อมูลลักษณะพิเศษจะถูกส่งไปเพื่อคำนวณร่วมกับวิธี K-Nearest Neighbor (KNN) และวิธี Support Vector Machine นอกไปจากนั้น งานวิจัยฉบับนี้ได้ทำการปรับปรุงโครงสร้างของวิธี CNN ซึ่งประกอบด้วยโครงสร้างแบบ LeNet-5 และ AlexNet โดยโครงสร้างแบบ LeNet-5 ได้ปรับปรุงโครงสร้างด้วยการเพิ่มจำนวนของโหนดในชั้นเชื่อมโยงสมบูรณ์ และปรับปรุงโครงสร้างของ AlexNet ปรับปรุงโดยลดขนาดของโหนดในชั้นเชื่อมโยงสมบูรณ์ สุดท้ายแล้ว ประเมินประสิทธิภาพระหว่างวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ วิธี CNNs และวิธี CNNs ที่ได้ปรับปรุงโครงสร้างใหม่ จากการทดลองพบว่า วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่เมื่อนำไปคำนวณร่วมกับวิธี KNN มีประสิทธิภาพสูงกว่าวิธี CNN

คำสำคัญ: วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ขั้นตอนวิธีการคำนวณเพื่อนบ้านใกล้ที่สุด k ตำแหน่ง ซัพพอร์ตเว็กเตอร์แมชชีน การเรียนรู้เชิงลึก โครงข่ายประสาทเทียมแบบคอนโวลูชัน

Abstract

This paper aims to do a comparative study of local feature descriptor techniques and convolutional neural networks (CNN) for retrieving Thai silk pattern images. Two feature descriptor techniques, the histogram of oriented gradients and the scale-invariant feature transform, are compared to extract feature vectors from the silk pattern images. We combined the feature vectors extracted from feature descriptor techniques with k-nearest neighbors (KNN) and support vector machine. Then we modified CNN architectures: LeNet-5 and AlexNet. The LeNet-5 was modified by increasing the number of neurons in each layer of the fully connected layers. The AlexNet architecture was modified by reducing the neurons in each layer of the fully connected layers. Finally, we evaluated the local descriptor techniques, the existing CNN architectures and our modified CNN architectures on Thai silk pattern dataset. The results of the study showed that the local descriptor techniques combined with KNN algorithm significantly outperform the CNN methods.

Keywords: local descriptor technique, k-nearest neighbors algorithm, support vector machine, deep learning, convolutional neural networks

---

[1] *นิสิตปริญญาโท, ห้องปฏิบัติการมัลติเอเจนท์ ระบบอัจฉริยะ และการจำลองสถานการณ์, สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150*

[2] *อาจารย์, ห้องปฏิบัติการมัลติเอเจนท์ ระบบอัจฉริยะ และการจำลองสถานการณ์, คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150*

[1] *Master Student, Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology, Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand. E-mail: natthawat.rak@msu.ac.th*

[2] *Lecturer, Multi-agent Intelligent Simulation Laboratory (MISL), Department of Information Technology, Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand. E-mail: olarik.s@msu.ac.th*

บทนำ

การทอผ้าไหมสะท้อนให้เห็นถึงวิถีชีวิตความเป็นอยู่ของคนสมัยก่อนจนถึงปัจจุบัน โดยรูปแบบการทอผ้านั้นสะท้อนให้เห็นถึงลักษณะเด่นของลวดลายของแต่ละท้องถิ่น จึงทำให้ลายผ้าไหมที่ทอนั้นมีเอกลักษณ์ และมีคุณค่า แต่เนื่องด้วยลายผ้าไหมที่ทอนั้นมีมากมายหลายสิบชื่อ เช่นลายสร้อยดอกหมาก ลายประตูทอง และลายนกยูง เป็นต้น อีกทั้งบางลวดลายยังมีความใกล้เคียงกัน จึงทำให้ผู้วิจัยมีความสนใจในการศึกษาเพื่อการวิธีการการค้นคืนรูปภาพลายผ้าไหม เพื่อให้การค้นคืนรูปภาพมีความถูกต้องสูงที่สุด

การค้นคืนรูปภาพโดยใช้คอนเทนท์ (Content-Based Image Retrieval: CBIR) โดยทั่วไปแล้วสามารถคำนวณได้จากลักษณะเฉพาะของรูปภาพ เช่น สี รูปร่าง เส้นขอบ และพื้นผิว เป็นต้น[1] วิธีที่สามารถนำมาใช้ในการหาคุณลักษณะพิเศษ ได้แก่ Scale-Invariant Feature Transform (SIFT), Histograms of Oriented Gradients (HOG), Local Binary Pattern (LBP) และ Bag of Visual Words (BOW)[2-5] เป็นต้น ซึ่งกล่าวได้ว่าเป็นวิธีการหาคุณลักษณะพิเศษแบบเฉพาะพื้นที่ (Local Descriptor) เพื่อใช้เป็นตัวแทนของรูปภาพในระดับล่าง (Low-level Feature)[6] ซึ่ง Local Descriptor สามารถนำไปใช้ในงานวิจัยด้านอื่น เช่น การรู้จำใบหน้า (Face Recognition) และการค้นหาวัตถุ (Objection Detection) เป็นต้น

การค้นคืนรูปภาพสามารถทำได้โดยเปรียบเทียบหาค่าความคล้ายคลึง (Similarity Measure) ระหว่างคุณลักษณะพิเศษ (Feature Extraction) ของรูปภาพที่ต้องการค้นคืน (Query Image) และคุณลักษณะพิเศษของรูปภาพที่อยู่ในฐานข้อมูล รูปภาพที่มีค่าความคล้ายคลึงสูง (High Similarity Score) จะเป็นรูปภาพที่มีความคล้ายคลึงกับ Query Image มากที่สุด ในการแสดงผลลัพธ์ระบบ CBIR จะจัดเรียงลำดับ (Ranking) รูปภาพที่ค้นคืน (Retrieve Image) ตามค่าความคล้ายคลึง หรือเรียกว่า Top N โดยที่ N คือจำนวนของรูปภาพที่ค้นคัน[7]

ในปัจจุบัน การเรียนรู้เชิงลึก (Deep Learning) เป็นวิธีที่ได้รับความนิยม และสามารถนำไปใช้กับงานวิจัยได้หลายประเภท เช่น การจำแนกประเภท (Classification) และการจัดกลุ่ม (Clustering) เป็นต้น อีกทั้งยังมีงานวิจัยที่ใช้วิธีการเรียนรู้เชิงลึก[6, 7] กับงานวิจัยทางด้าน CBIR ซึ่งเรียกว่าเป็นวิธีการหาคุณลักษณะพิเศษระดับสูง (High-level Feature) วิธีการเรียนรู้เชิงลึก ที่ถูกนำไปใช้อย่างแพร่หลาย ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ซึ่งสามารถกำหนดโครงสร้าง (Architecture) ได้ตามไม่จำกัด เช่น โครงสร้างแบบ LeNet-5 และ AlexNet [8, 9] ที่ถูกออกแบบให้โครงสร้างมีจำนวน 5 และ 8 Layer ตามลำดับ เป็นต้น

งานวิจัยฉบับนี้ได้มุ่งเน้นศึกษาเกี่ยวกับวิธีการค้นคืนรูปภาพด้วยวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ร่วมกับวิธีการเรียนรู้เครื่องจักร (Machine Learning) และวิธีการเรียนรู้เชิงลึก

