การใช้กลุ่มของภาพฉากเพื่อจำแนกวิดีโอจากรายการโทรทัศน์

นายอิทธิศักดิ์ เผือกศรี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Using Clustered Frames to Classify Videos from Television Programs

Mr. Itthisak Phueaksri

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2019

หัวข้อวิทยานิพนธ์	การใช้กลุ่มของภาพฉากเพื่อจำแนกวิดีโอจากรายการโทรทัศน์
โดย	นายอิทธิศักดิ์ เผือกศรี
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.นัทที นิภานันท์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ)

..... กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.เด่นดวง ประดับสุวรรณ)

อิทธิศักดิ์ เผือกศรี : การใช้กลุ่มของภาพฉากเพื่อจำแนกวิดีโอจากรายการโทรทัศน์. (Using Clustered Frames to Classify Videos from Television Programs) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สินธุภิญโญ

งานวิจัยนี้นำเสนอวิธีการจำแนกวิดีโอ ด้วยเทคนิคแบบจำลองคอนโวลูชันสองมิติ และการเรียนรู้แบบกึ่งกำกับ โดยทั่วไปการจำแนกวิดีโอที่มีประสิทธิภาพสูง ถูกนำเสนอโดยใช้วิธีการเรียนรู้แบบลึก อย่างไรก็ตามจากการเพิ่มขึ้นของจำนวนวิดีโอในปัจจุบัน การเรียนรู้ของแบบจำลองเพื่อจำแนกวิดีโอจำเป็นต้องใช้ประสิทธิภาพในการประมวลผลสูง งานวิจัยนี้จึงนำเสนอวิธีการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติโดยใช้การซ้อนทับกันของภาพฉาก และการจัดกลุ่มของภาพฉากด้วยแผนที่จัดระเบียบด้วยตนเองก่อนนำไปสร้างแบบจำลองจำแนกประเภทรายการ โดยการสร้างแบบจำลองประเภทรายการถูกนำเสนอใน 4 รูปแบบ ประกอบด้วย การออกเสียง การคำนวณค่าความวุ่นวาย การเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม การเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว อีกทั้งยังประเมินจำนวนภาพฉากสำหรับการประมวลผลในการจัดกลุ่มโดยเปรียบเทียบระหว่างระยะเวลาการเรียนรู้และความแม่นยำ วิธีการในงานวิจัยนี้ถูกนำเสนอด้วยประเมินจากการเรียนรู้ด้วยชุดข้อมูลวิดีโอจำนวน 18 ประเภท 912 วิดีโอ จากรายการโทรทัศน์ ในการประเมินด้วยการประเมินผลแบบไขว้ จำนวน 5 โฟลด์ วิธีการในงานวิจัยนี้มีความแม่นยำเฉลี่ยร้อยละ 71.98 และใช้เวลาในการเรียนรู้โดยเฉลี่ยประมาณ 40 นาที นอกจากนี้ยังเปรียบเทียบกับการเรียนรู้ด้วยแบบจำลองอื่นๆ อาทิ แบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันร่วมกับหน่วยความจำระยะสั้นแบบยาว รวมถึงประเมินผลกับชุดข้อมูลพื้นฐาน Hollywood2 ซึ่งการเรียนรู้มีความแม่นยำเฉลี่ยร้อยละ 93.72

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2562	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6170982521 : MAJOR COMPUTER SCIENCE
KEYWORD:

Itthisak Phueaksri : Using Clustered Frames to Classify Videos from Television Programs. Advisor: Asst. Prof. SUKREE SINTHUPINYO, Ph.D.

This research presents techniques, including Convolutional Neural Network and Semi-Supervised Learning, to classify video clips. Usually, many tasks are done by categorizing video clips using deep learning techniques. However, based on the number of online videos today, it is necessary to use high computing power to accomplish this task. We present a traditional technique using a two-dimensional Convolutional Neural Network by stacking frames and propose using the Self-Organizing Map (SOM) to cluster video frames. We then classified them using simple voting, calculating entropy, neural networks, and Long-Short Term Memory (LSTM). We also show finding frame numbers that are used to cluster video frames according to accuracy and training time. The results of this approach are presented based on testing 18 specific classes of real-world datasets from TV-programs containing 912 videos. The authors evaluated the techniques using five-fold cross-validation that our method archived 71.98% of average accuracy. Their computing time was then assessed, which achieved approximately 40 minutes of average computing time. Moreover, we also compared the present proposal to other baseline models, including C3D and CNN-LSTM, and also evaluate the technique with Hollywood2 that archived 93.72% of average accuracy.

Field of Study:	Computer Science	Student's Signature .....
Academic Year:	2019	Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์อย่างดียิ่งของผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ อาจารที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้กรุณาสละเวลาให้ความรู้ ให้คำปรึกษาตรวจสอบ ให้คำแนะนำแนวทางการวิจัย และสนับสนุนจนทำให้การวิจัยในครั้งนี้สำเร็จลุล่วงด้วยดี ข้าพเจ้าจึงขอกราบระลึกพระคุณผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ ไว้ ณ ที่นี้

ข้าพเจ้าขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ ผู้ช่วยศาสตราจารย์ ผู้ช่วยศาสตราจารย์ ดร.เด่นดวง ประดับสุวรรณ ผู้ช่วยศาสตราจารย์ ดร.นัททที นิภานันท์ และ ผู้ช่วยศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาสละเวลา ให้คำแนะนำตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น

ท้ายนี้ข้าพเจ้าขอขอบพระคุณผู้บังคับบัญชาในสายงาน เพื่อนร่วมงาน และมิตรสหาย ที่คอยติดตามให้กำลังใจให้การสนับสนุนและความช่วยเหลือในด้านต่าง ๆ และท่านอื่น ๆ ที่มิได้กล่าวชื่อไว้ ณ ที่นี้ที่มีส่วนช่วยให้วิทยานิพนธ์ของข้าพเจ้าสำเร็จไปได้ด้วยดี

อิทธิศักดิ์ เผือกศรี

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

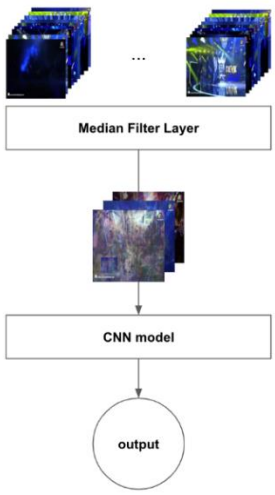
หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ความก้าวหน้าของเทคโนโลยี และความสะดวกสบายในการเข้าถึงเทคโนโลยี อาทิ สมาร์ทโฟน กล้องดิจิตอล ส่งผลให้การเพิ่มขึ้นของวิดีโอคลิปในปัจจุบันเป็นไปอย่างรวดเร็ว เนื่องจากทุกคนสามารถผลิตสื่อจำพวกคลิปวิดีโอ ผ่านอุปกรณ์ใกล้ตัว อย่างสมาร์ทโฟน อีกทั้งสามารถเผยแพร่คลิปวิดีโอ ผ่านช่องทางต่าง ๆ อาทิ สังคมออนไลน์ ดังนั้นการเข้าใจ และจำแนก คลิปวิดีโอ ได้ทันท่วงทีโดยใช้เทคนิคการเรียนรู้ด้วยเครื่อง (Machine learning) จึงเป็นความท้าทายสำหรับนักวิจัย

ความพยายามในการจำแนก และเข้าใจคลิปวิดีโอด้วยเทคนิคการเรียนรู้ด้วยเครื่อง ได้ถูกพัฒนาขึ้นอย่างต่อเนื่อง โดยใช้เทคนิคหลากหลายในการเรียนรู้ และเข้าใจ หนึ่งในเทคนิคที่ได้รับความนิยมมาก คือ การเรียนรู้เชิงลึก (Deep Learning) นักวิจัยนำเสนอแบบจำลองต่าง ๆ เพื่อจำแนกและเข้าใจวิดีโอ อาทิ การเรียนรู้จากคุณลักษณะของรูปภาพด้วยแบบจำลองคอนโวลูชันเพื่อจำแนกวิดีโอ [1] การจำแนกวิดีโอ โดยการเรียนรู้ขนาดใหญ่ (Large-Scale) จากคุณลักษณะเชิงพื้นที่ (Spatial Temporal Feature) ของภาพฉากในวิดีโอทั้งหมด ด้วยแบบจำลองคอนโวลูชันสามมิติ (3D Convolutional Neural Network) [2] และการจำแนกโดยพิจารณาวิดีโอโดยพิจารณาความต่อเนื่องของการเกิดของคุณลักษณะเชิงพื้นที่ (Feature) และการไหลของแสง (Optical Flow) ที่เกิดขึ้นในแต่ละภาพฉาก ด้วยเทคนิค Long Short-Term Memory [3]



ภาพ 1 แผนภาพการดำเนินการเรียนรู้เพื่อจำแนววิดีโอคลิปด้วยเทคนิคคอนโวลูชันสองมิติ

งานวิจัยนี้นำเสนอวิธีการจำแนกวิดีโอรายการโทรทัศน์ โดยมุ่งเน้นนำเสนอแนวทางการเรียนรู้ของแบบจำลองโดยลดขนาดของชุดข้อมูลวิดีโอ เพื่อลดการใช้งานทรัพยากรสำหรับการเรียนรู้ งานวิจัยนี้ได้มีการจัดเก็บข้อมูลรายการโทรทัศน์ จำนวน 18 รายการ จำนวน 912 วิดีโอ เพื่อใช้เป็นชุดข้อมูลสำหรับการทดลอง รวมทั้งนำเสนอการเรียนรู้ด้วย 2 เทคนิค คือ เทคนิคการเรียนรู้จากการซ้อนทับข้อมูลภาพฉากวิดีโอ และการเพิ่มชั้นตัวกรองมัธยฐาน (Median Filter) เพื่อลดมิติของชุดข้อมูล เพื่อนำไปเรียนรู้ด้วยแบบจำลองคอนโวลูชันแบบสองมิติ (ภาพที่ 1) และการใช้เทคนิคการแบ่งกลุ่มรูปภาพ (Image Clustering) ด้วยวิธีแผนที่การจัดระเบียบด้วยตนเอง (Self-Organizing Map) เพื่อจำแนกกลุ่มของรูปภาพก่อนนำไปจำแนกด้วยวิธีต่าง ๆ โดยนำเสนอการเรียนรู้จากข้อมูลคุณลักษณะของข้อมูลจากกลุ่มรูปภาพด้วย 4 วิธี คือ การจำแนกประเภทวิดีโอด้วยเสียงส่วนมาก (Majority Vote) ของการเกิดขึ้นของกลุ่มข้อมูล การจำแนกวิดีโอโดยพิจารณาค่าน้ำหนักจากค่าเอ็นโทรพีของชุดข้อมูล (Entropy) ของประเภทชุดข้อมูล การเรียนรู้จากคุณลักษณะของชุดข้อมูลด้วยโครงข่างประสาทเทียม (Neural Network) และการเรียนรู้จากลำดับการเกิดขึ้นของคุณลักษณะของกลุ่มข้อมูล ด้วยเทคนิคหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) งานวิจัยชิ้นนี้ได้แบ่งการวัดผลเป็นสองชุด คือ วัดผลโดยการเรียนรู้ชุดข้อมูลรายการโทรทัศน์ที่ถูกจัดทำขึ้นในงานวิจัยนี้ เทียบกับแบบจำลองที่นำเสนอและแบบจำลองพื้นฐานอื่น โดยแบ่งชุดข้อมูลที่ใช้ทดสอบเป็น 5 ชุดข้อมูล เพื่อวัดผลแบบจำลองด้วยวิธีวัดผลแบบไขว้ (Cross-Validation)

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อวิเคราะห์และจำแนกรายการโทรทัศน์ ด้วยวิธีพิจารณาภาพ
2. นำเสนอชุดข้อมูลรายการโทรทัศน์ 18 ประเภท จำนวน 912 วิดีโอ
3. นำเสนอแบบจำลองคอนโวลูชันสองมิติสำหรับจำแนกรายการโทรทัศน์
4. นำเสนอแบบจำลองแผนที่จัดการจัดระเบียบด้วยตนเองสำหรับจัดกลุ่มรูปภาพ
5. นำเสนอแบบจำลองสำหรับจำแนกรายการโทรทัศน์ด้วยกลุ่มรูปภาพ

1.3 ขอบเขตงานวิจัย

1. งานวิจัยนี้รองรับการจำแนกรายการโทรทัศน์เท่านั้น
2. งานวิจัยนี้จะสร้างแบบจำลองคอนโวลูชันสองมิติ สำหรับจำแนกรายการโทรทัศน์จำนวน 18 ประเภทข้อมูล จากชุดข้อมูลทั้งหมด 912 ชุดข้อมูล แบบการจำแนกฐานสอง (Binary Classification) เท่านั้น

3. งานวิจัยนี้จะสร้างแบบจำลองการจัดกลุ่มฉากด้วยแผนที่จัดระเบียบด้วยตนเอง สำหรับจัดกลุ่มฉากจากชุดข้อมูลรายการโทรทัศน์เท่านั้น
4. งานวิจัยนี้จะสร้างแบบจำลองการจำแนกรายการโทรทัศน์จากการจัดกลุ่มภาพฉาก ด้วยการลงคะแนน ค่าเอ็นโทรพี แบบจำลองโครงข่ายประสาทเทียม และหน่วยความจำระยะสั้นแบบยาวเท่านั้น

1.4 ขั้นตอนและวิธีการดำเนินงานวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการจำแนกวิดีโอ และจัดกลุ่มภาพฉาก
2. ศึกษาความรู้และทฤษฎีที่เกี่ยวข้องกับงานวิจัย
3. ศึกษาเครื่องมือที่ใช้สำหรับสร้างแบบจำลอง
4. เก็บรวบรวมข้อมูลรายการโทรทัศน์
5. ออกแบบ และสร้างแบบจำลอง
6. วัดประสิทธิภาพแบบจำลอง
7. วิเคราะห์ผลการวัดประสิทธิภาพ
8. สรุปผลและจัดทำเล่มวิทยานิพนธ์

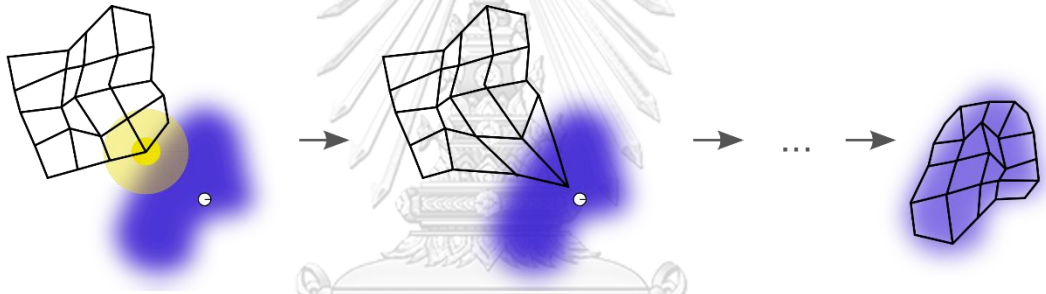
1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถใช้แบบจำลองคอนโวลูชันสองมิติเพื่อจำแนกรายการโทรทัศน์ได้
2. สามารถใช้แบบจำลองแผนที่จัดระเบียบด้วยตนเองจัดกลุ่มฉากได้
3. สามารถใช้กลุ่มฉากเพื่อจำแนกรายการโทรทัศน์ได้
4. ลดระยะเวลาการเรียนรู้สำหรับจำแนกรายการโทรทัศน์
5. สามารถนำแนวคิดการจำแนกรายการโทรทัศน์และการจัดกลุ่มฉากไปประยุกต์ใช้กับวิดีโออื่น ๆ ได้