งานวิจัยของ Karaaba et al.[10] ได้นำเสนอวิธีการระบุใบหน้า (Face Identification) ที่รูปภาพใบหน้ามีจำนวนจำกัด (Small Sample Sizes) โดยคำนวณหาคุณลักษณะพิเศษด้วยวิธี Bag of Words (BOW) ร่วมกับวิธี Histogram of Oriented Gradients (HOG) ซึ่งเรียกว่า HOG-BOW เพื่อใช้สำหรับการเรียนรู้ข้อมูลที่มีจำนวนจำกัด เนื่องจากชุดข้อมูล FERET (Face Recognition Technology) และ LFW (Labeled Faces in the Wild) มีจำนวนใบหน้าในแต่ละกลุ่มจำนวนจำกัด เช่น บางบุคคลมีตัวอย่างใบหน้าเพียง 3 รูปภาพ เป็นต้น และใช้วิธีการเรียนรู้ด้วย L2 Support Vector Machine (L2-SVM) เพื่อใช้สร้างโมเดลของใบหน้า ในงานวิจัยยังได้เปรียบเทียบวิธี HOG-BOW กับวิธีอื่น เช่น HOG, Scale Invariant Feature Transform (SIFT), Multi-Subregion based Correlation Filter Bank (MS-CFB), Discriminative Multi-Manifold Analysis (DMMA) จากการทดลองพบว่าวิธี HOG-BOW ให้อัตราการรู้จำใบหน้าสูงที่สุด โดยทดสอบกับข้อมูลชุด FERET มีอัตราการรู้จำใบหน้าที่ 92.62% และข้อมูลชุด LFW มีอัตราการรู้จำใบหน้าที่ 48.92%

งานวิจัยของ Ahonen et al.[11] นำเสนอการรู้จำใบหน้า (Face Recognition) โดยพิจารณาจาก รูปร่าง และพื้นผิว (Texture) โดยรูปภาพจะถูกแบ่งพื้นที่ออกเป็นส่วนย่อย (Small Region) ที่มีขนาดเท่ากัน จากนั้นส่วนย่อยนั้นจะถูกนำไปคำนวณด้วยวิธี Local Binary Pattern (LBP) และจะถูกใช้เพื่อเป็นตัวแทนของใบหน้า ซึ่งการรู้จำใช้วิธี Nearest Neighbor (NN) ใช้วิธี Chi Square ในการคำนวณ และนำไปทดสอบกับชุดข้อมูล FERET ซึ่งใช้ชุดข้อมูลย่อย ประกอบด้วยชุดข้อมูล fb และ fc จากการทดลองสรุปได้ว่าเมื่อใช้วิธี LBP ร่วมกับ NN กับข้อมูลชุด fb มีความถูกต้อง 97% และชุดข้อมูล fc มีความถูกต้อง 79%

สำหรับการค้นคืนรูปภาพโดยใช้คอนเทนท์ (Content-Based Image Retrieval: CBIR) งานวิจัย[12] ได้นำเสนอวิธีการใช้ค่าฮิสโตแกรมของค่าสี (Color Histogram) ที่มีขนาด 256 ซึ่งคือค่าสีแบบ RGB ที่แต่ละพิกเซลมีค่าความสว่าง (Intensity) ตั้งแต่ 0-255 มาทำการเปรียบเทียบ ดังนั้นรูปภาพที่ต้องการค้นคืน และรูปภาพจากฐานข้อมูล จะถูกนำมาเปรียบเทียบโดยใช้ค่าความคล้ายคลึง (Similarity Function)

เป็นค่าที่ใช้เพื่อกำหนดความคล้ายคลึงระหว่างรูปภาพ โดยงานวิจัย[13] ได้นำเสนอวิธีการพิจารณาน้ำหนักการกระจายของสีด้วยการกระจายตัวแบบเกาส์เซียน (Gaussian Distribution) โดยใช้แบบจำลองสี HSV เพื่อใช้สำหรับการค้นคืนรูปภาพ ข้อมูลที่ใช้ในการทดลองได้มาจาก Corel Stock Photo Gallery และดาวน์โหลดจากอินเทอร์เน็ต จำนวนทั้งสิ้น 10,297 รูปภาพ ซึ่งฮิสโตแกรมสี (Color Histogram) และฮิสโตแกรมสีข้างเคียงถูกนำไปเปรียบเทียบความแตกต่างของสีโดยคำนวณจากการกระจาย น้ำหนักแบบเกาส์เซียน และนำไปคำนวณเพื่อหาค่าระยะห่างของฮิสโตแกรม (Distance Histogram) โดยใช้วิธีการหาค่าเฉลี่ยของการปรับปรุงตำแหน่งของการค้นคืนให้อยู่ในช่วงปกติ (Average Normalized Modified Retrieval Rank: ANMRR) เป็นเครื่องมือชี้วัดประสิทธิภาพ โดยพิจารณาสีข้างเคียงจากสีหลักขนาด 7 สี ให้ประสิทธิภาพที่ดีที่สุดเท่ากับ 0.452

Hazra et al.[14] นำเสนอวิธีการหาคุณลักษณะพิเศษโดยใช้วิธี Wavelet Moment และ Gabor Filter เพื่อเข้ารหัสรูปภาพ โดยรูปภาพสีจะถูกแบ่งออกเป็น 3 ช่อง (Channel) ตามค่าสีแบบ RGB โดยในแต่ละช่อง จะถูกแบ่งออกเป็นบล็อกที่มีขนาดเล็กเพื่อใช้สำหรับนำไปคำนวณ คุณลักษณะพิเศษที่ได้จะถูกนำไปเรียนรู้ด้วยวิธีการเรียนรู้เครื่องจักร ได้แก่วิธี K-Nearest Neighbor (KNN) และวิธี SVM เพื่อทำการค้นคืนรูปภาพที่มีความใกล้เคียง จากนั้นนำรูปภาพที่ค้นคืนได้มาคำนวณหาค่าความคล้ายคลึงระหว่างภาพที่ค้นคืน และรูปภาพที่นำไปเปรียบเทียบ (Query Image) ผลลัพธ์ที่ได้ก็คือค่าสัมประสิทธิ์ของค่าความคล้ายคลึง โดยวิธี SVM มีประสิทธิภาพสูงกว่าวิธี KNN โดยมีความถูกต้องมากกว่า 80%

Singh[7] นำวิธี CNN ที่ใช้โครงสร้าง LeNet-5 มาใช้ในงานด้าน CBIR เพื่อจำแนกประเภทข้อมูลที่อยู่ในชุดข้อมูลย่อยของ SUN ซึ่งมีรูปภาพจำนวน 3,000 รูปภาพ ที่ประกอบด้วย 8 Class ได้แก่ น้ำ รถ ภูเขา พื้นดิน ต้นไม้ ตึก หิมะ และท้องฟ้า ในการทดลองได้แบ่งข้อมูลออกเป็น 80% สำหรับข้อมูลชุดเรียนรู้ และ 20% สำหรับข้อมูลชุดทดสอบ ข้อมูลรูปภาพที่ใช้ในการทดลองถูกแปลงให้เป็นสีเทา และเปลี่ยนขนาดเป็น 28x28 พิกเซล จากการทดลองพบว่าวิธี CNN มีความผิดพลาด (Error Rate) ในการจำแนกประเภทข้อมูล 27.97% ซึ่งน้อยกว่าวิธี Bag of Words ที่มีความผิดพลาดสูงถึง 47.44%