บทที่ 2
งานวิจัยและทฤษฎีที่เกี่ยวข้อง

2.1 แผนที่การจัดระเบียบด้วยตนเอง

แผนที่การจัดระเบียบด้วยตนเอง [4] เป็นวิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) สร้างโครงข่ายประสาทเทียม เพื่อแสดงพื้นที่ของชุดข้อมูลที่ถูกนำมาเรียนรู้ ซึ่งถูกเรียกว่าแผนที่ (Map) มักถูกใช้เป็นวิธีหนึ่งในการลดมิติของข้อมูล โดยชุดข้อมูลที่ใช้สำหรับเรียนรู้จะอยู่ในรูปแบบของเวกเตอร์ กระบวนการของแผนที่การจัดระเบียบตัวเอง ประกอบไปด้วย 2 กระบวนการ คือการเรียนรู้ (Training) ซึ่งคือการสร้างพื้นที่ตามชุดข้อมูลเรียนรู้ (ตัวอย่างการเรียนรู้และกระบวนการปรับพื้นที่ของการเรียนรู้ด้วยตนเอง ถูกแสดงในภาพที่ 2.1) และการแมป (Mapped) ซึ่งคือการจำแนกเวกเตอร์ใหม่



ภาพ 2 ตัวอย่างกระบวนการจัดกลุ่มด้วยแผนที่การจัดระเบียบด้วยตนเอง

กระบวนการเรียนรู้ของแผนที่การจัดระเบียบด้วยตนเองใช้ประโยชน์ จากการเรียนรู้แบบแข่งขัน (Competitive learning) โดยนำเข้าตัวอย่างเรียนรู้ไปยังโครงข่าย และคำนวณหาระยะทางแบบยุคลิด (Euclidean distance) ของเวกเตอร์ทั้งหมด โดยนิวรอนที่น้ำหนักเวกเตอร์ใกล้เคียงข้อมูลนำเข้ามากที่สุดจะถูกเรียกว่า การจับคู่ที่ดีที่สุด (Best Matching Unit: BMU) ซึ่งค่าน้ำหนักของการจับคู่ที่ดีที่สุดจะถูกปรับด้วยเวกเตอร์นำเข้า โดยสูตรที่ใช้สำหรับการปรับค่าน้ำหนักของการเรียนรู้ด้วยตนเองถูกแสดงในสูตรที่ 2.1

$$W_v(s+1) = W_v(s) + \theta(u,v,s) \cdot \alpha(s) \cdot (D(t) - W_v(s))$$

โดย s คือ ลำดับการเรียนรู้

t คือ ลำดับของตัวอย่าง

น คือ ลำดับของการจับคู่ที่ดีที่สุดของ เวกเตอร์นำเข้า D(t)

$\alpha(s)$ คือ ค่าสัมปสิทธิ์เรียนรู้ (Learning rate)

$\theta(u, v, s)$ คือ ฟังก์ชันคำนวณระยะทางระหว่างเวกเตอร์ U และ เวกเตอร์ V

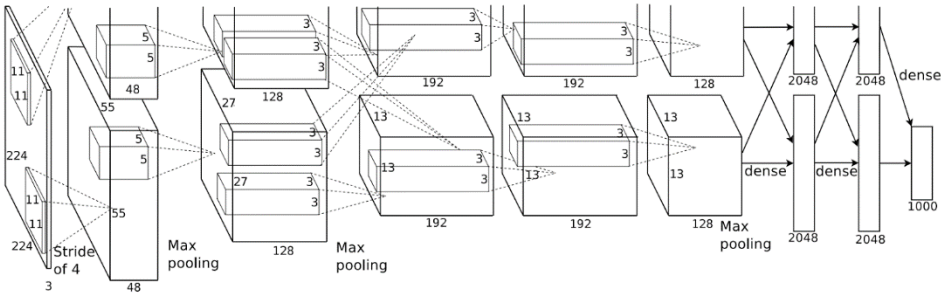
ระยะทางระหว่างเวกเตอร์ถูกคำนวณด้วยฟังก์ชันเกาส์ (Gaussian Function) ดังสมการ 2.5.2

$$\theta(u, v, s) = \exp\left(-\frac{[d(u, v)]^2}{2\sigma^2(t)}\right)$$

โดย σ คือ ความกว้างของโค้ง

2.2 การจำแนกรูปภาพ (Image Classification)

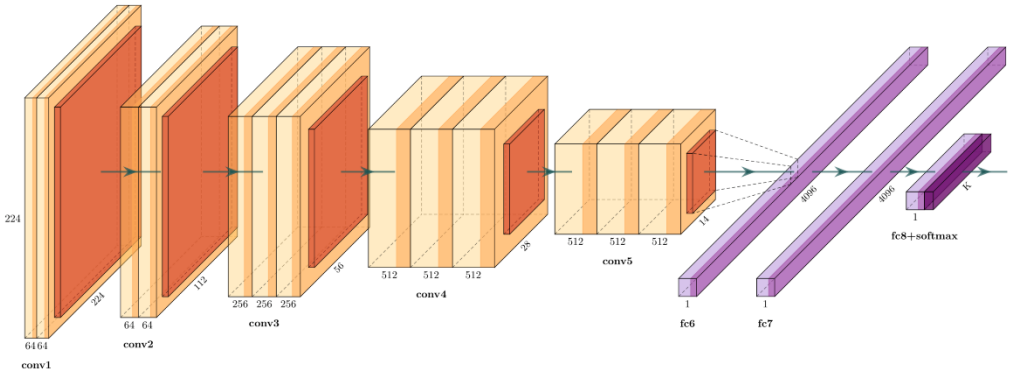
กระบวนการจำแนกรูปภาพถูกค้นคว้าและนำเสนอในหลากหลายวิธี หนึ่งในวิธีที่ได้รับความนิยมอย่างมากในปัจจุบัน คือ การจำแนกรูปภาพด้วยแบบจำลองโครงข่ายคอนโวลูชันแบบลึก (Deep Convolutional Neural Network) อาทิ AlexNet เป็นแบบจำลองสำหรับจำแนกรูปภาพสำหรับชุดข้อมูลขนาดใหญ่ (Large-scale dataset) รูปแบบคอนโวลูชันสองมิติ ประกอบด้วย 1 ชั้นข้อมูลนำเข้า 8 ชั้นคอนโวลูชัน (Convolutional layer) และ 3 ชั้นเชื่อมต่อสมบูรณ์ (Fully-connected layer) และ 1 ชั้นซอฟต์แม็กซ์ (Softmax) (ภาพที่ 2.2) สามารถจำแนกข้อมูลภาพกว่า 8.9 ล้านภาพที่มีกว่า 10,184 ประเภท ได้ความแม่นยำสูงสุด 67.4% [5]



ภาพ 3 สถาปัตยกรรมแบบจำลอง AlexNet

จากความแม่นยำของแบบจำลอง AlexNet นักวิจัยยังคงพัฒนาแบบจำลอง เพื่อเพิ่มความแม่นยำสำหรับจำแนกรูปภาพ ซึ่งหนึ่งในวิธีที่ถูกนำเสนอ คือ การเพิ่มจำนวนชั้นคอนโวลูชัน แบบจำลอง

VGG-16 และ VGG-19 [6] จึงถูกพัฒนาโดยเพิ่มจำนวนชั้นคอนโวลูชัน แบบจำลอง VGG-16 นำเสนอแบบจำลอง ประกอบด้วย 1 ชั้นข้อมูลนำเข้า 16 ชั้นคอนโวลูชัน และ 3 ชั้นเชื่อมต่อสมบูรณ์ (ภาพที่ 2.3) และแบบจำลอง VGG-19 นำเสนอแบบจำลอง ประกอบด้วย 1 ชั้นข้อมูลนำเข้า 19 ชั้นคอนโวลูชัน และ 3 ชั้นเชื่อมต่อสมบูรณ์ ทั้งสองแบบจำลองได้ทดลองปรับขนาดของตัวกรอง (Filter หรือ Kernel) และการเลื่อนตัวกรอง (Stride) พบว่าตัวกรองขนาด 3x3 ด้วยการเลื่อนทีละ 1 ขั้น ให้ผลลัพธ์ที่มีประสิทธิภาพมากกว่า

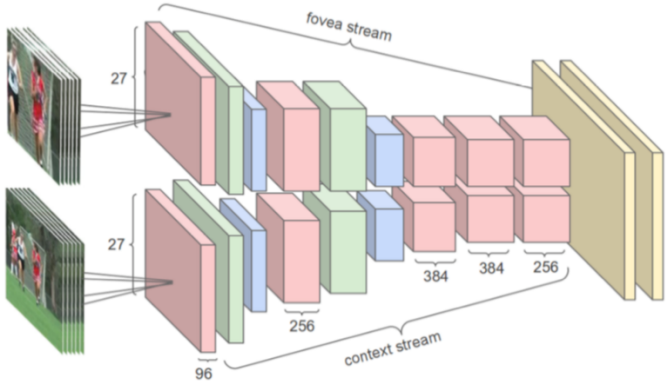


ภาพ 4 แบบจำลอง VGG-16

จากการวิเคราะห์เปรียบเทียบแบบจำลองล้ำสมัย (State-of-the-art) [7] ของแบบจำลองจำแนกรูปภาพประกอบด้วย BN-NN, GoogLeNet, Inception-v3, Inception-v4, AlexNet, BN-AlexNet, VGG16-19, ResNet8-34-50-101-152 และ Enet พบว่าแบบจำลอง E-net มีประสิทธิภาพในการจำแนกรูปภาพดีที่สุด อย่างไรก็ตามแบบจำลอง E-net เป็นแบบจำลองการเรียนรู้แบบลึก โดยถูกออกแบบเพื่อใช้จำแนกรูปภาพ สำหรับการติดตามแบบเรียลไทม์บนแอปพลิเคชันบนโทรศัพท์เคลื่อนที่บนพื้นฐานของแบบจำลอง ResNet [8]

2.3 การจำแนกวิดีโอ (Video Classification)

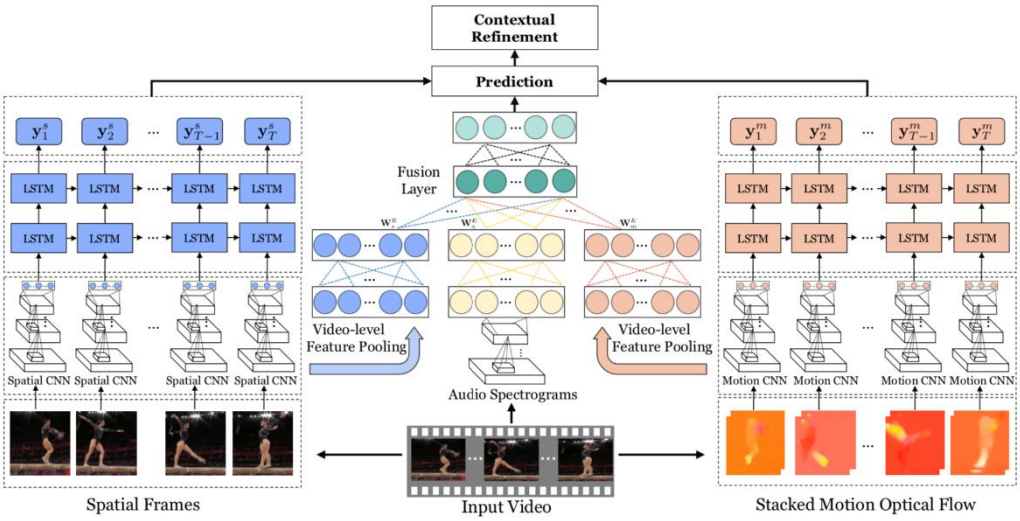
การจำแนกวิดีโอถูกพัฒนาเพื่อตอบปัญหาในหลากหลายรูปแบบ อาทิ Karpathy et al. นำเสนอแบบจำลองคอนโวลูชัน เพื่อจำแนกวิดีโอด้วยความพยายามใช้แบบจำลองจำแนกรูปภาพ โดยเลือกพิจารณาจากความละเอียดหลากหลาย (Multi-resolution) แบบ 2 สตรีม (ภาพที่ 2.4) และใช้วิธีการหลอมรวมข้อมูล (Fusing information) จากผลลัพธ์ของแบบจำลองคอนโวลูชัน โดยงานวิจัยชิ้นนี้ใช้ข้อมูล UCF-101 จำนวน 487 ประเภทในการวัดผล [9]



ภาพ 5 แบบจำลองคอนโวลูชันสำหรับการจำแนกรูปภาพหลากหลายมิติ

แบบจำลองคอนโวลูชันถูกพัฒนาและเสนอเพื่อตอบสนองความต้องการในการจำแนกวิดีโออย่างต่อเนื่อง หนึ่งในกระบวนการพัฒนาแบบจำลองคอนโวลูชันสำหรับจำแนกวิดีโอ คือ การเพิ่มมิติของแบบจำลองเป็นแบบจำลองคอนโวลูชันสามมิติ เพื่อจำแนกชุดข้อมูลขนาดใหญ่อย่าง UCF-101 สามารถจำแนกได้ด้วยความแม่นยำ 85.2% และ 90.4 เมื่อประมวลผลร่วมกับการปรับปรุงความวิถีหนาแน่น (Improved dense trajectories or iDT) [2]

อย่างไรก็ตามนอกเหนือจากการพิจารณาจากเนื้อหาวิดีโอเพียงอย่างเดียว นักวิจัยยังคงนำเสนอแบบจำลองสำหรับจำแนกวิดีโอโดยพิจารณาถึงความต่อเนื่องของวิดีโอ รูปแบบการจำแนกวิดีโอโดยพิจารณาความต่อเนื่องของวิดีโอถูกนำเสนอด้วยวิธีเพิ่มหน่วยความจำระยะสั้นแบบยาว เพื่อพิจารณาลำดับของเนื้อหา Wu, Z., et al. นำเสนอแบบจำลองคอนโวลูชันสองมิติ ร่วมกับหน่วยความจำระยะสั้นแบบยาวเพื่อจำแนกวิดีโอ โดยพิจารณาจากรูปภาพ การไหลของแสง และเสียง รูปแบบแบบจำลองที่ถูกนำเสนอถูกแสดงในรูปภาพที่ 2.5 และนำมาผลลัพธ์จากการพิจารณาสองสิ่งมาหลอมรวมที่ชั้น หลอมรวม (Fusion Layer) โดยทดสอบแบบจำลองกับชุดข้อมูล UCF-101 สามารถจำแนกได้ด้วยความแม่นยำ 93.1% และ 84.5% บนชุดข้อมูล CCV [3]



ภาพ 6 แสดงแบบจำลองจำแนกวิดีโอของ Wu, Z., et al. ด้วยเทคนิคการพิจารณาภาพ การไหลของแสง และเสียง

อย่างไรก็ตามนอกเหนือจากความพยายามในการสร้างแบบจำลองเพื่อจำแนกวิดีโอ การเลือกภาพเพื่อนำเข้าสู่กระบวนการจำแนก ซึ่งเป็นปัจจัยที่ส่งผลต่อความแม่นยำในการจำแนก การพิจารณาภาพเพื่อใช้สำหรับจำแนกวิดีโอมีหลากหลายวิธี อาทิ การพิจารณาภาพทั้งหมด การลงคะแนนตามเสียงส่วนมาก [10] การเลือกภาพแรกเท่านั้น และการเลือกแบบสุ่ม [11]

2.4 การจัดกลุ่มรูปภาพ (Image Clustering)

การจัดกลุ่มรูปภาพเป็นเทคนิคหนึ่งในการจำแนกรูปภาพไปในกลุ่มต่าง ๆ หนึ่งในเทคนิคที่ได้รับความนิยมของการจัดกลุ่มคือ K-means Clustering ซึ่งเป็นเทคนิคการเรียนรู้แบบไม่มีผู้สอน Dhanachandra, N., K. Manglem, and Y.J. Chanu นำเสนอวิธีการจัดกลุ่มรูปภาพด้วยเทคนิค K-means เพื่อใช้จัดกลุ่มรูปภาพทางการแพทย์ (เม็ดเลือด) พร้อมทั้งเพิ่มคุณภาพของภาพระหว่างกระบวนการจัดกลุ่ม บนพื้นฐานภาพโทนสีเทา (Gray Scale) [12]

อย่างไรก็ตามการจัดกลุ่มรูปภาพที่มีความซับซ้อนของค่าสี ที่ไม่เป็นข้อมูลเชิงเส้นด้วยเทคนิคพื้นฐาน Zhu, W., J. Lu, and J. Zhou นำเสนอกระบวนการจัดกลุ่มรูปภาพด้วยวิธีการจัดกลุ่มพื้นที่แบบไม่เชิงเส้น (Non-linear Subspace Clustering: NSC) ด้วยกระบวนการโครงข่ายประสาทเทียมแบบไปข้างหน้าเพื่อระบุพื้นที่ (Feed-Forward Neural Network) [13]

การใช้โครงข่ายประสาทเทียมเพื่อจัดกลุ่มรูปภาพ ถูกนำมาใช้อย่างต่อเนื่องหนึ่งในเทคนิคที่ได้รับความนิยมสำหรับจัดกลุ่มรูปภาพและเป็นการเรียนรู้แบบไม่มีผู้สอน คือ แผนที่จัดระเบียบด้วยตนเอง Mo, H., et al นำเสนอกระบวนการจัดกลุ่มรูปภาพด้วยเทคนิคแผนที่จัดกลุ่มด้วยตนเอง มุ่งเน้นที่การจำแนกตามความหลากหลายของค่าสี โดยเพิ่มประสิทธิภาพของกลุ่มข้อมูลด้วยเทคนิคการปรับแต่งกลุ่มข้อมูล (Cluster refinement) และการรวมกลุ่มข้อมูล (Cluster merging) [14]

2.5 มัธยฐาน (Median)

มัธยฐาน คือ ค่าที่อยู่ระหว่างกลางค่าที่มากที่สุด และค่าที่น้อยที่สุด ของชุดข้อมูลที่ถูกเรียงลำดับ อาจถูกเรียกว่า ค่ากลาง (Middle Value) ยกตัวอย่างเช่น ชุดข้อมูล [1, 2, 3, 4, 5, 6, 7] ค่ากึ่งกลางของชุดข้อมูลชุดนี้ คือ 4 หรือ ชุดข้อมูล [11, 12, 13, 14] ค่ากึ่งกลางของชุดข้อมูลชุดนี้ คือ (12+13) / 2 เท่ากับ 12.5 โดยค่ากลางมักถูกนำมาใช้เป็นเครื่องมือพื้นฐานสำหรับอธิบายคุณสมบัติของชุดข้อมูลทางสถิติ และความน่าจะเป็นเทียบกับค่าเฉลี่ย (Mean) ค่ากลางสามารถนำมาใช้กับชุดข้อมูลที่มีมากกว่าหนึ่งมิติ เพื่อลดขนาดมิติของข้อมูลได้

ตัวอย่างการใช้ค่ากลางเพื่อลดมิติของข้อมูลสองมิติ

ชุดข้อมูล [[11, 12, 13, 14, 15],
[16, 17, 18, 19, 20],
[21, 22, 23, 24, 25]]

ผลลัพธ์จากการหาค่ากลาง [16, 17, 18, 19, 20]

ตัวอย่างการใช้ค่ากลางเพื่อลดมิติของข้อมูลสามมิติ

ชุดข้อมูล [[[11, 12, 13], [14, 15, 16], [17, 18, 19]],
[[20, 21, 22], [23, 24, 25], [26, 27, 28]],
[[29, 30, 31], [32, 33, 34], [35, 36, 37]]]

ผลลัพธ์จากการหาค่ากลางทางลึก [[20, 21, 22],
[23, 24, 25],
[26, 27, 28]]

2.6 ฮิสโตแกรม (Histogram)

ฮิสโตแกรมเป็นกราฟแสดงแสดงความสัมพันธ์และความถี่ของข้อมูล เพื่อตรวจสอบการกระจายของข้อมูล ฮิสโตแกรมสามารถถูกนำมาประยุกต์ใช้พิจารณารูปภาพ เพื่อนดูการกระจายของค่าสี และโทนสีภาพได้เช่นกัน โดยพิจารณาจากค่าความสว่างทั้งหมด 256 ระดับ (0 - 255) เทียบกับพิกเซลของภาพ

2.7 การเลือกคุณสมบัติ (Feature Selection)

การเลือกคุณสมบัติเป็นวิธีหนึ่งในกระบวนการเรียนรู้ด้วยเครื่อง โดยการเลือกชุดข้อมูลเพื่อเป็นตัวแทนของชุดข้อมูลทั้งหมด ในการสร้างแบบจำลองเลือกคุณสมบัติมักใช้ลดขนาดมิติของข้อมูลเพื่อหลีกเลี่ยงคำสาปของข้อมูล (Curse of dimensionality) อีกทั้งยังมีข้อดีในการเลือกคุณสมบัติ คือ ใช้เวลาประมวลผลกับข้อมูลน้อยลง สามารถช่วยลดความซ้ำซ้อนของข้อมูล และช่วยลดความพอดีกับข้อมูลเรียนรู้เกินไป (Overfitting) เทคนิคการเลือกคุณสมบัติที่ถูกนำมาใช้ในงานวิจัยชิ้นนี้ ได้แก่ การลงคะแนนตามเสียงส่วนมาก การลงคะแนนตามเสียงตามจำนวนมาก และค่าเอ็นโทรพี

2.8 การลงคะแนน (Voting)

การลงคะแนน คือ การกำหนดคุณสมบัติของชุดข้อมูลตามคุณสมบัติที่มีมากที่สุดของแต่ละชุดข้อมูลย่อย งานวิจัยนี้พิจารณาการลงคะแนน เพื่อกำหนดคุณสมบัติของชุดข้อมูลย่อยด้วย 2 รูปแบบ คือ

2.8.1 การเลือกตามเสียงส่วนมาก

การเลือกตามเสียงส่วนมาก คือ การกำหนดคุณสมบัติของชุดข้อมูลตามคุณสมบัติของข้อมูลที่มีมากที่สุดในชุดข้อมูลนั้น ๆ แต่จำนวนคุณสมบัติที่มากที่สุดจะต้องมีจำนวนมากกว่ากึ่งหนึ่งของจำนวนข้อมูลทั้งหมดในชุดข้อมูลนั้น ๆ

ตัวอย่างของการเลือกตามเสียงข้างมากของตัวอย่างข้อมูลตัวอักษร A, B และ C ในชุดข้อมูล ABBBCABBBA

ตัวอักษร	จำนวน
A	3
B	6
C	1

จากข้อมูลทั้งหมด 10 ตัวอักษร พบว่า ตัวอักษร A มีจำนวนมากที่สุด คือ 6 และมากกว่า กึ่งหนึ่งของจำนวนข้อมูลทั้งหมด จึงสามารถอธิบายได้ว่ากลุ่มของชุดข้อมูลมีคุณสมบัติ A ด้วยการเลือกตามเสียงส่วนมาก

ตัวอย่างของการเลือกตามเสียงข้างมากของตัวอย่างข้อมูลตัวอักษร A, B และ C ของชุดข้อมูล ABBBCACBBA

ตัวอักษร	จำนวน
A	3
B	5
C	2

จากข้อมูลทั้งหมด 10 ตัวอักษร พบว่า ตัวอักษร A มีจำนวนมากที่สุด คือ 5 แต่ไม่มีค่าใดที่มีค่าเกินกึ่งหนึ่งจึงไม่สามารถอธิบายได้หาคุณสมบัติที่เป็นเสียงส่วนมากได้

2.8.2 การเลือกตามจำนวนมาก

การเลือกคุณสมบัติตามจำนวนมากที่สุด เป็นการเลือกโดยไม่ต้องคำนึงถึงคุณสมบัติที่มีมากที่สุดจำเป็นต้องได้เสียงมากกว่ากึ่งหนึ่งของจำนวนชุดข้อมูลทั้งหมดหรือไม่ ยกตัวอย่างเช่น ชุดข้อมูล ABBBCACBBA พบว่าตัวอักษร A มีจำนวนเท่ากับ 3, ตัวอักษร B มีจำนวนเท่ากับ 5 และตัวอักษร C มีจำนวนเท่ากับ 2 สามารถอธิบายได้ว่า ชุดข้อมูลมีคุณสมบัติเป็น A ตามจำนวนที่มากที่สุด แต่ไม่ได้เป็นคุณสมบัติตามเสียงส่วนมาก

2.9 ค่าเอ็นโทรพี

ค่าเอ็นโทรพี เป็นค่าที่ใช้อธิบายความน่าจะเป็นของชุดข้อมูล เพื่อเลือกหรือกำหนดคุณสมบัติของชุดข้อมูล โดยค่าเอ็นโทรพียิ่งสูง จะบอกถึงความไม่แน่นอนหรือความวุ่นวายของคุณสมบัติของชุดข้อมูลนั้น ๆ ค่าเอ็นโทรพีถูกนำมาประยุกต์ใช้ในกระบวนการจำแนกชุดข้อมูล อาทิ ต้นไม้ตัดสินใจ (Decision Tree) เป็นต้น การคำนวณค่าเอ็นโทรพีของชุดข้อมูลจะคิดจากค่าความน่าจะเป็นของชุดข้อมูล โดยใช้สูตร 2.2

$$Entropy = \sum_{i=1}^{C} -p_i \log_2 p_i$$

โดย p_i คือ ค่าความน่าจะเป็นของคุณสมบติที่ i

C คือ จำนวนคุณสมบัติทั้งหมด

บทที่ 3
ระเบียบวิธีการ

3.1 การรวบรวมและเตรียมชุดข้อมูล

งานวิจัยชิ้นนี้มุ่งเน้นที่การจำแนกวิดีโอจากรายการโทรทัศน์ วิเคราะห์ภาพรายการโทรทัศน์ต่าง ๆ จากวิดีโอรายการโทรทัศน์ด้วยวิธีการแสดงผล พบว่ารายการโทรทัศน์แต่ละช่อง มีสิ่งที่เป็นคุณลักษณะของรายการโทรทัศน์ที่ใกล้เคียงกัน คือ สัญลักษณ์ของช่องที่ออกอากาศ โดยพื้นฐานจะมีสัญลักษณ์เดียวกันในทุกรายการ และสัญลักษณ์ของรายการโทรทัศน์นั้น ๆ งานวิจัยชิ้นนี้จึงได้เก็บรวบรวมวิดีโอรายการโทรทัศน์ที่ออกอากาศ และวิดีโอรายการอื่น ๆ ที่ปรากฏบนแพลตฟอร์มสาธารณะ

3.1.1 ชุดข้อมูล

ชุดข้อมูลเป็นเครื่องมือสำหรับการทดสอบกระบวนการจำแนกวิดีโอ ถูกนำเสนอเพื่อใช้เป็นตัวชี้วัดประสิทธิภาพในการจำแนกวิดีโอของแบบจำลองต่าง ๆ ซึ่งชุดข้อมูลแต่ละประเภทถูกออกแบบมาเพื่อวัตถุประสงค์ในการวัดผลที่ต่างกัน อาทิ วัดผลการรู้จำเคลื่อนไหว (Action Recognition) และ วัดผลการรู้จำประเภท (Categorizes Recognition) เป็นต้น ชุดข้อมูลสำหรับวัดผลการรู้จำการเคลื่อนไหวที่ได้รับความนิยมอย่างมาก คือ ชุดข้อมูล UCF-101 ซึ่งประกอบไปด้วย 13,320 วิดีโอคลิป จาก 101 ประเภท [15] และชุดข้อมูลขนาดใหญ่ที่สุด ณ ตอนนี้ ซึ่งถูกนำเสนอโดย YouTube ซึ่งเป็นแพลตฟอร์มที่ให้บริการวิดีโอคอนเทนต์ คือ YouTube8m [16]

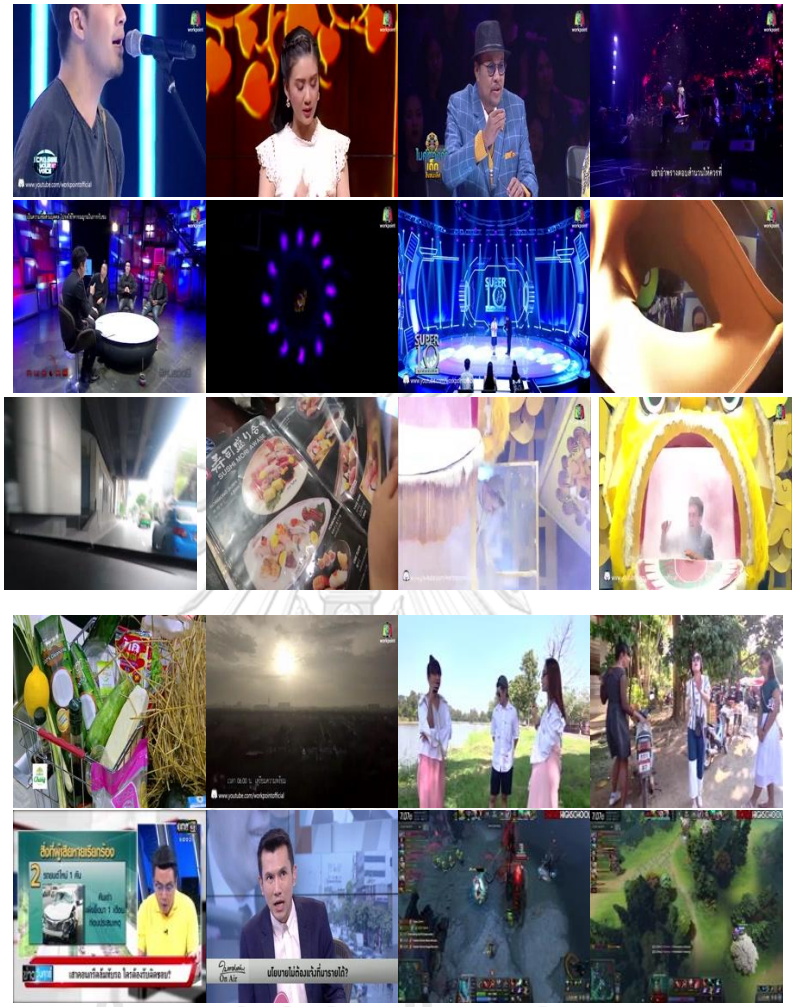
3.1.2 การรวบรวมข้อมูล

งานวิจัยชิ้นนี้รวบรวมข้อมูลงานรายการโทรทัศน์จำนวน 18 รายการ จำนวน 912 รายการ ซึ่งทุกรายการโทรทัศน์มีความยาวมากกว่า 60 นาที จากแพลตฟอร์มยูทูบ (YouTube) โดยใช้ไพทอนไลบรารี่ (Python Library) ชื่อ youtube-dl ด้วยวิธีการสร้างเพลยลิสต์ (Playlist) ของรายการโทรทัศน์ทั้งหมดบนยูทูบ และดำเนินการดาวโหลดเพลย์ลิสต์ที่ต้องการทั้งหมด

ตัวอย่างชุดคำสั่งสำหรับการดาวโหลดวิดีโอรายการโทรทัศน์จากแพลตฟอร์มยูทูป

```
youtube-dl -f  format -i -o "to_path" --download-archive archive.txt
```

ข้อมูลวิดีโอคลิปรายการโทรทัศน์ จำนวน 18 รายการ รวมทั้งสิ้น 912 ตัวอย่างของวิดีโอรายการถูกแสดงในรูปภาพ 3.1



ภาพ 7 ตัวอย่างชุดข้อมูลรายการโทรทัศน์

3.1.3 การเตรียมชุดข้อมูล

การเตรียมชุดข้อมูลของงานวิจัยชิ้นนี้ เป็นกระบวนการสกัดภาพฉากจากวิดีโอรายการโทรทัศน์ โดยสกัดเฉพาะภาพฉากแรกของแต่ละวินาทีเท่านั้น เนื่องจากการวิเคราะห์ความแตกต่างของภาพฉากในวินาทีเดียวกันด้วยวิธีแสดงผล พบว่าแต่ละฉากไม่มีความแตกต่างกันมากนัก ในกระบวนการสกัดภาพฉาก แต่ละภาพฉากจะถูกลดขนาดของภาพ เหลือที่ขนาดความกว้าง 224 พิกเซล และความสูง 224 พิกเซล งานวิจัยชิ้นนี้ใช้เฟรมเวิร์ก ffmpeg สำหรับสกัดภาพฉากจากวิดีโอ

ตัวอย่างชุดคำสั่งที่ถูกใช้เพื่อสกัดภาพฉากของงานวิจัยชิ้นนี้

```
ffmpeg -i "file_name" -s 224x224 -vf fps=1 to_path/%d.jpg
```