งานวิจัยฉบับนี้ ได้ทำการเปรียบเทียบวิธีการค้นคืนรูปภาพด้วยวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ประกอบด้วยวิธี SIFT และ HOG ซึ่งเป็น Low-level Feature ร่วมกับวิธี SVM และการหาค่าระยะห่างแบบ Euclidean (Euclidean Distance) และวิธีการเรียนรู้เชิงลึก แบบ CNN โดยใช้โครงสร้าง LeNet-5 และ AlexNet ซึ่งเป็นการหาคุณลักษณะพิเศษระดับสูง เพื่อทดสอบกับข้อมูลชุดผ้าไหมไทย (Thai Silk Pattern Dataset) ทั้งสิ้น 10 ลาย จำนวน 300 รูปภาพ ที่เก็บอยู่ในรูปแบบภาพสี (Color Image)

วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ (Local Descriptor Technique)

วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ที่ใช้ในงานวิจัยประกอบด้วยวิธี Histogram of Oriented Gradients (HOG) และวิธี Scale-Invariant Features Transform (SIFT)

Histogram of Oriented Gradients (HOG)

วิธี HOG ถูกนำเสนอในงานวิจัย[3] การคำนวณหาคุณลักษณะพิเศษด้วยวิธีนี้ รูปภาพจะถูกแปลงให้เป็นสีเทา และนำไป Convolution กับเคอร์นอล (kernel) เพื่อทำการหาภาพขอบ (Edge Detection)[15] สามารถคำนวณกับ Kernel แบบง่าย เช่น[17] จากนั้นจึงแบ่งรูปภาพออกเป็นพื้นที่ย่อย หรือเรียกว่าบล็อก เพื่อนำพื้นที่ย่อยไปคำนวณหา Gradient ในแนวแกน และ ดังสมการ (1) และ (2)

$$G_x = I(x+1,y) - I(x-1,y) \quad (1)$$

$$G_x = I(x,y+1) - I(x,y-1) \quad (2)$$

โดยที่

$I(x,y)$ คือค่าของพิกเซล ณ ตำแหน่ง ของรูปภาพ I

จากนั้นจึงนำค่า G_x และ G_y ไปคำนวณหาค่า Gradient Orientationθ ดังสมการ (3)

$$\theta(x,y) = tan^{-1}\frac{G_y}{G_x} \quad (3)$$

สุดท้าย Gradient Orientation ของแต่ละบล็อกจะถูกนำไป Weighted และจัดเก็บลงในฮิสโตแกรมตามขนาดของ Orientation Binβ

ดังนั้น หากกำหนดให้รูปภาพถูกแบ่งออกเป็น 3 x 3 บล็อก และกำหนดให้ β มีจำนวน 8 Bin คุณลักษณะพิเศษที่ได้จากการคำนวณด้วยวิธี HOG จะมีคุณลักษณะพิเศษจำนวน 72 (3 x 3 x 8) คุณลักษณะ

Scale-Invariant Features Transform (SIFT)

วิธี SIFT ถูกนำเสนอโดย Lowe[2] เพื่อใช้สำหรับเปรียบเทียบจุดสำคัญ (Keypoint) ที่คล้ายคลึงกันระหว่างรูปภาพสองรูป ได้แก่ รูปที่นำไปค้นหา (Query Image) และรูปภาพที่ต้องการค้นคืน (Retrieve Image) โดยนำรูปภาพทั้ง

สองไปหา Keypoint เพื่อนำมาเป็นลักษณะเด่นของแต่ละรูปภาพ จากนั้นคำนวณหาคุณลักษณะพิเศษจากพื้นที่บริเวณรอบ Keypoint ขั้นตอนต่อไป นำคุณลักษณะพิเศษของแต่ละ Keypoint จาก Query Image ไปเปรียบเทียบกับคุณลักษณะพิเศษของ Keypoint จากรูปภาพที่ต้องการค้นคืน เพื่อหา Keypoint ที่มีความคล้ายคลึงกันมากที่สุด สุดท้ายแล้ว จะทำให้รู้ว่า Query Image ที่นำไปค้นคืนคล้ายคลึงกับบริเวณไหนของ Retrieve Image ที่สุด โดยคุณลักษณะพิเศษที่คำนวณได้ในแต่ละ Keypoint มีจำนวน 128 คุณลักษณะ เนื่องจากบริเวณพื้นที่รอบ Keypoint แต่ละจุด จะถูกแบ่งออกเป็นบล็อกขนาด 4 x 4 แต่ละบล็อกถูกกำหนดให้มี Orientation Bin จำนวน 8 Bin (4 x 4 x 8)[16]

ในการคำนวณหาคุณลักษณะพิเศษด้วยวิธี SIFT รูปภาพจะถูกแปลงให้เป็นภาพสีเทา และนำไป Convolution โดยใช้ Gaussian Kernel ดังสมการ (4)

$$L(x, y, \sigma) = G(x, y, \sigma)*I(x, y) \quad (4)$$

โดยที่

$G(x, y, \sigma)$ คือ Gaussian Kernel

$I(x, y)$ คือค่าของพิกเซล ณ ตำแหน่ง x, y ของรูปภาพ I

σ คือความกว้างของ Gaussian Kernel

จากนั้นคำนวณหาค่า Gradient Orientationθ ทั้งในแนวนอนและแนวตั้ง ดังสมการ (5) และสมการ (6)

$$G_x = I(x + 1, y, \sigma) - I(x\text{-}1, y, \sigma) \quad (5)$$
$$G_y = I(x, y + 1, \sigma) - I(x, y\text{-}1, \sigma) \quad (6)$$

จากนั้น ค่า G_x และ G_y ถูกนำไปคำนวณหาค่า Gradient Orientation$\theta\ (x,y)$ เพื่อนำค่า θ ไปจัดเก็บลงในฮิสโตแกรมโดยกำหนดให้ $\beta = 8$

ขั้นตอนสุดท้าย นำคุณลักษณะพิเศษที่ได้จากการคำนวณด้วยวิธี HOG และ SIFT ไปเรียนรู้และจำแนกประเภทด้วยวิธีการเรียนรู้เครื่องจักร โดยใช้วิธี SVM[17] ที่ใช้ RBF Kernel และ KNN[18] ที่กำหนดให้ K=1 และใช้วิธี Euclidean เพื่อหาค่าระยะห่าง

การเรียนรู้เชิงลึก (Deep Learning)

งานวิจัยฉบับนี้ นำเสนอการเรียนรู้เชิงลึกที่ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยเปรียบเทียบการทำงานระหว่างโครงสร้างแบบ LeNet-5 และ AlexNet (4, 8-9)

โครงสร้างแบบ LeNet-5 (LeNet-5 Architecture)

โครงสร้างแบบ LeNet-5 นำเสนอโดย LeCun et al.[20] โดยเพิ่มชั้นการคำนวณแบบคอนโวลูชัน (Convolutional) เข้าไปในโครงข่าย ส่งผลให้โครงข่ายสามารถสกัดลักษณะเด่นจากรูปภาพ และจำแนกประเภทได้ในคราวเดียวกัน โครงข่าย CNN ประกอบด้วย 3 ชั้นหลัก ดังต่อไปนี้

ชั้นคอนโวลูชัน (Convolutional Layer)