วิดีโอคลิปทั้ง 18 ประเภท ถูกแบ่งเป็น วิดีโอคลิปย่อยที่มีความยาวคลิปละ 500 วินาที โดยไม่สนใจความต่อเนื่องของเนื้อหา ติดป้ายกำกับ (Labeling) ชุดข้อมูลรายการหนึ่งรายการ เพื่อจำแนกวิดีโอด้วยวิธีจำแนกฐานสอง ดังแสดงในตารางที่ 3.1 และแยกส่วนของวิดีโอคลิปเพื่อวัดผล โดยแยกภาพลำดับแรกของแต่ละวินาที เป็นจำนวน 500 ภาพต่อชุด แต่ละภาพมีความกว้าง 224 พิกเซล และความยาก 224 พิกเซล รวมทั้งสิ้นเป็นจำนวนทั้งสิ้น 912 ชุดข้อมูล และ ติดป้ายกำกับสำหรับจำแนกตามคลาสของรายการ ซึ่งแสดงในตารางที่ 1 เพื่อใช้สำหรับทดสอบ

ตาราง 1 ตัวอย่างชุดข้อมูลที่ถูกติดป้ายกำกับเพื่อจำแนกแบบฐานสองในกรณีเป็นตัวอย่างรายการที่ 2

รายการ	**จำนวนคลิป**	**ป้ายกำกับ**
1	48	0
2	59	1
3	52	0
4	61	0
5	55	0
6	17	0
7	43	0
8	37	0
9	70	0
10	61	0
11	53	0
12	74	0
13	43	0
14	56	0
15	61	0
16	58	0
17	10	0
18	54	0

ตาราง 2 ชุดข้อมูลที่ถูกติดป้ายกำกับเพื่อจำแนกตามประเภท

รายการ	จำนวนคลิป	ป้ายกำกับ
1	48	0
2	59	1
3	52	2
4	61	3
5	55	4
6	17	5
7	43	6
8	37	7
9	70	8
10	61	9
11	53	10
12	74	11
13	43	12
14	56	13
15	61	14
16	58	15
17	10	16
18	54	17

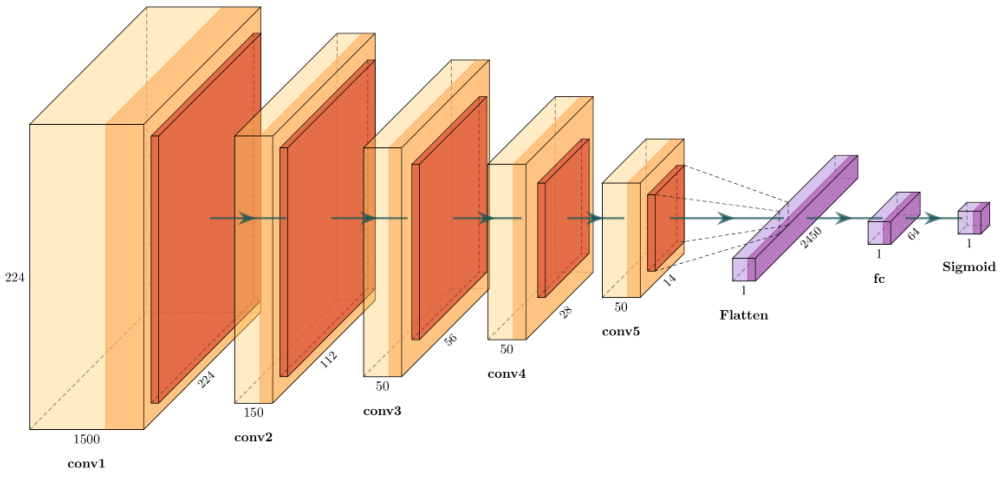
3.2 กระบวนการจำแนกวิดีโอด้วยแบบจำลองคอนโวลูชันสองมิติ

จากการศึกษาแบบจำลองล้ำสมัย สำหรับการจำแนกวิดีโอในปัจจุบันพบว่าหลายแบบจำลองถูกออกแบบมาเพื่อจำแนกชุดข้อมูลขนาดใหญ่ ซึ่งบางแบบจำลองใช้ระยะนานในกระบวนการเรียนรู้ ยกตัวอย่างเช่นแบบจำลองคอนโวลูชันสามมิติสำหรับการจำแนกวิดีโอ หรือแบบจำลองมิติสัมพันธ์ ที่ประกอบด้วยแบบจำลองคอนโวลูชันสองมิติ และหน่วยความจำแบบสั้นระยะยาว งานวิจัยชิ้นนี้จึงต้องการนำเสนอแบบจำลองคอนโวลูชันที่มีขนาดเล็ก ที่ยังคงสามารถจำแนกรายการโทรทัศน์ได้

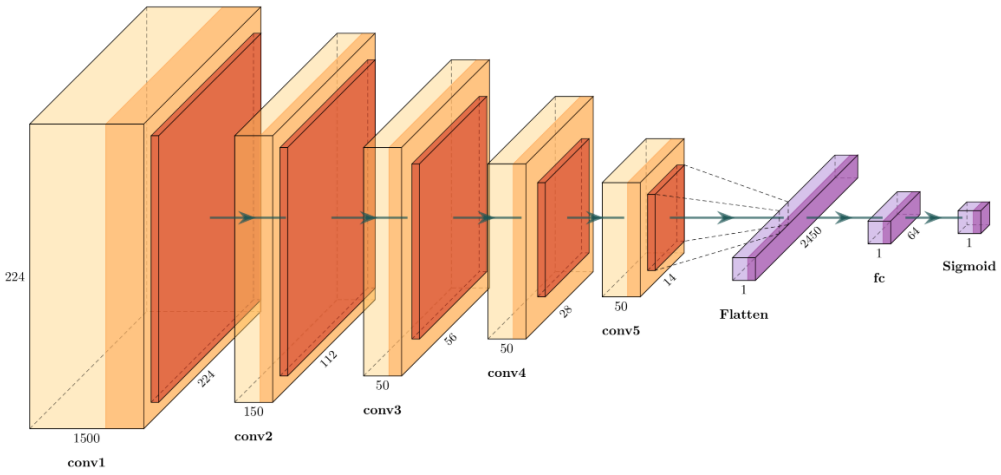
เช่นกัน จากชุดข้อมูลในงานวิจัยชิ้นนี้นำเสนอแบบจำลองสำหรับจำแนกรายการ 1 รายการจากรายการทั้ง 17 ด้วยการจำแนกแบบเลขฐานสอง ด้วยฟังก์ชันซิกมอยด์

3.2.1 แบบจำลองคอนโวลูชันสองมิติแบบคำนวณทุกฉาก

จากการศึกษาแบบจำลองคอนโวลูชันสองมิติ ซึ่งเป็นส่วนมากเป็นแบบจำลองล้ำสมัยสำหรับจำแนกรูปภาพ งานวิจัยชิ้นนี้เลือกนำแบบจำลอง VGG-16 (ภาพที่ 3.2) มาปรับปรุงชั้นคอนโวลูชันเพื่อจำแนกวิดีโอ โดยปรับปรุงชั้นข้อมูลนำเข้าจากจำนวน 3 ช่อง เป็น 1,500 ช่อง เพื่อรองรับภาพฉากจำนวน 500 ภาพ โดยแต่ละภาพประกอบด้วย 3 ช่องสี (RGB) และลดจำนวนชั้นข้อมูลคอนโวลูชันในทุก ๆ ชั้น ทำให้เหลือชั้นคอนโวลูชันจำนวน 5 ชั้นคอนโวลูชัน และชั้นเชื่อมต่อสมบูรณ์ จำนวน 1 ชั้น ก่อนเข้าสู่ชั้นฟังก์ชันซิกมอยด์ เพื่อทำนาย โดยรูปแบบโครงข่ายของแบบจำลองคอนโวลูชันแบบสองมิติถูกแสดงในภาพที่ 8



ภาพ 8 รูปแบบโครงข่ายแบบจำลองคอนโวลูชันสองมิติ สำหรับชั้นข้อมูลนำเข้าขนาด 224 x 224 x 3 โหนด



ภาพ 9 รูปแบบโครงข่ายแบบจำลองคอนโวลูชันสองมิติ สำหรับชั้นข้อมูลนำเข้าขนาด 224 x 224 x 1,500 โหนด

กระบวนการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติเลือกใช้การเพิ่มประสิทธิภาพด้วยอัลกอริทึมอดัม และดำเนินการเรียนรูปจำนวน 20 ยุค โดยกำหนดชุดการเรียนรู้ ที่ขนาด 50 ข้อมูล ให้มีการสลับชุดข้อมูลอยู่เสมอ และเลือกพิจารณาจากรอบที่มีผลค่าความผิดพลาดน้อยที่สุด

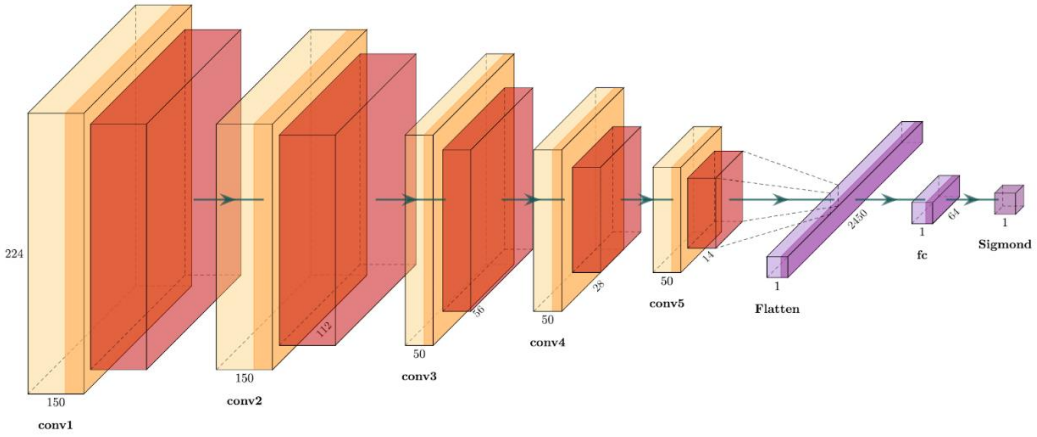
3.2.2 กระบวนการลดจำนวนภาพด้วยตัวกรองมัธยฐาน

จากแบบจำลองคอนโวลูชันสองมิติแบบคำนวณทุกภาพ งานวิจัยชิ้นนี้นำเสนอการพิจารณาเพิ่มการประมวลผลชุดข้อมูล เพื่อเป็นกระบวนการลดจำนวนชุดข้อมูลสำหรับการเรียนรู้ โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ของแบบจำลอง และลดระยะเวลาในการเรียนรู้ของแบบจำลอง ซึ่งงานวิจัยชิ้นนี้นำเสนอการลดภาพด้วยวิธีการใช้ตัวกรองมัธยฐานเพื่อลดจำนวนภาพ โดยการคำนวณค่ากลางของภาพที่นำชุดข้อมูลภาพมาซ้อนกันในแนวลึก ซึ่งจะพิจารณาแยกตามค่าสีที่ซ้อนกัน คือ ค่าสี จากการทดลองการใช้ตัวกรองมัธยฐานเพื่อหาจำนวนภาพเหมาะสมที่จะผ่านตัวกรองงานวิจัยชิ้นนี้ทดสอบการเลือกจำนวนภาพระหว่าง 10 และ 20 ภาพ และพิจารณาด้วยวิธีการแสดงผลเปรียบเทียบของผลลัพธ์ภาพภายหลังจากผ่านตัวกรอง (ภาพที่ 10) งานวิจัยชิ้นนี้เลือกใช้จำนวน 10 ภาพ เป็นสำหรับคำนวณในตัวกรองมัธยฐาน ทำให้สามารถลดจำนวนภาพจาก 500 ภาพ เหลือเพียง 50 ภาพ โดยแต่ละภาพประกอบด้วย 3 ช่อง

ภาพ 10 ภาพซ้าย ภาพที่เกิดจากการประมวลผลตัวกรองมัธยฐานด้วย 10 ภาพ
ภาพขวา ภาพที่เกิดจาการประมวลผลตัวกรองมัธยฐานด้วย 20 ภาพ

3.2.3 ปรับปรุงแบบจำลองคอนโวลูชันสองมิติหลังจากเพิ่มกระบวนการลดจำนวนภาพ

จากการนำเสนอการปรับปรุงข้อมูลนำเข้าด้วยตัวกรองมัธยฐาน งานวิจัยชิ้นนี้จึงดำเนินการปรับปรุงแบบจำลองคอนโวลูชันสองมิติจากการคำนวณทุกภาพ โดยลดจำนวนชั้นข้อมูลนำเข้าจาก 1,500 ช่องเหลือเพียง 150 ช่อง โดยคงชั้นคอนโวลูชัน ชั้นเชื่อมต่อสมบูรณ์ และชั้นปรับเทียบหรือชั้นเอาต์พุตไว้เท่าเดิม รูปแบบโครงข่ายของแบบจำลองคอนโวลูชันสำหรับจำแนกรายการโทรทัศน์ที่ผ่านตัวกรองมัธยฐานถูกแสดงในภาพที่ 11



ภาพ 11 รูปแบบโครงข่ายแบบจำลองคอนโวลูชันสองมิติ ขนาด 224 x 224 x 1500 โหนด

กระบวนการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติ ในงานวิจัยนี้เลือกใช้อัลกอริทึมการเพิ่มประสิทธิภาพ ด้วยอัลกอริทึมอดัม และดำเนินการเรียนรูปจำนวน 20 ยุค โดยกำหนดชุดกาเรียนรู้ ที่ขนาด 50 ข้อมูล ให้มีการสลับชุดข้อมูลอยู่เสมอ และเลือกพิจารณาจากรอบที่มีผลค่าความผิดพลาดน้อยที่สุด

3.3 การจัดกลุ่มภาพด้วยวิธีการจัดกลุ่ม

จากกระบวนการจัดการภาพด้วยตัวกรองมัธยฐาน งานวิจัยชิ้นนี้เลือกพิจารณาการจัดกลุ่มภาพฉากด้วยสมมติฐานที่ว่าการพิจารณาภาพที่มีคุณลักษณะใกล้เคียงกัน จะสามารถช่วยเพิ่มประสิทธิภาพในการจำแนกวิดีโอ งานวิจัยชิ้นนี้เลือกใช้วิธีการเรียนรู้ด้วยแผนที่จัดการตนเอง ซึ่งเป็นกระบวนการเรียนรู้โดยการปรับแผนที่โครงข่ายประสาทเทียม โดยการระบุจำนวนโหนดของโครงข่ายประสาทเทียมเพื่อสร้างพื้นที่สำหรับคำนวณและทำนายข้อมูลนำเข้าอื่น

จากการทดสอบการจัดกลุ่มรูปภาพด้วยวิธีแผนที่จัดการตนเอง งานวิจัยชิ้นนี้นำเสนอ ขนาดของแผนที่จัดการตัวเองที่ขนาดความกว้าง 40 โหนด และความยาว 40 โหนด ซึ่งทำให้เกิดพื้นที่ขนาด 1,600 โหนด งานวิจัยชิ้นนี้เลือกใช้ฟังก์ชันเกาส์ เพื่อคำนวณหาโหนดเพื่อนบ้าน

3.3.1 การจัดกลุ่มภาพด้วยค่าสี (RGB)

การจัดกลุ่มภาพด้วยค่าสีในงานวิจัยชิ้นนี้ ดำเนินการปรับปรุงขนาดภาพ โดยลดขนาดความกว้างและความสูงของภาพจาก 224 x 224 จำนวน 3 ค่าสี เหลือเพียง 28 x 28 และปรับปรุงรูปภาพจากเมตริกซ์ขนาด 28 x 28 x 3 เป็นเวกเตอร์ขนาด 2,352 x 1 เพื่อนำมาจัดกลุ่มด้วยวิธีแผนที่การจัดกลุ่มด้วยตนเอง

งานวิจัยนี้ดำเนินการปรับปรุงทั้งสิ้น 20 ยุค ด้วยขนาดกลุ่มข้อมูลนำเข้า 50 ข้อมูล ด้วยวิธีสลับ ชุดข้อมูลในทุก ๆ ยุค

3.3.2 การจัดกลุ่มภาพด้วยค่าฮิสโตแกรม (Histogram)

งานวิจัยนี้พิจารณาค่าฮิสโตแกรมซึ่งสามารถแสดงความหลากหลายของค่าสีของภาพได้ ด้วยสมมติฐานว่ารายการโทรทัศน์รายการเดียวกัน มักจะมีโทนสีภาพที่ใกล้เคียงกันเสมอ งานวิจัยนี้คำนวณฮิสโตแกรมจากค่ารูปภาพของชุดข้อมูลแต่ละชุด ได้ผลลัพธ์เป็นเวกเตอร์ขนาด 255 x 1 หลังจากนั้นจึงนำมาจัดกลุ่มด้วยวิธีแผนที่จัดกลุ่มด้วยตนเอง งานวิจัยนี้ดำเนินการปรับปรุงทั้งสิ้น 20 ยุค ด้วยขนาดกลุ่มข้อมูลนำเข้า 50 ข้อมูล ด้วยวิธีสลับ ชุดข้อมูลในทุก ๆ ยุค

3.4 การจำแนกวิดีโอรายการโทรทัศน์จากกลุ่มภาพ

การจำแนกวิดีโอรายการโทรทัศน์โดยพิจารณาจากกลุ่มภาพฉาก ที่เป็นผลลัพธ์มาจากการหากลุ่มข้อมูลด้วยแบบจำลองแผนที่การจัดระเบียบด้วยตนเอง งานวิจัยนี้นำเสนอกระบวนการพิจารณาใน 4 รูปแบบ คือ จำแนกวิดีโอด้วยการลงคะแนน จำแนกวิดีโอด้วยการพิจารณาจากค่าเอ็นโทรพี จำแนกวิดีโอด้วยการสร้างแบบจำลองโครงข่ายประสาทเทียม และจำแนกวิดีโอด้วยหน่วยความจำระยะสั้นแบบยาว

3.4.1 จำแนกด้วยการลงคะแนน

การจำแนกด้วยการลงคะแนนในงานวิจัยนี้ ถูกพิจารณาด้วยการลงคะแนนสองแบบ คือ เสียงส่วนมาก และเสียงตามจำนวนมาก โดยจะใช้การลงคะแนนในในสองระดับ คือ การลงคะแนนเพื่อจำแนกประเภทของกลุ่มข้อมูล และการลงคะแนนเพื่อจำแนกประเภทของวิดีโอ

การลงคะแนนเพื่อจำแนกประเภทของกลุ่มข้อมูลที่เป็นผลลัพธ์มากจากกระบวนการจัดกลุ่มด้วยวิธีแผนที่การจัดระเบียบด้วยตนเอง โดยทดสอบกับผลลัพธ์ด้วยการลงคะแนนทั้งวิธีเสียงส่วนมาก และเสียงตามจำนวนมาก

3.4.2 จำแนกด้วยการพิจารณาค่าเอ็นโทรพี

การพิจารณาค่าเอ็นโทรพีในงานวิจัยนี้ ถูกนำมาใช้เพื่อเป็นค่าน้ำหนักในการจำแนกประเภทวิดีโอภายหลังจากการกำหนดประเภทของกลุ่มด้วยวิธีการลงคะแนน โดยงานวิจัยนี้พิจารณาค่าเอ็นโทรพีในสองมิติของค่าเอ็นโทรพี

คำนวณค่าเอ็นโทรพีโดยตรง คือการให้น้ำหนักกับกลุ่มข้อมูลที่มีความวุ่นวายมาก มากกว่ากลุ่มข้อมูลที่มีความวุ่นวายน้อย แสดงในสูตรที่ 3.1

$$Entropy = \sum_{i=1}^{C} -p_i \log_2 p_i$$

โดย p คือ ค่าความน่าจะเป็นของประเภทข้อมูล

c คือ จำนวนประเภทประเภทของชุดข้อมูล

ตัวอย่างแสดงการคำนวณประเภทของกลุ่มข้อมูลด้วยภาพที่ได้จากการจัดระเบียบด้วยตนเอง

รายการ	**จำนวนภาพ**
1	5
2	6
3	7
4	8
5	9
6	2
7	1
รวม	38

จากชุดข้อมูลสามารถคำนวณค่าเอ็นโทรพีได้ 0.777

และพิจารณาจากการกลับค่าเอ็นโทรพี คือการให้น้ำหนักกับกลุ่มที่มีค่าเอ็นโทรพีน้อย มากกว่ากลุ่มข้อมูลที่มีความวุ่นวายมาก แสดงในสูตรที่ 3.2

$$Inverted = 1 - Entropy$$

จากชุดข้อมูลก่อนหน้านี้ จะสามารถมารถคำนวณค่าส่วนกลับได้ 0.223

การนำค่าเอ็นโทรพีและส่วนกลับค่าเอ็นโทรพีมาใช้จำแนกประเภทวิดีโอ จะปรับค่าตามน้ำหนักของแต่ละกลุ่มข้อมูลด้วยกระบวนการสูงสุดต่ำสุด ตามอัตราส่วนของภาพในแต่ละกลุ่มข้อมูล แสดงในสูตรที่ 3.3

$$weight = \frac{C}{F} \times V$$

โดย C คือ จำนวนภาพในกลุ่มข้อมูลนั้นๆ

F คือ จำนวนภาพทั้งหมดที่ถูกนำมาจัดกลุ่ม

V คือ ค่าเอ็นโทรพี หรือค่ากลับความวุ่นวาย

3.4.3 จำแนกด้วยแบบจำลองโครงข่ายประสาทเทียม

การจำแนกรายการโทรทัศน์ด้วยแบบจำลองโครงข่ายประสาทเทียม ในงานวิจัยนี้นำเสนอแบบจำลองโครงข่ายประสาทเทียมแบบหนึ่งมิติ ใช้วิธีการสร้างข้อมูลนำเข้าขนาด 40 x 40 จากแผนที่การจัดระเบียบด้วยตนเอง โดยพิจารณาค่าของแต่ละตำแหน่งในข้อมูลนำเข้าจากตำแหน่งของภาพบนแผนที่การจัดระเบียบด้วยตนเอง ให้ตำแหน่งของข้อมูลนำเข้ามีค่าเป็นหนึ่งก็ต่อเมื่อภาพนั้นอยู่บนโหนดใด ๆ บนแผนที่การจัดระเบียบด้วยตนเอง และให้ตำแหน่งของข้อมูลนำเข้ามีค่าเป็น 0 เมื่อไม่มีตำแหน่งปรากฏบนแผนที่จัดระเบียบ ยกตัวอย่างเช่น ภาพที่ 1 และ 2 จัดกลุ่มได้ตำแหน่งที่ (1,2) ของแผนที่การจัดระเบียบตนเอง ดังนั้น ตำแหน่ง (1,2) ของเมตริกซ์นำเข้า จะมีค่าเป็น 1

การออกแบบโครงข่ายประสาทเทียมเพื่อจำแนกวิดีโอจากข้อมูลผลลัพธ์ที่ได้จากการจัดกลุ่มในรูปแบบหนึ่งมิติ งานวิจัยนี้กำหนดชั้นข้อมูลนำเข้าขนาด 1,600 โหนด และปรับปรุงมิติข้อมูลนำเข้าจากขนาด 40 x 40 เป็น 1,600 x 1

3.4.4 จำแนกด้วยหน่วยความจำระยะสั้นแบบยาว

การจำแนกวิดีโอในขั้นตอนสุดท้ายด้วยหน่วยความจำระยะสั้นแบบยาวในงานวิจัยนี้ได้นำเสนอแบบจำลองหน่วยความจำระยะสั้นแบบยาว โดยพิจารณาจากลำดับการเกิดขึ้นของกลุ่มของชุดข้อมูลโดยเข้ารหัสประเภทชุดข้อมูลก่อนนำไปเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว

3.5 การวัดผลและประเมินผลแบบจำลองด้วยชุดข้อมูลรายการโทรทัศน์

การวัดผลชุดข้อมูลที่งานวิจัยนี้จัดทำขึ้นใช้วิธีการตรวจสอบไขว้ จำนวน 5 ชุดข้อมูล เพื่อวัดผลประสิทธิภาพของแบบจำลองทั้งหมด คือ แบบจำลองคอนโวลูชันสองมิติ และการเรียนรู้แบบกึ่งกำกับเทียบกับแบบจำลองพื้นฐานอื่น ประกอบด้วย แบบจำลองคอนโวลูชันสามมิติ (C3D) และแบบจำลองคอนโวลูชันสองมิติร่วมกับแบบจำลองหน่วยความจำระยะสั้นแบบยาว (CNN+LSTM) โดยกำหนดทรัพยาการเพื่อใช้สำหรับการเรียนรู้แบบจำลองต่าง ๆ ด้วยคอมพิวเตอร์ 16vCPU, 104GB of Memory and 2 NVIDIA Tesla T4 for GPUs บนระบบปฏิบัติการ Ubuntu 16.04

3.5.1 ความถูกต้องแม่นยำในการเรียนรู้

งานวิจัยชิ้นนี้วัดผลความถูกต้องแม่นยำเฉลี่ยจากการเรียนรู้แบบ โดยแสดงในสูตรที่ 3.4

$$Accuracy = \frac{\sum_{i=1}^{K} Accuracy_k}{K}$$

โดย K คือ จำนวนชุดข้อมูล

3.5.2 ขนาดของแบบจำลองและระยะเวลาในการเรียนรู้

การพิจารณาประสิทธิภาพการเรียนรู้ของแบบจำลอง ที่ถูกใช้ในงานวิจัยชิ้นนี้จะพิจารณาที่ขนาดของแบบจำลอง และระยะเวลาในการเรียนรู้บนทรัพยากรที่จำกัด โดยเทียบกับแบบจำลองล้ำสมัย ที่ออกแบบเพื่อจำแนกวิดีโอโดยทดสอบความแม่นยำกับ การเรียนรู้ด้วยแบบจำลองคอนโวลูชันสามมิติ และการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติร่วมกับหน่วยความจำระยะสั้นแบบยาว

3.5.3 ความถูกต้องแม่นยำในการเรียนรู้ร่วมกับชุดข้อมูล Hollywood2

การวัดผลความถูกต้องแม่นยำของงานวิจัยนี้ได้ จะเลือกแบบจำลองที่มีประสิทธิภาพในการเรียนรู้ที่สุดกับชุดข้อมูลวิดีโอที่งานวิจัยนี้จัดทำ มาใช้เพื่อประเมินผลความถูกต้องแม่นยำ กับชุดข้อมูลที่ใช้วัดประสิทธิภาพ (Benchmark dataset) ซึ่งเลือกใช้ชุดข้อมูล Hollywood2 ทั้งสองประเภท คือ ชุดการรู้จำฉาก (Scene Recognition) และชุดการรู้จำการเคลื่อนไหว (Action Recognition) โดยเทียบผลความแม่นยำกับแบบจำลองอื่น ประกอบด้วย SalCLSTM และ ACLNet

บทที่ 4
การทดลองและผลการทดลง

4.1 ระบบที่ใช้ในการทดลอง

4.1.1 คอมพิวเตอร์ที่ใช้ในการทดลอง

การประมวลผลงานวิจัยชิ้นนี้ ถูกแบ่งเป็น 2 ระยะ คือ ระยะพัฒนา และระยะทดสอบ เนื่องจากงานวิจัยชิ้นนี้ ดำเนินการประเมินผลการทดลองโดยเปรียบเทียบจากระยะเวลาในจำแนกวิดีโอ ซึ่งทั้ง 2 ระยะจะดำเนินการบนระบบกูเกิลคลาวด์ (Google Cloud Platform) ในระยะพัฒนาจะพัฒนาของงานวิจัยชิ้นนี้กำหนดคอมพิวเตอร์ที่ใช้สำหรับพัฒนาโดยมีหน่วยประมวลผลกลางจำลอง (vCPU) 32 คอร์ หน่วยความจำ 104 GB หน่วยประมวลผลกราฟฟิก Nvidia Tesla T4 หน่วยความจำ 14 GB จำนวน 2 หน่วย ระบบปฏิบัติการ Ubuntu 16.04

เนื่องจากงานวิจัยนี้ มีการประเมินผลระยะเวลาในการประเมินผล และข้อจำกัดของแบบจำลองที่นำมาประเมินผลร่วม อาทิ C3D และ CNN+LSTM จึงดำเนินการปรับคอมพิวเตอร์ที่ใช้ในระยะทดสอบของงานวิจัยชิ้นนี้ โดยคุณสมบัติของระบบซึ่งใช้หน่วยประมวลผลกลางจำลอง (vCPU) จำนวน 16 คอร์ หน่วยความจำ 64 GB หน่วยประมวลผลกราฟฟิก Nvidia Tesla T4 หน่วยความจำ 14 GB จำนวนหนึ่งหน่วย ระบบปฏิบัติการ Ubuntu 16.04

การเรียนรู้ของแบบจำลองทั้งหมดในงานวิจัยชิ้นนี้จะถูกเรียนรู้ด้วยหน่วยประมวลผลกราฟฟิก ซึ่งชุดข้อมูลสำหรับการเรียนรู้จะถูกประมวลผลและเก็บไว้ในหน่วยความจำของเครื่องทั้งหมด และการออกแบบอัลกอลิทึมของงานวิจัยชิ้นนี้จะมุ่งเน้นที่การประมวลผลข้อมูลวิดีโอ โดยอาศัยการประมวลผลแบบคู่ขนาน

4.1.2 การเขียนโปรแกรมและเฟรมเวิร์คที่ใช้

การเขียนโปรแกรมของงานวิจัยชิ้นนี้จะถูกแบ่งออกเป็น 4 ส่วน โดยส่วนที่หนึ่งคือ การเตรียมชุดข้อมูลวิดีโอ การเตรียมชุดข้อมูลรูปภาพ และการสร้างแบบจำลอง และการประมวลผลลัพธ์จากแบบจำลอง การประมวลผลชุดข้อมูลวิดีโอในงานวิจัยนี้ เป็นการเตรียมข้อมูลสำหรับการทดลองและประมวลผล โดยสกัดข้อมูลภาพฉาก และการไหลของแสง ซึ่งการสกัดรูปภาพจากวิดีโอถูกเขียนด้วยชุดคำสั่งแบช (Bash scripts) โดยใช้ไลบราลี FFMpeg และการสกัดการไหลของแสงถูกเขียนด้วยภาษาไพทอน (Python Language) โดยใช้ไลบราลี OpenCV การประมวลผลรูปภาพ การสร้าง

แบบจำลอง และการประมวลผลผลลัพธ์จากแบบจำลอง ใช้ภาษาไพทอน (Python Language) ซึ่งการประมวลผลรูปาภพประกอบด้วยการปรับขนาดรูปภาพ และปรับรูปร่างของข้อมูล โดยใช้ไลบราลี Pillow, Numpy การสร้างแบบจำลอง ประกอบด้วยการสร้างแบบจำลองคอนโวลูชัน แผนที่จัดการตนเอง และหน่วยความจำระยะสั้นแบบยาว ถูกเขียนด้วยภาษาไพทอน โดยใช้ไลบราลี TensorFlow 1.15

4.2 ชุดข้อมูลที่ใช้ในการทดลอง

4.2.1 ชุดข้อมูล

ชุดข้อมูลดิบของงานวิจัยนี้ได้รวบรวมข้อมูลวิดีโอจากช่องต่าง ๆ เพื่อใช้เป็นชุดข้อมูลในการทดสอบ โดยประกอบไปด้วย 18 ชุดข้อมูล จำนวน 912 ชุดข้อมูล ซึ่งชุดข้อมูลทั้งหมดถูกบันทึกไว้ในรูปแบบ MP4 (H.264) ที่รายละเอียดความกว้างไม่น้อยกว่า 480p ซึ่งกระบวนการเตรียมชุดข้อมูลของงานวิจัยนี้ จะดำเนินการสกัดข้อมูลรูปภาพ และข้อมูลการไหลของแสง เพื่อใช้สำหรับการทดสอบแบบจำลองต่างๆ ซึ่งกระบวนการสกัดรูปภาพจากชุดข้อมูลใช้ชุดคำสั่งแบช ที่เรียกใช้งานไลบราลี ffmpeg ถูกแสดงในภาพที่ 12 และกรบวนการสกัดการไหลของแสงจะใช้ชุดคำสั่งไพทอน ที่เรียกใช้งานไลบราลี OpenCV ถูกแสดงในภาพที่ 13

```
for f in videos/$program/*.mp4; do
    mkdir frames_224/${program}/${i} -p
    ffmpeg -i "${f}" -s 224x224 -vf fps=1 frames_224/$program/$i/%d.jpg
    i=$((i + 1))
done
```