ลักษณะเด่นของโครงข่ายแบบ CNN ก็คือการทำงานของ Convolutional Layer ที่คำนวณเพื่อหาชั้นของผลลัพธ์ซึ่งเรียกว่า Feature Map ด้วยการนำพื้นที่ส่วนย่อยรูปภาพ (Sub-region) ไปคำนวณแบบ dot product กับเคอร์นอล (Kernel) โดย Kernel ที่นำมาคำนวณจะต้องมีขนาดเล็กกว่ารูปภาพ การคำนวณของ Convolutional Layer แสดงดัง Figure 1

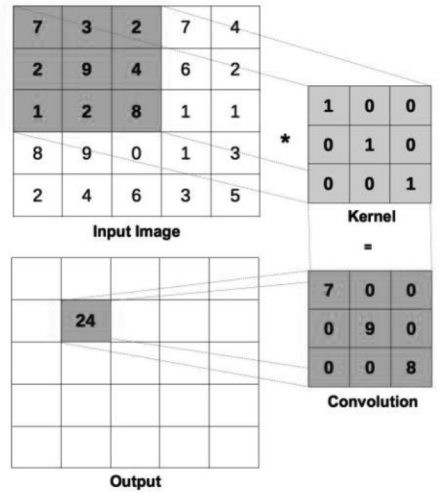


Figure 1 Convolution method with the dot product calculation between kernel and sub-region of the image.

ชั้นพูลลิ่ง (Pooling Layer)

Pooling Layer เป็นชั้นที่เชื่อมจาก Convolutional Layer โดยมีเป้าหมายคือทำให้ขนาดของ Feature Map ลดลง ในการคำนวณสามารถใช้ค่าต่ำสุด (Min Pooling) ค่าสูงสุด (Max Pooling) ผลรวม (Sum Pooling) และค่าเฉลี่ย (Average Pooling)[22] ในการคำนวณ Feature Map จะถูกแบ่งออกเป็นบล็อกขนาด ซึ่งหากใช้วิธี Max Pooling ในการคำนวณ ค่าที่ได้ก็คือค่าสูงสุด (Max Value) ของแต่ละบล็อก

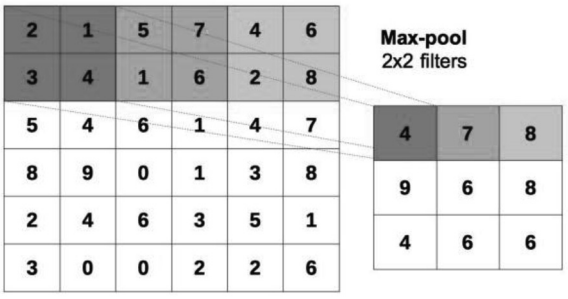


Figure 2 The illustration of max pooling with 2x2 filter and stride 2.

Figure 2 แสดงวิธีการคำนวณ Max Pooling จาก Feature Map ที่มีขนาด 6 x 6 บล็อก ในการคำนวณกำหนดให้ Pool มีขนาด 2 x 2 บล็อก ข้อมูลที่อยู่ในบล็อกที่ $F(m_1,n_1)$ ประกอบด้วย $\begin{Bmatrix} 2 & 1 \\ 3 & 4 \end{Bmatrix}$ ดังนั้น ผลลัพธ์ที่ได้จากการทำ Max Pooling คือ 4 จากนั้นจึงเลื่อน Pool ไปยังบล็อกถัดไป $F(m_i,n_{j+1})$ และทำไปจนกระทั่งบล็อกสุดท้าย

ชั้นเชื่อมโยงสมบูรณ์ (Fully-Connected Layer)

Fully-Connected Layer ก็คือ Hidden Layer และ Output Layer ของโครงข่ายประสาทเทียม ดังนั้น Fully-Connected Layer จึงทำหน้าที่ในการเรียนรู้ (Training) และการจำแนกประเภทของวัตถุ โดยผลลัพธ์ ที่ได้ก็คือจำนวนของ Class ที่ต้องการจำแนก

โครงข่ายแบบ CNN สามารถที่จะเพิ่ม Convolutional Layer และ Pooling Layer ได้อย่างไม่จำกัด จากงานวิจัย[9] โครงสร้างแบบ LeNet-5 ถูกกำหนดให้มีโครงสร้างดังต่อไปนี้

- Convolutional Layer 1 (Conv1) จำนวน 6 Feature Map, Filter ขนาด 5x5 และ Stride=1
- Avg-Pooling Layer 2 (Max-Pool2) จำนวน 6 Layer, Pool ขนาด 2x2 และ Stride=2
- Convolutional Layer 3 (Conv3) จำนวน 16 Feature Map, Filter ขนาด 5x5 และ Stride=1
- Avg-Pooling Layer 4 (Max-Pool4) จำนวน 16 Layer, Pool ขนาด 2x2 และ Stride=2
- Fully-Connected (FC) ที่ชั้น FC5 มีจำนวน 120 โหนด (Node) ชั้น FC6 มีจำนวน 84 Node และชั้นผลลัพธ์ (Output Layer) จำนวน 10 Node

ในงานวิจัยฉบับนี้ได้ปรับปรุงโครงสร้างแบบ LeNet-5 โดยโครงสร้างใหม่ แสดงดังต่อไปนี้

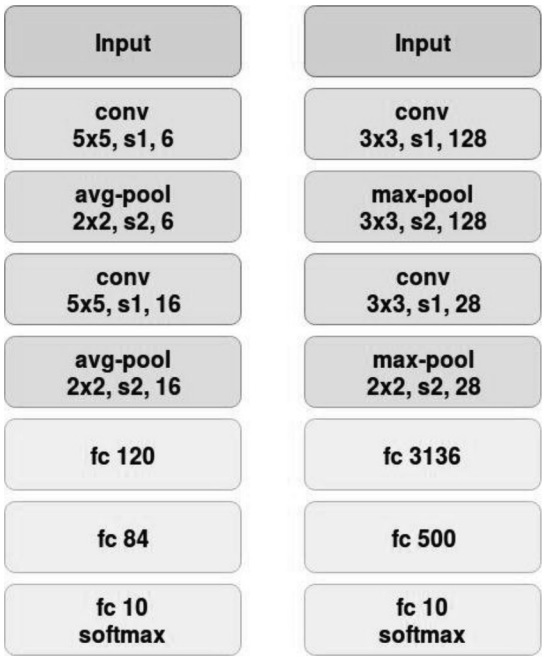
- Conv1 จำนวน 128 Feature Map, Filter ขนาด 3x3 และ Stride=1
- Max-Pool2 จำนวน 128 Layer, Pool ขนาด 3x3 และ Stride=2
- Conv3 จำนวน 28 Feature Map, Filter ขนาด 3x3 และ Stride=1
- Max-Pool4 จำนวน 28 Layer, Pool ขนาด 2x2 และ Stride=2
- Fully-Connected (FC) ที่ชั้น FC5 มีจำนวน 3,136 Node ชั้น FC6 มีจำนวน 500 Node และผลลัพธ์ (Output Layer) จำนวน 10 Node

โครงสร้างแบบ LeNet-5[9] แสดงดัง Figure 3a) และโครงสร้าง LeNet-5 ที่ถูกปรับปรุงและใช้ในงานวิจัยฉบับนี้ แสดงดัง Figure 3b)

โครงสร้างแบบ AlexNet (AlexNet Architecture)

โครงสร้างแบบ AlexNet ถูกนำเสนอในงานวิจัย[8] โดยโครงสร้างมีจำนวน Layer ทั้งสิ้น 8 Layer ประกอบไปด้วย Convolutional Layer จำนวน 5 Layer และ Fully-Connected Layer จำนวน 3 Layer รายละเอียดของโครงสร้างแบบ AlexNet แสดงดังต่อไปนี้

- Conv1 จำนวน 96 Feature Map, Filter ขนาด 11x11x3 และ Stride=4
- Max-Pool1 จำนวน 96 Layer, Pool ขนาด 2x2 และ Stride=2
- Conv2 จำนวน 256 Feature Map, Filter ขนาด 5x5x48 และ Stride=2
- Max-Pool2 จำนวน 256 Layer, Pool ขนาด 2x2 และ Stride=2
- Conv3 จำนวน 384 Feature Map, Filter ขนาด 3x3x256 และ Stride=2
- Conv4 จำนวน 384 Feature Map, Filter ขนาด 3x3x192 และ Stride=2
- Conv5 จำนวน 256 Feature Map, Filter ขนาด 3x3x192 และ Stride=2
- Max-Pool3 จำนวน 28 Layer, Pool ขนาด 2x2 และ Stride=2
- Fully-Connected (FC) ที่ชั้น FC1 มีจำนวน 4,096 Node ชั้น FC2 มีจำนวน 4,096 Node และ FC3 หรือชั้นผลลัพธ์ (Output Layer) จำนวน 10 Node



(a) (b)

Figure 3 Architectures of the convolutional neural network (CNN). a) The original LeNet-5 architecture (9) and b) the LeNet-5 architecture used in our experiments.

โครงสร้างแบบ AlexNet[8] แสดงดัง Figure 4a) และในงานวิจัยฉบับนี้ได้ปรับปรุงโครงสร้างแบบ AlexNet ในส่วนของ Fully-Connected Layer โดยลดจำนวนของ Node ลง เพื่อลดเวลาในการประมวลผล อีกทั้งยังทำให้ประสิทธิภาพของโครงสร้างแบบ AlexNet เพิ่มขึ้น โครงสร้างใหม่แสดงดัง Figure 4b)

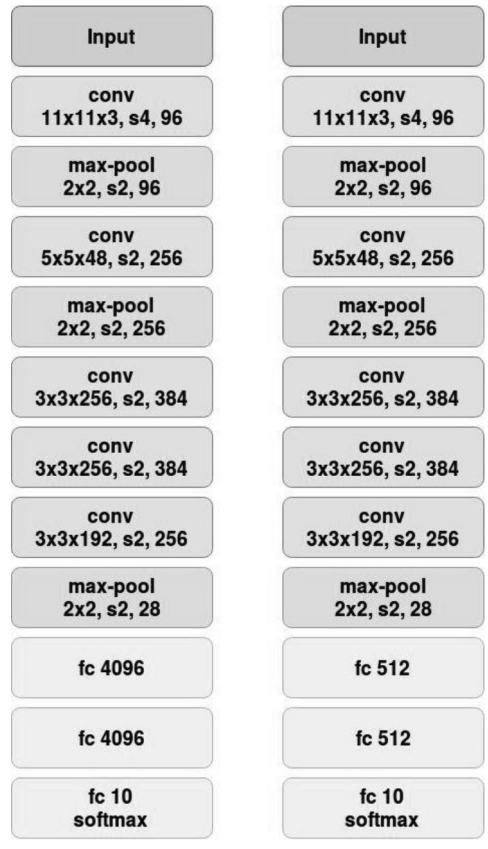
ชุดข้อมูลลายผ้าไหมไทยและผลการทดลอง (Thai Silk Pattern Dataset and Experimental Results)

ในส่วนนี้อธิบายถึงชุดข้อมูลลายผ้าไหมที่ใช้ในการทดลอง วิธีการเลือกข้อมูลชุดเรียนรู้ และผลลัพธ์ และการอภิปรายผลที่ได้จากการทดลอง

ชุดข้อมูลลายผ้าไหมไทย (Thai Silk Pattern Dataset)

รูปภาพลายผ้าไหม ที่ใช้ในงานวิจัยเก็บรวบรวมมาจากศูนย์จำหน่ายสินค้า OTOP บ้านหนองเขื่อนช้าง จำนวนทั้งสิ้น 10 ลาย ชื่อของลายผ้าไหมที่แสดงใน Figure 5 ประกอบด้วย ลายกระจับจั่ว ลายนกยูง ลายกระจับหนาม ลายกุญแจใจ ลายน้ำฟ้าคาดทอง ลายนาคน้อย ลายตะขอ ลายสร้อยดอกหมาก ลายสร้อยดอกหมากเล็ก และลายไข่มดแดง ตามลำดับ

ชุดข้อมูลลายผ้าไหมมีจำนวน 300 รูปภาพ เก็บรวบรวมลายละ 30 รูปภาพ โดยทุกรูปจัดเก็บเป็นภาพสีแบบ RGB (RGB Color Space) จากนั้น รูปภาพลายผ้าไหมทุกรูปถูกเลือกเฉพาะส่วน (Crop) ที่เป็นลายผ้าไหมเท่านั้น และเปลี่ยนรูปภาพให้มีขนาด (Normalized) 450x650 พิกเซล ข้อมูลที่ใช้ในการทดสอบ (Test Set) ได้มาจากการสุ่มเลือก (Random Crop) จากรูปภาพลายผ้าไหม โดยรูปภาพลายผ้าไหมหนึ่งรูปจะถูกสุ่มเลือกจำนวน 3 ครั้ง ดังนั้น รูปภาพใน Test Set จะมีจำนวนทั้งสิ้น (10 ลายx 30 รูปภาพ x 3 ครั้ง) 900 รูปภาพต่อการ Crop หนึ่งครั้ง



a) b)

Figure 4 The architecture of AlexNet. a) The AlexNet architecture presented in (8) and b) the AlexNet architecture used in our experiments.