ภาพ 12 ตัวอย่างชุดคำมั่งแบชสำหรับสกัดรูปภาพจากวิดีโอ

```
for file in videos:
    video = cv2.VideoCapture("/videos" + file)
    print(video.isOpened())
    framerate = video.get(5)
    os.makedirs("/Users/.../" + "video_" + str(int(count)))
    while (video.isOpened()):
        frameId = video.get(1)
```

```
        success,image = video.read()
        if( image != None ):
            image=cv2.resize(image,(224,224), interpolation = cv2.INTER_AREA)
        if (success != True):
            break
        if (frameId % math.floor(framerate) == 0):
            filename = "/video_" + str(int(count)) + "/image_" + str(int(frameId / math.floor(framerate))+1) + ".jpg"
            print(filename)
            cv2.imwrite(filename,image)
    video.release()
    print('done')
    count+=1
```

ภาพ 13 ภาพตัวอย่างชุดคำสั่งไพทอนสำหรับสกัดรูปภาพและการไกลของแสงจากวิดีโอ

4.2.2 การแบ่งชุดข้อมูล

ชุดข้อมูลในงานวิจัยนี้ประกอบไปด้วย 912 วิดีโอ ทั้งสิ้น 18 ประเภทรายการโดยการรวบรวมชุดข้อมูล จะพิจารณาจากประเภทของเนื้อหาในวิดีโอ เนื้อหาของชุดข้อมูลวิดีโอที่ถูกใช้ในงานวิจัยนี้จะประกอบไปด้วยวิดีโอที่มีลักษณะฉากการถ่ายทำ และลักษณะเนื้อหาที่แตกต่างกัน อาทิ ฉากการถ่ายทำในสตูดิโอ (Studio Production) ซึ่งมีทั้งลักษณะของรายการโทรทัศน์ และรายการข่าว, ฉากการถ่ายทำภายนอกสถานที่ (Outdoor Production) ฉากการถ่ายทำแบบผสมระหว่างสตูดิโอ และภายนอกสถานที่ (Mixed Production) การ Vlog และฉากภายในเกมส์ (Games Casting) โดยชุดข้อมูลนี้จะถูกแบ่งเพื่อใช้เพื่อใช้ในการวัดผลแบบจำลองต่างๆ ที่ถูกนำเสนอ และเปรียบเทียบ การแบ่งชุดข้อมูลจะทำเพื่อทดสอบหรือทำแบบไขว้จำนวน 5 โฟลด์ โดยการแบ่งชุดข้อมูลถูกแสดงในตารางที่ 3 ซึ่งเป็นการแบ่งตามจำนวนโฟลด์ และตารางที่ 4 แสดงรายละเอียดประเภทของรายการตามลักษณะของเนื้อหาในวิดีโอของเนื้อหารายการ

ตาราง 3 การแบ่งชุดข้อมูลสำหรับการทดสอบ

Programs/Fold	1	2	3	4	5
Program 1	9	9	10	10	10
Program 2	12	12	12	11	11
Program 3	10	10	10	11	11
Program 4	13	12	12	12	12
Program 5	11	11	11	11	11
Program 6	4	4	3	3	3
Program 7	9	9	8	8	8
Program 8	7	7	7	8	8
Program 9	14	14	14	14	14
Program 10	13	12	12	12	12
Program 11	10	10	11	11	11
Program 12	15	14	14	14	14
Program 13	8	8	9	9	9
Program 14	11	11	11	11	12
Program 15	12	12	12	12	13
Program 16	11	11	12	12	12
Program 17	2	2	2	2	2
Program 18	10	11	11	11	11
รวม	**181**	**179**	**181**	**182**	**184**

ตาราง 4 การแบ่งชุดข้อมูลสำหรับการทดสอบ

Programs	รายละเอียด
Program 1	Studio-based Production
Program 2	Studio-based Production
Program 3	Studio-based Production
Program 4	Studio-based Production
Program 5	Studio-based Production

Programs	รายละเอียด
Program 6	Studio-based Production
Program 7	Studio-based Production
Program 8	Outdoor Production
Program 9	Outdoor Production
Program 10	Mixed Production
Program 11	Mixed Production
Program 12	Mixed Production
Program 13	Mixed Production
Program 14	Mixed Production
Program 15	Game casting
Program 16	Vlog
Program 17	News
Program 18	News

4.2.3 การเพิ่มจำนวนชุดข้อมูล

การเพิ่มจำนวนชุดข้อมูลสำหรับการเรียนรู้ในงานวิจัยนี้ ถูกนำไปใช้เพื่อเพิ่มประสิทธิภาพการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติ เนื่องจากข้อจำกัดทางจำนวนชุดข้อมูลของรายการโทรทัศน์ อย่างไรก็ตามการจำแนกวิดีโอเป็นไปได้ยากที่จะใช้ภาพฉากทั้งหมดในการจำแนกวิดีโอ งานวิจัยนี้จึงมีแนวคิดการเพิ่มจำนวนข้อมูล โดยการสร้างชุดข้อมูลเพิ่มเติมจากภาพฉากที่ไม่ได้ถูกเลือกมาใช้เป็นชุดข้อมูล ซึ่งแนวคิดในการเพิ่มชุดข้อมูลของงานวิจัยนี้ คือการเลือกภาพฉากให้ครอบคลุมวิดีโอทั้งหมด ซึ่งจะข้ามบางฉากโดยเลือกภาพฉากทั้งสิ้น 500 ภาพฉากในแต่ละวิดีโอ ดังนั้นเพื่อให้ครอบคลุมเนื้อหาทั้งหมด การสร้างชุดข้อมูลจะต้องข้ามบางฉากสำคัญไป งานวิจัยนี้จึงเสนอการเพิ่มชุดข้อมูลด้วยการสร้างชุดข้อมูลจากจำนวนความยาวของวิดีโอโดยใช้จำนวนภาพฉากที่ต้องการหารด้วยความยาวทั้งหมดของวิดีโอ แสดงอัลกอริทึมที่ใช้ในการเพิ่มชุดข้อมูลในภาพที่ 14

Declare an array of results as an empty array Gather a total of frames number	
Step 1:	Extract image frames from videos
Step 2:	Set the number of needed frames to 500 frames
Step 3:	Calculate a number skipping from dividing a total of frames number by several needed frames.
Step 4:	Calculate several datasets from dividing the number of total frames by a skipping number.
Step 5	Set several query dataset as zero.
Step 6:	Declare features an array of features as an empty array, and an index of frames as zero.
Step 7	Calculate an index of frames from adding a multiplying of adding an index of frames to several query dataset by a skipping number
Step 8	Get a frame from a video at an index of frames and append it to an array of features
Step 9	Add an index of frames by one and do the step 7 if an index of frames is less than several total frames and a skipping number.
Step 10	Normalize a minimum value and a maximum value of feature array from 0 and 255 to 0 and 1.
Step 11	Stack a feature array in a sequence of depth-wise.
Step 12	Append a feature array to a dataset array.
Step 13	Add a query number by one and do step 6 if a query number is less than several datasets.

ภาพ 14 อัลกอริทึม ในการเพิ่มชุดข้อมูล

4.2.4 การเตรียมชุดข้อมูล

การเตรียมชุดข้อมูลในงานวิจัยนี้เป็นกระบวนการปรับขนาดของภาพฉาก และลักษณะของฟีเจอร์ให้เหมาะกับแบบจำลองต่าง ๆ ซึ่งจะประกอบด้วย 2 วิธี วิธีแรก คือ การปรับขนาดมิติของภาพฉากสำหรับใช้ในการเรียนรู้ของแบบจำลองคอนโวลูชันสองมิติ โดยปรับขนาดของรูปภาพโดยไม่คำนึงถึงอัตราส่วน ให้ภาพฉากมีขนาด 224 x 224 พิกเซล แสดงในรูปภาพที่ 15 และการปรับขนาดมิติของภาพฉากสำหรับใช้ในการเรียนรู้ของแผนที่จัดการตนเอง แสดงในรูปภาพที่ 16

```
img.thumbnail(size, Image.ANTIALIAS)
```

ภาพ 15 ตัวอย่างชุดคำสั่งเพื่อปรับขนาดรูปภาพสำหรับแบบจำลองคอนโวลูชัน

```
np.divide(load_image_resize(d[0], (size,size)).reshape(-1,size*size*3), 255.)
```

ภาพ 16 ตัวอย่างชุดคำสั่งเพื่อคำสั่งสำหรับปรับรูปภาพของแผนที่จัดการตนเอง

4.3 การดำเนินการทดลอง

4.3.1 การจัดการกลุ่มรูปภาพด้วยฟิลเตอร์ค่ากลาง

ฟิลเตอร์ค่ากลางถูกนำมาใช้เพื่อรวมกลุ่มของรูปภาพหลายๆ ภาพให้เป็นภาพเดียวกัน โดยการใช้ค่ากลางของภาพแต่ละช่อง จำนวน 3 ช่อง โดยใช้ไลบลาลี numpy แสดงในภาพที่ 17 ซึ่งในการทดลองหาจำนวนกลุ่มของรูปภาพที่เหมาะสมในงานวิจัยนี้ได้มีการทดลองค่าระหว่าง 10 ภาพฉาก และ 20 ภาพฉาก ซึ่งตัวอย่างของการรวมภาพฉากแสดงใน ภาพที่ 18

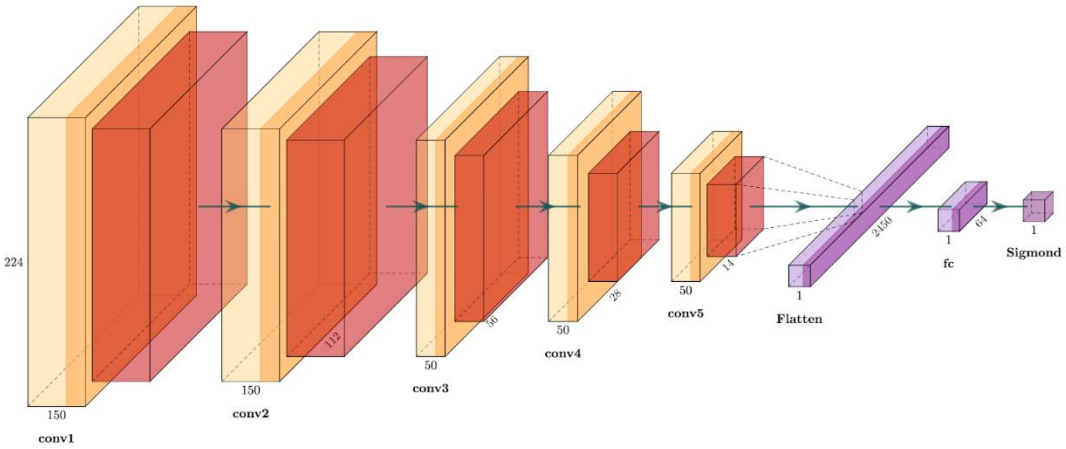
```
np.dstack(images)
```

ภาพ 17 ตัวอย่างชุดคำสั่งเพื่อสำหรับการซ้อนทับของรูปภาพ.

ภาพ 18 ตัวอย่างภาพที่เกิดจากการรวมกันด้วยตัวกรองมัธยฐาน

4.3.2 การสร้างแบบจำลองคอนโวลูชันสองมิติ (C2D)

การสร้างแบบจำลองคอนโวลูชันสองมิติในงานวิจัยนี้ สร้างแบบจำลองคอนโวลูชันสองมิติโดยการปรับปรุงจากแบบจำลองวีจีจี 16 (ภาพที่ 19) และเพิ่มชั้นค่ากลางเพื่อลดขนาดชุดข้อมูลก่อนที่จะดำเนินการเรียนรู้ โดยการทดสอบเพื่อหาค่าพารามิเตอร์ที่เหมาะสม ในขั้นตอนการสร้างแบบจำลองงานวิจัยนี้ดำเนินการทดสอบกับชุดข้อมูลทั้งหมด ในทุกประเภทของวิดีโอ



ภาพ 19 แสดงแบบจำลองคอนโวลูชันสองมิติที่ถูกสร้างเพื่อใช้สำหรับทดลอง

กระบวนการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติในการทดลอง กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เลือกใช้อัลกอริทึมการเพิ่มประสิทธิภาพ ด้วยอัลกอริทึมอดัม และดำเนินการเรียนรูป จำนวน 20 ยุค โดยกำหนดชุดการเรียนรู้ ที่ขนาด 100 ข้อมูล ให้มีการสลับชุดข้อมูลอยู่เสมอ และเลือกพิจารณาจากรอบที่มีผลค่าความผิดพลาดน้อยที่สุด

4.3.3 การสร้างแผนที่จัดระเบียบด้วยตนเอง

แผนที่จัดระเบียบด้วยตนเองที่ถูกใช้ในงานวิจัยนี้ ถูกสร้างด้วยวิธีการทดสอบสร้างแผนที่ด้วยจำนวนโหนด ความกว้างและความสูง ระหว่าง 5, 10, 20, 40, 60, 80 และ 100 โหนด ตามลำดับ ในการทดสอบการสร้างแผนที่จัดระเบียบด้วยตนเอง ในการทดสอบการสร้างแผนที่จัดระเบียบด้วยตนเองในงานวิจัยนี้ ได้ดำเนินการทดสอบกับชุดข้อมูลทั้งหมด จึงได้เลือกใช้จำนวนโหนดทั้งสิ้น 40 โหนด ทั้งความกว้างและความสูง

กระบวนการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติในการทดลอง กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เลือกใช้อัลกอริทึมการเพิ่มประสิทธิภาพ ด้วยอัลกอริทึมอดัม และดำเนินการเรียนรูป จำนวน 20 ยุค โดยกำหนดชุดกาเรียนรู้ ที่ขนาด 100 ข้อมูล ให้มีการสลับ (Shuffle) ชุดข้อมูลอยู่เสมอ และเลือกพิจารณาจากรอบที่มีผลค่าความผิดพลาดน้อยที่สุด

4.3.4 การสร้างแบบจำลองเพื่อทำนายผลลัพธ์จากแผนที่จัดการตนเอง

การสร้างแบบจำลองเพื่อทำนายประเภทพวิดีโอจากผลลัพธ์ของแผนที่จัดการตนเองในงานวิจัยนี้ จะสร้างแบบจำลองเพื่อเปรียบเทียบกันทั้งหมด 4 แบบ ซึ่งประกอบด้วย แบบจำลองโดยการโหวต แบบจำลองโดยการคำนวณค่าเอ็นโทรพี แบบจำลองโครงข่ายประสาทเทียม และแบบจำลองหน่วยความจำระยะสั้นแบบยาว

แบบจำลองการโหวตในงานวิจัยนี้ ถูกใช้เพื่อประเมินจำนวนภาพฉากที่จะใช้ในการทดสอบ โดยจำแนกกลุ่มข้อมูลทดสอบด้วยแผนที่จัดระเบียบตัวเอง เพื่อหาจัดกลุ่มและหาตำแหน่งของกลุ่มรูปภาพนั้น หลังจากนั้นทำการคำนวณหาค่าคลาสของวิดีโอในกลุ่มข้อมูลนั้น ๆ โดยระบุตามคลาสของวิดีโอตามจำนวนภาพฉากที่ปรากฏในกลุ่มนั้น ๆ มากที่สุด ซึ่งอัลกอริทึมที่ใช้ในการระบุคลาสของกลุ่มข้อมูลถูกแสดงในภาพที่ 20

Start	Collect all weights of the SOM map and all map locations in the same index.
Step I	Select input vector
Step II	Find an index of minimum value from calculating Euclidian norms by subtraction of an input vector and all weights
Step III	Get a map location from an index in step II

ภาพ 20 อัลกอริทึมที่ใช้สำหรับการหากลุ่มข้อมูลของภาพฉาก

การคำนวณความวุ่นวาย ในงานวิจัยนี้ ใช้แสดงความวุ่นวายของประเภทของข้อมูล โดยการคำนวณหาค่าเอ็นโทรพีของคลาสข้อมูล ที่เกิดขึ้นในแต่ละกลุ่มของข้อมูลภาพ หลังจากกระบวนการระบุกลุ่มข้อมูลภาพ จากเรียนรู้ด้วยแผนที่จัดระเบียบด้วยตนเอง ซึ่งอัลกอริทึมที่ใช้ในการคำนวณหาค่าเอ็นโทรพีของกลุ่มข้อมูลถูกแสดงในอัลกอริทึมที่ 21

Start	Collect all clusters
Step I	Select a cluster
Step II	Count all appeared classes in the cluster
Step III	Calculate entropy of each class and define not appear class to zero.
Step IV	Multiple each entropy values with the ratio of counting class in this cluster to a total of train sets.