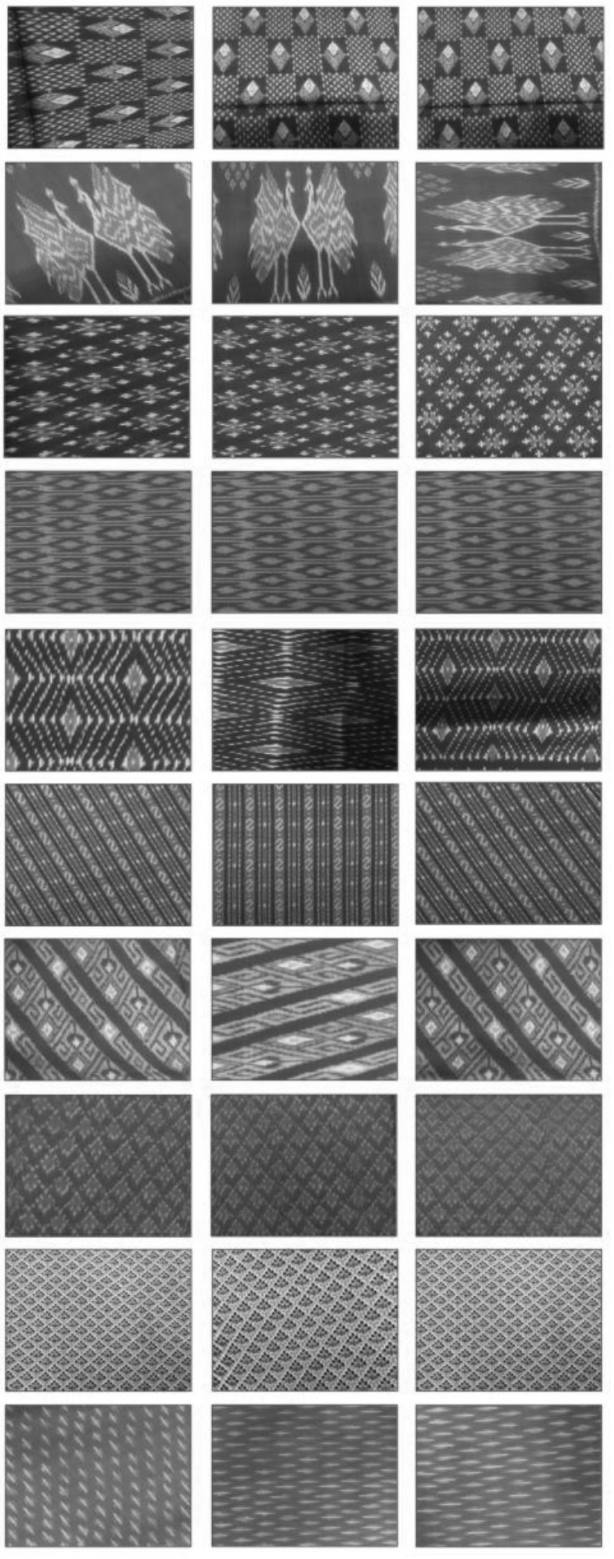


Figure 5 Sample images from the Thai silk pattern dataset. Note that, the images on each row represent one class.

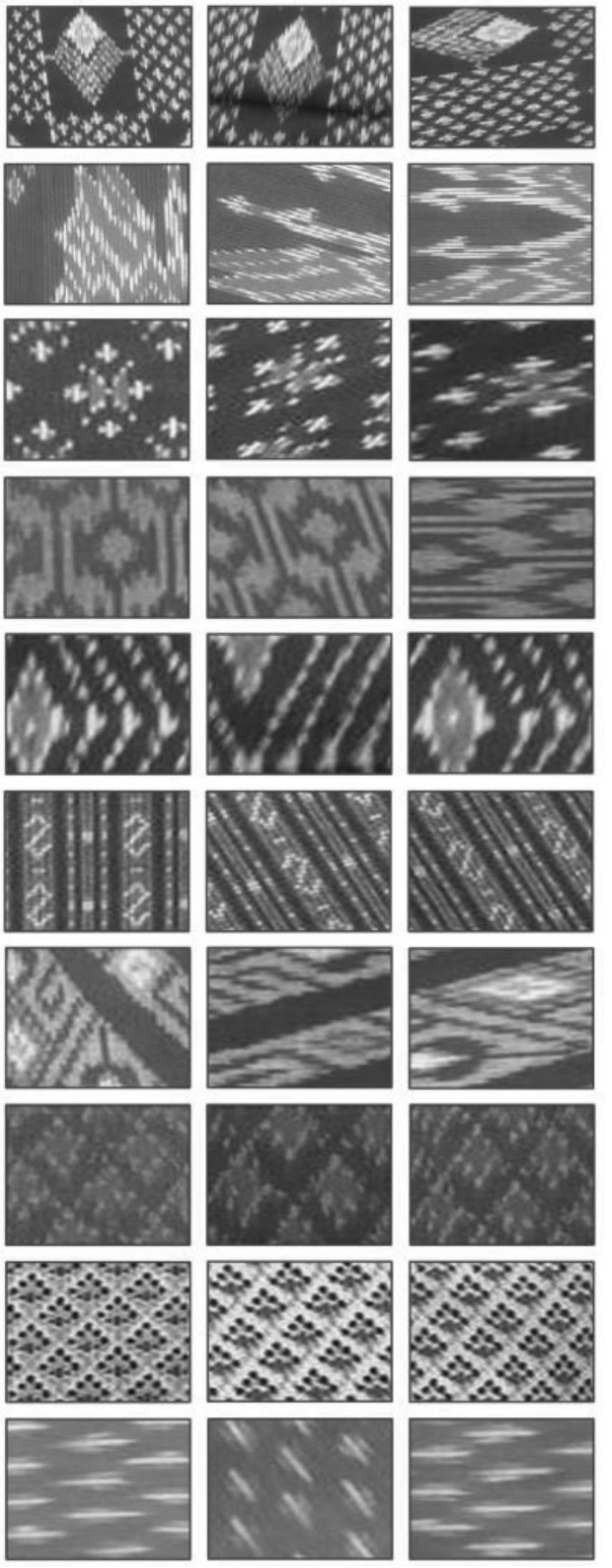


Figure 6 Test images random cropping from the whole image. In these sample images, size of the test image is 30 percent smaller than the original image.

ผลลัพธ์ที่ได้จากการทดลอง (Experimental Results)

งานวิจัยฉบับนี้ ใช้วิธี K-Fold Cross Validation เพื่อแบ่งข้อมูลออกเป็น 2 ส่วน ประกอบด้วย ข้อมูลชุดเรียนรู้ (Training Set) และข้อมูลชุดทดสอบ (Test Set) โดยกำหนดให้ k =5

การทดลองกำหนดให้มี Test Set จำนวน 2 ชุด ข้อมูลชุดที่ 1 กำหนดให้การ Crop รูปภาพมีขนาดเป็น 30% (Crop-30) และข้อมูลชุดที่ 2 กำหนดให้ Crop รูปภาพขนาด 40% (Crop-40) ดังนั้น รูปภาพที่ได้จากการ Crop ขนาด 30% จะมีขนาดของภาพเป็น 135x180 พิกเซล และการ Crop ขนาด 40% จะมีขนาดของภาพเป็น 180x240 พิกเซล Figure 6 แสดงตัวอย่างของรูปภาพที่ได้จากการ Crop ขนาด 30%

วิธีที่ใช้ในการทดสอบการค้นคืนลายผ้าไหมแบ่งออกเป็น 2 วิธี ได้แก่ 1) วิธีการเรียนรู้เชิงลึกแบบ Convolutional Neural Network (CNN) โดยใช้โครงสร้างแบบ LeNet-5 และ AlexNet และ 2) วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ประกอบด้วยวิธี SIFT และ HOG ร่วมกับวิธี SVM และการหาค่าระยะห่างแบบ Euclidean

การทดลองด้วยวิธีการเรียนรู้เชิงลึก รูปภาพที่ใช้การการเรียนรู้และการทดสอบจะถูกเปลี่ยนให้มีขนาด 128x128 พิกเซล โดยกำหนดพารามิเตอร์ ดังต่อไปนี้ Learning rate กำหนดเป็น 0.001 จำนวนรอบ (Iteration) ที่ใช้ในการเรียนรู้จำนวน 200 รอบ จำนวนข้อมูลต่อครั้งที่ใช้ในการเรียนรู้ (Batch Size) จำนวน 32

สำหรับการทดลองด้วยวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ รูปภาพที่ใช้ในการเรียนรู้จะมีขนาด 450x650 พิกเซล และรูปภาพที่ใช้ในการทดสอบมีสองขนาดคือ 135x180 และ 180x240 พิกเซลสำหรับการ Crop ขนาด 30 และ 40% ตามลำดับ