ภาพ 21 อัลกอริทึมที่ใช้สำหรับคำนวณค่าเอ็นโทรพีของกลุ่มข้อมูล

การสร้างแบบจำลองโครงข่ายประสาทเทียมใน จะสร้างแบบจำลองโดยกำหนดจำนวนโหนดชั้นนำเข้า เท่ากับจำนวนโหนดทั้งหมดของแผนที่จัดระเบียบตนเอง ซึ่งมีทั้งสิ้น 1,600 โหนด และมีชั้นซ่อนจำนวน 3 ชั้นซ่อน ประกอบด้วย 1,600 โหนด, 800 โหนด และ 128 โหนด ตามลำดับ และมี 5 โหนด ของชั้นปรับเทียบหรือชั้นเอาต์พุต เพื่อทำนายคลาสของวิดีโอ ซึ่งรายละเอียดของแบบจำลองโครงข่ายประสาทเทียม ถูกแสดงในภาพที่ 22

Input layer	1600 nodes
Activation	ReLu
Hidden layer 1	1600 nodes
Activation	ReLu
Hidden layer 2	800 nodes
Activation	ReLu
Hidden layer 3	128 nodes
Activation	ReLu
Output layer	Five nodes
Activation	Softmax

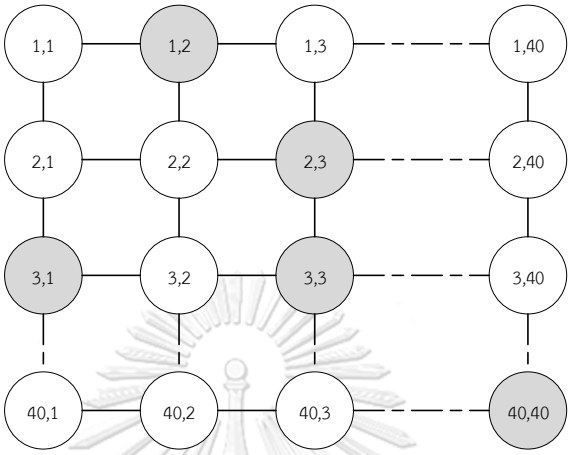
ภาพ 22 แบบจำลองโครงข่ายประสาทเทียมสำหรับจำแนกคลาสของวิดีโอ

เพื่อให้ผลลัพธ์จากแผนที่จัดระเบียบตนเอง สามารถนำมาจำแนกด้วยโครงข่ายประสาทเทียมได้ ดังนั้นหลังจากระบุภาพฉากไปยังกลุ่มต่าง ๆ จะนำผลลัพธ์ที่ได้มาเปลี่ยนรูปเป็น เวกเตอร์ขนาด 1,600 x 1 เพื่อให้อยู่ในรูปแบบที่เหมาะสมกับการนำไปใช้ โดยอัลกอริทึมแสดงในภาพที่ 23 และเข้ารหัสคลาสของวิดีโอเพื่อให้ใช้สำหรับการเรียนรู้ของแบบจำลองโครงข่ายประสาทเทียม

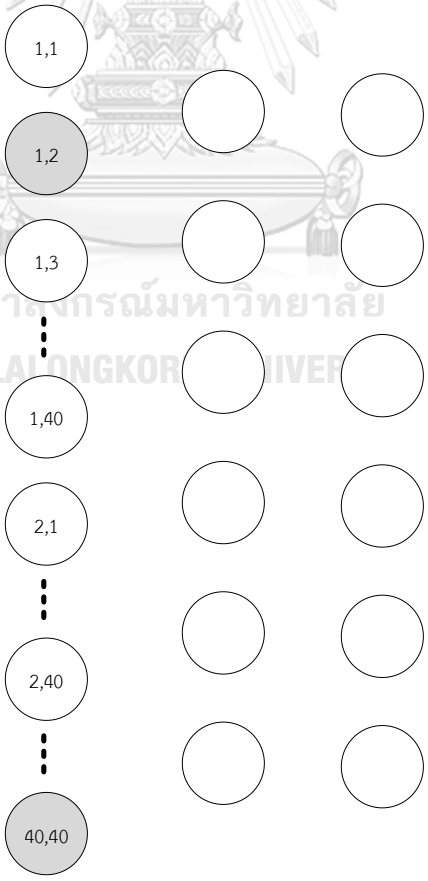
Start	Define empty 2d array and set m, n to zero
Step I	Set 2d array
Step II	Count all appeared classes in the cluster
Step III	Calculate entropy of each class and define not appear class to zero.
Step IV	Multiple each entropy values with the ratio of counting class in this cluster to a total of train sets.
Step V	Do *Step I* for all input features

ภาพ 23 อัลกอริทึมสำหรับปรับปรุงข้อมูลนำเข้าผลลัพธ์จากการเรียนรู้ด้วยแผนที่จัดการตนเอง

การแปลงผลลัพธ์ตามอัลกอริทึมที่ 23 ด้วยผลลัพธ์จากแผนที่จัดระเบียบด้วยตนเองของวิดีโอ ซึ่งโหนดที่มีสีเทาแสดงถึงโหนดของกลุ่มข้อมูลที่ปรากฏในวิดีโอนั้น ๆ (ภาพที่ 24) เป็นเวกเตอร์ขนาด 1,600 x 1 ด้วยการปรับรูปภาพ แสดงในภาพที่ 25



ภาพ 24 ตัวอย่างผลลัพธ์ของวิดีโอจากการจัดกลุ่มด้วยแผนที่จัดระเบียบด้วยตนเอง



ภาพ 25 ตัวอย่างของเวกเตอร์สำหรับการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม

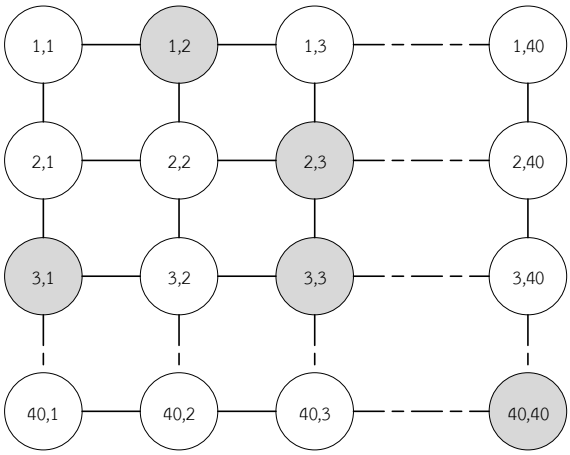
โดยขั้นตอนในการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล โดยทำการเรียนรู้ 20 ยุค ด้วยอัลกอริทึมอดัม ในการปรับปรุงค่าน้ำหนัก และเลือกยุคที่มีค่าความผิดพลาดน้อยที่สุด

นอกจากนี้งานวิจายนี้ยังนำเสนอวิธีการใช้หน่วยความจำระยะสั้นแบบยาว เพื่อพิจารณารูปแบบของกลุ่มข้อมูลอย่างเป็นลำดับ โดยจากผลลัพธ์ของแผนที่จัดระเบียบด้วยตนเองจำนวน 1,600 โหนด ในรูปแบบเวกเตอร์ขนาด 1,600 x 1 จะถูกเข้ารหัสจำนวน 16 โหนด เพื่อใช้เป็นชุดข้อมูลนำเข้า และเข้ารหัสข้อมูลคลาสของวิดีโอจำนวนเพื่อใช้เป็นข้อมูลในชั้นปรับเทียบหรือชั้นเอาต์พุตสำหรับการเรียนรู้ เช่นเดียวกับการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม ซึ่งอัลกอลิทึมของการปรับปรุงข้อมูลนำเข้าถูกแสดงในภาพที่ 26

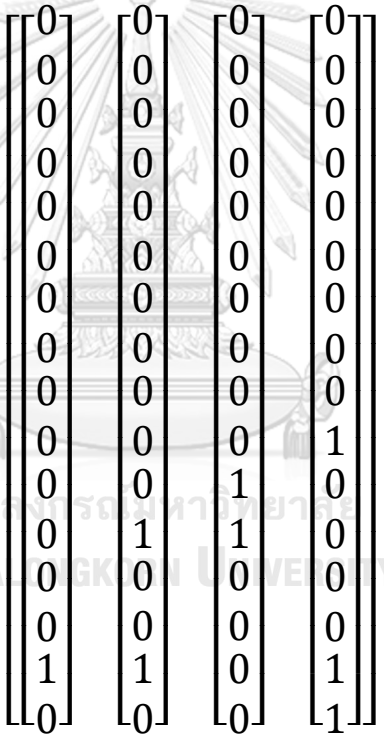
Start	Encode 1,600 neuron nodes to 16 bits
Step I	Select video from all input features
Step II	Collect all video clusters with frames sorted.
Step III	Map all sorted clusters to encoded nodes
Step IV	Repeat *Step I* for all videos

ภาพ 26 อัลกอริทึมสำหรับปรับปรุงชุดข้อมูลนำเข้าจากผลลัพธ์ของแผนที่จัดระเบียบตนเอง

การแปลงผลลัพธ์ตามอัลกอริทึมที่ 23 ด้วยผลลัพธ์จากแผนที่จัดระเบียบด้วยตนเองของวิดีโอ ซึ่งโหนดที่มีสีเทาแสดงถึงโหนดที่ปรากฏในวิดีโอนั้นๆ (ภาพที่ 27) เป็นเวกเตอร์ขนาด 16 x 1 ด้วยการเข้ารหัส และนำมาประกอบเป็นเมตริกซ์ของชุดข้อมูลสำหรับหน่วยความจำระยะสั้นแบบยาว แสดงในภาพที่ 28

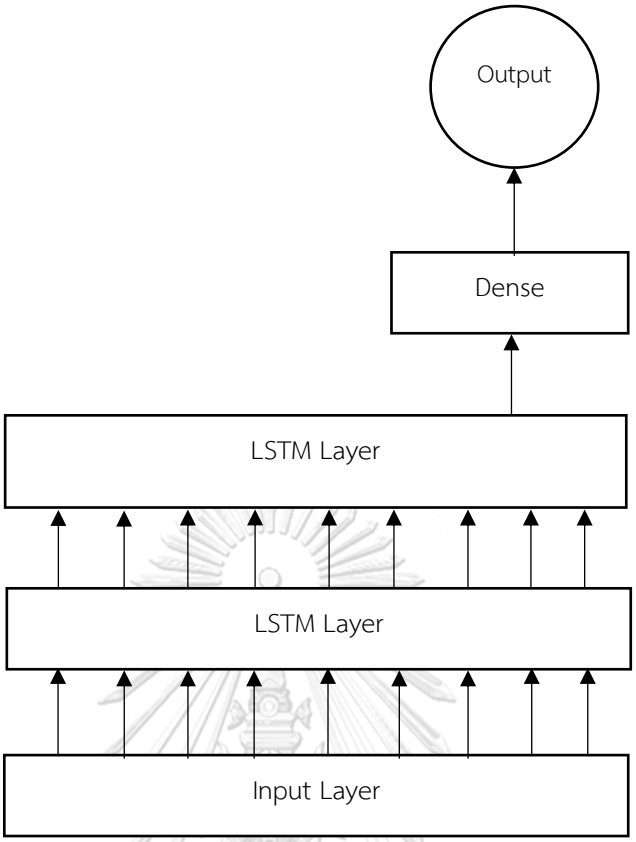


ภาพ 27 ตัวอย่างผลลัพธ์ของวิดีโอจากการจัดกลุ่มด้วยแผนที่จัดระเบียบด้วยตนเอง



ภาพ 28 ตัวอย่างของเวกเตอร์สำหรับการเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว

การสร้างแบบจำลองหน่วยความจำระยะสั้นแบบยาว ในงานวิจัยนี้จะประกอบด้วยชั้นข้อมูลนำเข้า, 2 ชั้นข้อมูลหน่วยความจำระยะสั้นแบบยาว (LSTM Layers) ขนาด 1,000, 500 โหนด ตามลำดับ และ 5 โหนดของชั้นปรับเทียบหรือชั้นเอาต์พุต แสดงในภาพที่ 29



ภาพ 29 หน่วยความจำระยะสั้นแบบยาวสำหรับจำแนกวิดีโอ

กระบวนการเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เลือกใช้อัลกอริทึมการเพิ่มประสิทธิภาพ ด้วยอัลกอริทึมอดัม และดำเนินการเรียนรูปจำนวน 20 ยุค และเลือกพิจารณาจากยุคที่มีผลค่าความผิดพลาดน้อยที่สุด.

4.4 ผลการทดลอง

4.4.1 ผลการทดลองของแบบจำลอง C2D

ผลการทดลองของแบบจำลองคอนโวลูชันสองมิติ กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เพื่อจำแนกรายการโทรทัศน์ จะประเมินจากการเรียนรู้แบบไขว้จำนวน 5 โฟลด์ ซึ่งแต่ละรอบการเรียนรู้จะบันทึกแบบจำลองที่มีค่าสูญเสียการตรวจสอบความถูกต้องน้อยที่สุดเสมอโดยผลการทดลองแสดงในตารางที่ 5

ตาราง 5 ตารางแสดงผลการทดลองของแบบจำลองคอนโวลูกชันสองมิติ

K	C2D
1	84.38%
2	76.56%
3	87.50%
4	70.31%
5	78.13%
	79.38%

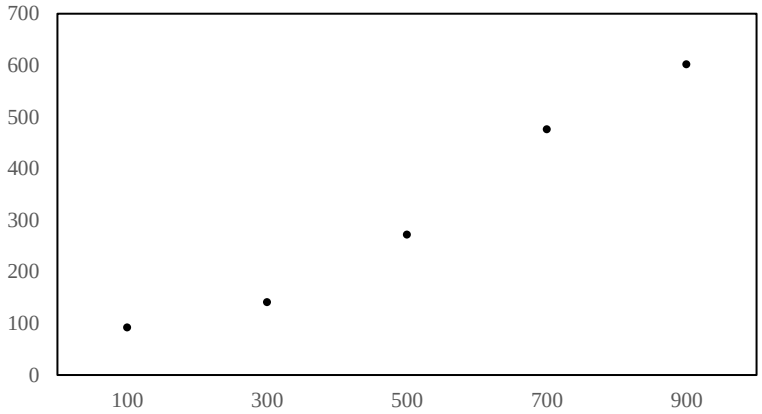
4.4.2 ผลการทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับการลงคะแนน

ผลการทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับการลงคะแนนเสียงตามจำนวนมาก (SOM + Voting) กับชุดข้อมูลทั้งหมด 912 ชุดข้อมูล ถูกนำมาใช้เพื่อเลือกจำนวนภาพฉากที่จะใช้ในวิธีอื่น ๆ ดังนั้นในการทดลองจะทดสอบด้วยจำนวนภาพฉาก 100, 300, 500, 700 และ 900 ตามลำดับ โดยผลลัพธ์แสดงในตารางที่ 6