ในการทดสอบประสิทธิภาพด้วยวิธี HOG พารามิเตอร์ที่ใช้ในการทดสอบ ประกอบด้วย ขนาดของบล็อกที่กำหนดให้มีขนาด 1x1 บล็อก และกำหนดให้ Orientation bin มีจำนวน 128 Bin ดังนั้น คุณลักษณะพิเศษที่ได้จากการคำนวณด้วยวิธี HOG จึงมีจำนวน 128 คุณลักษณะต่อรูปภาพ 1 รูป

สำหรับวิธี SIFT พารามิเตอร์ที่ใช้ในการทดสอบประกอบด้วย จำนวนของ Keypoint ต่อรูปภาพ 1 รูป ในงานวิจัยนี้ได้กำหนดให้มีจำนวนรูปภาพละ 1 Keypoint ในการคำนวณ 1 Keypoint จะคำนวณคุณลักษณะพิเศษได้ 128 คุณลักษณะ

สำหรับการทดลองด้วยวิธี SVM กำหนดให้ใช้ Kernel แบบ RBF และใช้วิธี Grid Search เพื่อค้นหาพารามิเตอร์ C และ gamma โดยค้นหาตั้งแต่ช่วง $\{2^{-3}, 2^{-2},..., 2^{3}, 2^{4}\}$ จากนั้นเลือก C และ gamma ที่ให้ผลในการทดสอบสูงที่สุด

ตัวอย่างผลลัพธ์ที่ได้จากการค้นคืนรูปภาพลายผ้าไหมด้วยวิธีที่แตกต่างกัน 6 วิธี แสดงดัง Figure 7

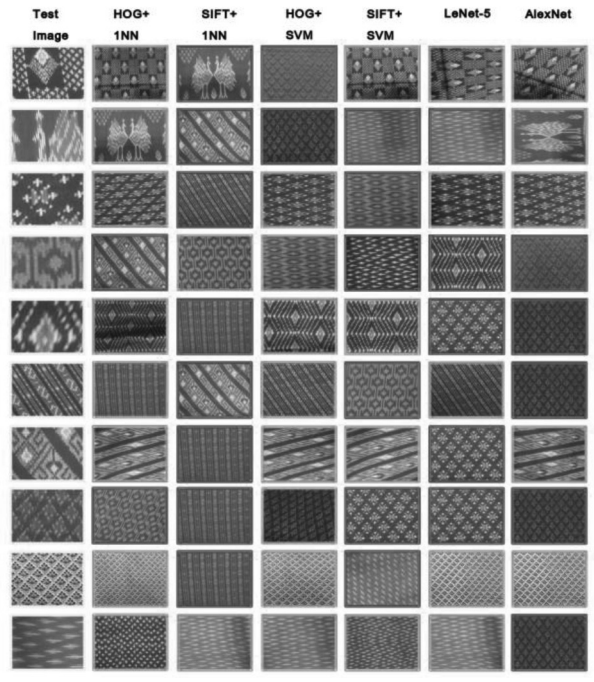


Figure 7 Example of retrieve images. Each row shows the retrieved results for each technique given a test image (first column).

Table 1 แสดงให้เห็นถึงอัตราความถูกต้องในการค้นคืนรูปภาพลายผ้าไหมด้วยการเรียนรู้เชิงลึกโดยใช้วิธี CNN และกำหนดโครงสร้างแบบ LeNet-5 ในงานวิจัยได้ทดสอบโครงสร้างแบบ LeNet-5 ที่อ้างอิงตามงานวิจัย (9) (โครงสร้างแสดงดัง Figure 3a) และโครงสร้างที่แสดงใน Figure 3b โดยโครงสร้างทั้งสองมีความแตกต่างกัน ดังนี้ 1) จำนวนของ Convolution Kernel 2) จำนวนของ Feature Map 3) วิธีการทำ Pooling และ 4) จำนวน Node ในชั้น Fully-Connected

Table 1 Accuracy results (accuracy and standard deviation) using different LeNet-5 parameters for the Thai silk pattern dataset.

Methods	Test Accuracy	
	Crop-30	Crop-40
Ori-Avg-pool (Top-1) (9)	50.19±2.90	61.94±2.29
Ori-Max-pool (Top-1) (9)	49.07±4.62	55.06±3.40
Our-Avg-pool (Top-1)	63.18±1.41	71.44±2.08
Our-Max-pool (Top-1)	**64.06±2.25**	**76.98±2.29**

จากการทดลองพบว่า โครงสร้างแบบ LeNet-5 ที่ผู้วิจัยได้ออกแบบ (Our-Max-pool) มีอัตราการค้นคืนรูปภาพสูงที่สุด เมื่อเปรียบเทียบกับโครงสร้างแบบ LeNet-5 ที่นำเสนอใน[9] โดยมีอัตราการค้นคืน 64.06% และ 76.98% ในข้อมูล Crop-30 และ Crop-40 ตามลำดับ ซึ่งมีอัตราการค้นคืนสูงกว่าที่นำเสนอใน[9] มากกว่า 10% โดยผลการค้นคืนแสดงเฉพาะ Top-1 เท่านั้น

ในส่วนของโครงสร้างแบบ AlexNet ผู้วิจัยได้ทดสอบด้วยการลดจำนวนของ Node ในชั้น Fully-Connected[21] จากจำนวน 4,096 โหนด เป็น 1,024 โหนด 512 โหนด และ 256 โหนด ตามลำดับ

ผลการทดลองใน Table 2 แสดงให้เห็นว่าการลดขนาดของโหนดส่งผลให้อัตราการค้นคืนภาพสูงขึ้น และเมื่อทดสอบกับชุดข้อมูลลายผ้าไหมไทยพบว่าจำนวนโหนด 512 โหนด มีผลการทดลองสูงที่สุดที่ 44.70% และ 55.58% ในชุดข้อมูล Crop-30 และ Crop-40 ตามลำดับ แต่เมื่อเปรียบเทียบระหว่างโครงสร้างแบบ LeNet-5 และ AlexNet พบว่าโครงสร้างแบบ LeNet-5 มีอัตราการค้นคืนสูงกว่าโครงสร้างแบบ AlexNet

Table 2 Test Accuracy comparison among different numbers of nodes in the AlexNet architecture on the Thai silk pattern dataset.

Number of nodes	Test Accuracy	
	Crop-30	Crop-40
256	41.52±1.46	47.54±1.54
512	**44.70±0.94**	**55.58±1.04**
1024	34.71±2.07	40.44±1.12
4096	27.42±0.84	32.65±1.62

Table 3 Performances of the 6 different techniques on the Thai silk pattern dataset.

Methods	Test Accuracy	
	Crop-30	Crop-40
HOG+1NN	**92.05±0.31**	**89.73±0.79**
SIFT+1NN	23.03±0.46	57.99±0.33
HOG+SVM	74.92±1.94	82.68±4.67
SIFT+SVM	42.8±6.98	40.24±1.08
Our-LeNet-5 (Top-1)	64.06±2.25	76.98±2.29
AlexNet-512 (Top-1)	44.70±0.94	55.58±1.04

Table 3 แสดงให้เห็นถึงอัตราการค้นคืนรูปภาพด้วยลายผ้าไหมด้วยวิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ที่ประกอบด้วยวิธี HOG+1NN, SIFT+1NN, HOG+SVM และ SIFT+SVM และเปรียบเทียบอัตราการค้นคืนกับวิธี CNN โดยใช้โครงสร้างแบบ Our-LeNet-5 (Top-1) และ AlexNet-512 (Top1)

จากการทดลองพบว่าวิธี HOG+1NN มีอัตราการค้นคืนสูงที่สุดทั้งในชุดข้อมูล Crop-30 และ Crop-40 โดยมีอัตราการค้นคืน 92.05% และ 89.73% ตามลำดับ ในทางกลับกันวิธี SIFT+1NN และ SIFT+SVM มีอัตราการค้นคืนต่ำที่สุด 23.03% สำหรับชุดข้อมูล Crop-30 และ 57.09% สำหรับชุดข้อมูล Crop-40

สรุปผล

วิธีการที่ใช้ในการค้นคืนรูปภาพลายผ้าไหมที่นำเสนอในงานวิจัยฉบับนี้มีทั้งสิ้น 2 วิธี 1) วิธีการหาคุณลักษณะพิเศษเฉพาะพื้นที่ ประกอบด้วยวิธี Scale-Invariant Feature Transform (SIFT) และ Histogram of Oriented Gradients (HOG) ร่วมกับวิธีการเรียนรู้เครื่องจักร ด้วยวิธี Support Vector Machine (SVM) และการหาค่าระยะห่างแบบ Euclidean และ 2) วิธีการเรียนรู้เชิงลึกแบบ Convolutional Neural Network (CNN) โดยใช้โครงสร้างแบบ LeNet-5 และ AlexNet

โดยทั้งสองวิธีข้างต้นถูกนำไปทดสอบกับชุดข้อมูลลายผ้าไหมไทย (Thai Silk Pattern Dataset)

จากการทดลองกับชุดข้อมูลลายผ้าไหมไทย โดยมี Test Set ทั้งสิ้น 2 ชุด ประกอบด้วย Crop-30 และ Crop-40 ปรากฏว่าวิธี HOG+1NN มีอัตราการค้นคืนสูงกว่าวิธีอื่นทั้งหมด โดยมีอัตราการค้นคืนในข้อมูล Crop-30 และ Crop-40 ที่ 92.05% และ 89.73% ตามลำดับ

เมื่อเปรียบเทียบวิธี CNN โดยใช้โครงสร้างแบบ LeNet-5 และ AlexNet ปรากฏว่า LeNet-5 มีอัตราการค้นคืนมากกว่า 20% โดยประมาณ ทั้งนี้เนื่องจากจำนวนของรูปภาพลายผ้าไหมที่ใช้ในการทดสอบมีจำนวนจำกัด

งานวิจัยฉบับต่อไป การเพิ่มจำนวนของ Training Set หรือที่เรียกว่า Data Augmentation และวิธีการ Transfer Learning[20-22] อาจส่งผลทำให้ประสิทธิภาพของวิธี CNN เพิ่มขึ้น

เอกสารอ้างอิง

1. Bhute A.N. Meshram B.B. Content Based Image Indexing and Retrieval. International Journal of Graphics & Image Processing 2013; 3(4):235-246.
2. Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 2004; 60(2): 91-110.
3. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. International Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
4. Heusch G, Rodriguez Y, Marcel S. Local Binary Patterns as an Image Preprocessing For Face Authentication. International Conference on Automatic Face and Gesture Recognition, 2006. pp. 6-14.
5. Shekhar R, Jawahar C.V. Word Image Retrieval using Bag of Visual Words. 10th International Workshop on Document Analysis Systems (DAS), 2012. pp. 297-301.
6. Wan J, Wang D, Hoi S, Wu P, Zhu J, Zhang Y, Li J. Deep Learning for Content-based Image Retrieval: A Comprehensive Study. the 22nd International Conference on Multimedia, 2014. pp. 157-166.
7. Singh A.V. Content-based Image Retrieval using Deep Learning. Rochester Institute of Technology, 2015.
8. Krizhevsky A, Sutskever I, Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 2012. pp. 1097-1105.
9. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based Learning Applied to Document Recognition. Proceedings of the Institute of Electrical and Electronic Engineers; 1998:86(11), pp. 2278-2324.
10. Karaaba M.F, Surinta O, Schomaker L.R.B, Wiering M.A. Robust Face Identification with Small Sample Sizes using Bag of Words and Histogram of Oriented Gradients. the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), 2016. p. 582-589.
11. Ahonen T, Hadid A, Pietikäinen, M. Face Recognition with Local Binary Patterns. European Conference on Computer Vision (ECCV), Berlin, Heidelberg; 2004. p. 469-481.
12. อนุมาศ แสงสว่าง. การสืบค้นรูปภาพจากการเปรียบเทียบค่าฮิสโตแกรมโดยใช้เวคเตอร์โมเดล. การประชุมวิชาการระดับชาติ เบญจมิตรวิชาการ ครั้งที่ 2, 2011, pp. 1–9.
13. ประภาพร กุลลิ้มรัตน์ชัย. การค้นคืนภาพโดยพิจารณาน้ำหนักการกระจายของสีด้วยการกระจายตัวแบบเกาส์เซียนสำหรับฮิสโตแกรมสีในแบบจำลองสี HSV. วารสารวิชาการมหาวิทยาลัยอีสเทิร์น เอเชียฉบับ วิทยาศาสตร์และเทคโนโลยี, vol. 6, no. 2, pp. 101–9, 2012.
14. Hazra T.K, Chowdhury S.R, Chakraborty A.K. Encrypted Image Retrieval System: A Machine Learning Approach. the 7th Annual Conference on Information Technology, Electronics and Mobile Communication, 2016. pp. 1-6.
15. Surinta O, Karaaba M.F, Schomaker L.R, Wiering M.A. Recognition of Handwritten Characters using Local Gradient Feature Descriptors. Engineering Applications of ArtificialIntelligence, 2015; 45: 405-414.
16. Surinta O, Karaaba, M.F, Mishra T.K, L.R, Wiering M.A. Recognizing Handwritten Characters with Local Descriptors and Bags of Visual Words. Engineering Applications of Neural Networks (EANN), 2015. p. 255-264.
17. Vapnik V. Statistical Learning Theory. Wiley, New York, 1998.
18. Kumar M, Jindal M, Sharma R. K-Nearest Neighbor Based Offline Handwritten Gurmukhi Character Recognition. International Conference on Image Information Processing, 2011. pp. 1–4
19. Le Cun Y, Matan O, Boser B, Denker J.S, Henderson D, Howard R.E, Baird H.S. Handwritten Zip Code

Recognition with Multilayer Networks. The 10th International Conference on Pattern Recognition (ICPR), 1990; 2:35-40.

20. Pawara P, Okafor E, Surinta O, Schomaker L.R, Wiering M.A. Comparing Local Descriptors and Bags of Visual Words to Deep Convolutional Neural Networks for Plant Recognition. the 6th International Conference on Pattern Recognition Applications and Model (ICPRAM), 2017.pp. 479-486.
21. Pawara P, Okafor E, Schomaker L.R, Wiering M.A. Data Augmentation for Plant Classification. International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS), 2017. pp. 615-626.
22. Okafor E, Pawara P, Karaaba F, Surinta O, Codreanu V, Schomaker L.R, Wiering M.A. Comparative Study Between Deep Learning and Bag of Visual Words for Wild-Animal Recognition. Symposium Series on Computational Intelligence (SSCI), 2016. pp. 1-8.