ตาราง 6 ตารางแสดงผลการทดลองของ SOM + Voting

Number of frames	Accuracy (%)
100	59.47
300	60.00
500	65.79
700	58.33
900	57.29

นอกจากนี้ ในการทดลองด้วย SOM + Voting ยังคงวัดผลในแง่ของระยะเวลาในการเรียนรู้เทียบกับความแม่นยำในการเรียนรู้ ซึ่งถูกแสดงผลเปรียบเทียบด้วยกราฟในภาพที่ 30 โดยแกน X แสดงข้อมูลจำนวนภาพฉาก และแกน Y แสดงระยะเวลาในการเรียนรู้



ภาพ 30 กราฟเปรียบเทียบระยะเวลาในการเรียนรู้ (Y) และจำนวนภาพฉาก (X)

4.4.3 ผลการทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับค่าเอ็นโทรพี

ผลการทดสอบการจำแนกวิดีโอด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับค่าเอ็นโทรพี (SOM + Entropy) กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล โดยพิจารณาจากค่าเอ็นโทรพีของข้อมูลผลลัพธ์จากแผนที่จัดระเบียบตนเองด้วยวิธีการวัดผลแบบไขว้ จำนวน 5 โฟลด์ แสดงในตารางที่ 7 ซึ่งมีค่าเฉลี่ยของความแม่นยำร้อยละ 59.18

ตาราง 7 ตารางแสดงผลการทดสอบการจำแนกวิดีโอด้วยวิธี SOM + Voting

K	SOM + Entropy
1	60.8
2	54.48
3	55.37
4	66.48
5	58.79
	59.18

4.4.4 ผลการทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม

ผลการทดสอบการจำแนกวิดีโอด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม (SOM + ANN) กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เพื่อพิจารณาผลลัพธ์จากแผนที่จัดระเบียบตนเอง ด้วยวิธีวัดผลแบบไขว้ จำนวน 5 โฟลด์ แสดงผลในตารางที่ 8 ซึ่งมีค่าเฉลี่ยร้อยละ 71.978

ตาราง 8 ตารางแสดงผลการทดสอบการจำแนกวิดีโอด้วยวิธี SOM + ANN

K	SOM + ANN
1	72.52
2	71.8
3	71.64
4	72.39
5	71.54
	71.978

4.4.5 ผลการทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับหน่วยความจำระยะสั้นแบบยาว

ผลการทดสอบการจำแนกวิดีโอด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับหน่วยความจำระยะสั้นแบบยาว (SOM + LSTM) กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล โดยใช้หน่วยความจำระยะสั้นแบบยาวเพื่อพิจารณาผลลัพธ์จากแผนที่จัดระเบียบตนเอง ด้วยวิธีวัดผลแบบไขว้ จำนวน 5 โฟลด์ แสดงผลในตารางที่ 9 ซึ่งมีค่าเฉลี่ยร้อยละ 70.09

ตาราง 9 ตารางแสดงผลการทดสอบการจำแนกวิดีโอด้วยวิธี SOM + ANN

K	SOM + LSTM
1	73.17
2	71.46
3	68.14
4	71.08
5	66.58
	70.09

4.4.6 ผลการทดลองการเปรียบเทียบการทำนาย (Confusion Matrix)

จากการทดลองด้วยแผนที่จัดระเบียบด้วยตนเองในบทที่ 4.4.3 ถึง บทที่ 4.4.4 พบว่าการทำนายด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม มีประสิทธิภาพในการจำแนกวิดีโอได้ดีที่สุด ในบทนี้จึงแสดงรายละเอียดการทำนายโดยเปรียบเทียบรายคลาส ของแผนที่จัดระเบียบด้วยตนเองร่วมกับโครงข่ายประสาทเทียม โดยจะแสดงจำนวนคลาสที่เป็นจริง เทียบกับคลาสที่สามารถทำนายได้ พบว่าคลาสที่ 0 ถึง 5, คลาสที่ 8 ถึง คลาสที่ 12, และคลาสที่ 16 ถึง 17 สามารถจำแนกได้ถูกต้องทั้งหมด แต่ไม่สามารถจำแนกรายการโทรทัศน์ในคลาสที่ 15 ได้ ซึ่งผลถูกแสดงในตารางที่ 10

ตาราง 10 ตารางแสดงผลการเปรียบเทียบการทำนาย

CLASS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	2	0	0	7	0	0	0	0	0	0	0	0	0	0	0

CLASS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7	0	0	0	1	4	0	0	3	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
13	0	4	2	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
14	0	3	8	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
15	0	0	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11

4.4.7 ผลการทดลองของแบบจำลองคอนโวลูชันสามมิติ

การทดลองด้วยแบบจำลองคอนโวลูชันสามมิติ กับชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล เป็นการทดสอบการเรียนรู้ด้วยวิธีวัดผลแบบไขว้ จำนวน 5 โฟลด์ โดยไม่มีการถ่ายทอดการเรียนรู้ (Transfer Learning) แสดงผลในตารางที่ 11 ซึ่งมีค่าเฉลี่ยร้อยละ 60.072

ตาราง 11 ตารางผลลัพธ์การจำแนกวิดีโอด้วยแบบจำลองคอนโวลูชันสามมิติ

K	C3D
1	60
2	60.62
3	59.96
4	60.11
5	59.67
	60.072

4.4.8 ผลการทดลองของแบบจำลองคอนโวลูชันร่วมกับหน่วยความจำระยะสั้นแบบยาว

การทดลองด้วยแบบจำลองคอนโวลูชันร่วมกับหน่วยความจำระยะสั้นแบบยาว (CNN + LSTM) กับชุดข้อมูลนำเข้าจำนวน 912 ชุดข้อมูล เป็นการทดสอบการเรียนรู้ด้วยวิธีวัดผลแบบไขว้ จำนวน 5 โฟลด์ แสดงผลในตารางที่ 12 ซึ่งมีค่าเฉลี่ยร้อยละ 60.64

ตาราง 12 ตารางผลลัพธ์การจำแนกวิดีโอด้วยแบบจำลองคอนโวลูชัน ร่วมกับหน่วยความจำระยะสั้นแบบยาว

K	C2D
1	61.71
2	62.24
3	60.13
4	59.07
5	60.07
	60.644

4.4.9 ผลการทดลองเปรียบเทียบระยะเวลาในการเรียนรู้ตามประเภทแบบจำลอง

การทดลองในงานวิจัยนี้ยังได้แสดงการเปรียบเทียบระยะเวลาในการเรียนรู้ของแบบจำลองต่าง ๆ แผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม ที่ถูกเสนอในงานวิจัยนี้เป็นแบบจำลองที่มีประสิทธิภาพมากที่สุด โดยแสดงการเปรียบเทียบระยะเวลาการประมวลผลเทียบกับแบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันร่วมกับหน่วยความจำระยะสั้นแบบยาว ซึ่งในกระบวนการเรียนรู้ของแต่ละแบบจำลองไม่มีการถ่ายโอนการเรียนรู้ การเปรียบเทียบจะเปรียบเทียบจากการเรียนรู้ด้วยชุดข้อมูลวิดีโอจำนวน 912 ชุดข้อมูล ด้วยวิธีการวัดผลแบบไขว้ จำนวน 20 ยุค โดยค่าเฉลี่ยการเรียนรู้ถูกแสดงในตารางที่ 13

ตาราง 13 การเปรียบเทียบระยะเวลาในการเรียนรู้ของแบบจำลองแต่ละประเภท

Models	Training time (minutes)
C2D	~44 minutes
SOM + ANN	~320 minutes
C3D	~480 minutes
LSTM	~720 minutes

4.4.10 ผลการทดลองแบบจำลองต่าง ๆ ด้วยชุดข้อมูล Hollywood2

กระบวนการทดลองในงานวิจัยนี้ ได้ทดสอบกระบวนการเทียบกับชุดข้อมูลพื้นฐานอื่น ๆ เพื่อตรวจสอบประสิทธิภาพของวิธีการที่นำเสนอ โดยงานวิจัยนี้เลือกใช้ชุดข้อมูล Hollywood2 ซึ่งประกอบด้วยชุดข้อมูลรู้จำภาพฉาก และชุดข้อมูลรู้จำการเคลื่อนไหว โดยได้เปรียบเทียบความแม่นยำในการเรียนรู้ด้วยวิธีแผนที่จัดระเบียบด้วยตนเองร่วม กับแบบจำลองโครงข่ายประสาทเทียม เทียบกับแบบจำลองอื่น ๆ ที่ถูกนำเสนอในปัจจุบัน อาทิ SelCLSTM และ ACLNet ซึ่งงานวิจัยนี้สามารถจำแนกได้ด้วยค่าเฉลี่ยความแม่นยำร้อยละ 93.72 โดยรายละเอียดการเปรียบเทียบ แสดงตารางที่ 14

ตาราง 14 ตารางแสดงการเปรียบเทียบผลลัพธ์ความแม่นยำในการรู้จำชุดข้อมูล Hollywood2

Models	Accuracy (%)
SOM + ANN	**93.72**
SalCLSTM	93.33
ACLNet	91.13

4.5 วิเคราะห์ผลการทดลอง

จากชุดข้อมูลที่ถูกใช้ทดสอบ การสร้างชุดข้อมูลด้วยวิธีการข้ามภาพฉากเพื่อให้ชุดข้อมูลมีความครอบคลุมเนื้อหาของวิดีโอนั้นสามารถใช้เพื่อเป็นกระบวนการหนึ่งในการเรียนรู้ได้ ทั้งยังสามารถช่วยเพิ่มจำนวนชุดข้อมูลจากการข้ามภาพฉาก ด้วยวิธีการใช้ความยาวของวิดีโอคำนวณกับจำนวนภาพฉากที่ต้องการที่แสดงในบทที่ 4.2.3

จากผลการทดลองในบทที่ 4.4.1 ด้วยวิธีการเรียนรู้แบบจำลองคอนโวลูชันสองมิติ โดยใช้ตัวกรองมัธยฐาน พบว่าสามารถเพิ่มประสิทธิภาพในรู้จำและลดระยะเวลาในการเรียนรู้ชุดข้อมูลเพิ่มขึ้น ดังแสดงในบทที่ 4.4.1 (ความแม่นยำ) และบทที่ 4.4.8 (ระยะเวลาในการเรียนรู้) เมื่อเทียบกับแบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันร่วมกับหน่วยความจำระยะสั้นแบบยาว ซึ่งเทคนิคตัวกรองมัธยฐานสามารถช่วยลดจำนวนภาพฉากที่ใช้ในใช้ในการเรียนรู้ได้ โดยเฉพาะกับข้อมูลรายการโทรทัศน์ และแบบจำลองคอนโวลูชันสองมิติสามารถใช้เพื่อเรียนรู้และจำแนกวิดีโอได้เช่นกัน อีกทั้งการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสองมิติ ยังใช้ระยะเวลาในการเรียนรู้น้อยกว่าแบบจำลองอื่น ๆ

จากผลการทดลองในบทที่ 4.4.2 ถึง บทที่ 4.4.5 ด้วยวิธีการเรียนรู้ด้วยแผนที่จัดระเบียบตนเองร่วมกับการลงคะแนน การใช้ค่าเอ็นโทรพี แบบจำลองโครงข่ายประสาทเทียม และหน่วยความจำระยะสั้นแบบยาว ซึ่งงานวิจัยนี้ใช้การเรียนรู้ด้วยแผนที่จัดระเบียบตนเองร่วมกับการลงคะแนนเป็นตัวชี้วัดจำนวนภาพฉากที่เหมาะสมในการจำแนกวิดีโอ จากผลการทดลองในบทที่ 4.4.2 พบว่า การใช้จำนวนภาพฉากจำนวน 500 ภาพฉากในการเรียนรู้มีประสิทธิภาพสูงสุด เมื่อเปรียบเทียบระหว่างความแม่นยำ และระยะเวลาในการเรียนรู้ ซึ่งการวิเคราะห์ผลการทดลองจำนวนภาพฉากที่ได้จากการทดลองที่บทที่ 4.4.2 ถูกนำมาใช้เป็นจำนวนภาพฉากที่ใช้กับการทดลองในบทที่ 4.4.3 ถึง บทที่ 4.4.5 ซึ่งจากผลการทดลองใน 3 วิธี ประกอบด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับการใช้ค่าเอ็นโทรพี แผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม และแผนที่จัดระเบียบด้วยตนเองร่วมกับหน่วยความจำระยะสั้นแบบยาว พบว่าการใช้แผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียมมีประสิทธิภาพดีที่สุด โดยเมื่อพิจารณาจากผลการทดลองของแบบจำลอง SOM + ANN จึงได้แสดงผลการเปรียบเทียบการทำนายของวิธีการนี้ โดยแสดงการเปรียบเทียบการทำนาย ในบทที่ 4.4.6 พบว่าแบบจำลองไม่สามารถจำแนกวิดีโอประเภท vlog และมีประสิทธิภาพต่ำในการจำแนกคลาสที่มีลักษณะการถ่ายทำภายนอกสถานที่เป็นจำนวนมาก

ผลการทดลองบทที่ 4.4.7 และบทที่ 4.4.8 เป็นการใช้แบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันสองมิติร่วมกับหน่วยความจำระยะสั้นแบบยาว เพื่อทดสอบประสิทธิภาพในการเรียนรู้กับชุดข้อมูลวิดีโอในงานวิจัยนี้ โดยไม่มีการถ่ายโอนการเรียนรู้ ด้วยวิธีการวัดผลแบบไขว้ นอกจากนี้ผลการทดลองบทที่ 4.4.10 ยังแสดงถึงการทดสอบประสิทธิภาพของการจำแนกวิดีโอด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม โดยวัดผลเทียบกับแบบจำลองอื่น ๆ ประกอบด้วย SalCLSTM และ ACLNet โดยทดสอบด้วยชุดข้อมูล Hollywood2 ซึ่งพบว่าการเรียนรู้ด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับแบบจำลองโครงข่ายประสาทเทียม สามารถจำแนกวิดีโอด้วยประสิทธิภาพที่ใกล้เคียงกับแบบจำลองพื้นฐานอื่น ๆ

บทที่ 5
สรุปผลงานวิจัยและข้อเสนอแนะ

5.1 สรุปผลงานวิจัย

งานวิจัยนี้นำเสนอกระบวนการจำแนกวิดีโอ และชุดข้อมูลวิดีโอจากชุดข้อมูลรายการต่าง ๆ ในปัจจุบัน จำนวน 18 ประเภทรายการ มีวิดีโอทั้งสิ้น 912 วิดีโอ ซึ่งเสนอวิธีการจำแนกโดยตระหนักถึงการลดขนาดของแบบจำลองที่ใช้ในการเรียนรู้ โดยเสนอการเรียนรู้เพื่อจำแนกวิดีโอในสองวิธีหลัก ประกอบด้วย แบบจำลองคอนโวลูชันสองมิติ และการเรียนรู้แบบกึ่งกำกับ (Semi-Supervised Learning) แบบจำลองแผนที่จัดระเบียบด้วยตนเอง

แบบจำลองคอนโวลูชันสองมิติเพื่อจำแนกวิดีโอในงานวิจัยนี้ นำเสนอการซ้อนรูปภาพในแนวลึกเพื่อลดขนาดของแบบจำลอง นอกจากนี้ยังนำเสนอตัวกรองมัธยฐานเพื่อลดจำนวนภาพฉาก โดยการรวมภาพฉากหลังผ่านตัวกรองมัธยฐาน อย่างไรก็ตามเนื่องจากข้อจำกัดของจำนวนวิดีโอรายการต่าง ๆ เรานำเสนอกระบวนการเพิ่มจำนวนข้อมูลโดยการพิจารณาใช้ภาพฉากทั้งหมดจากวิดีโอในขั้นตอนการสร้างชุดข้อมูลก่อนการเรียนรู้ จากผลการทดสอบพบว่าแบบจำลองคอนโวลูชันสองมิติสามารถเรียนรู้และจำแนกวิดีโอที่มีความแม่นยำร้อยละ 78.38 รวมถึงได้เปรียบเทียบประสิทธิภาพการจำแนกวิดีโอของแบบจำลองคอนโวลูชันสองมิติ กับแบบจำลองอื่น ๆ ประกอบด้วย แบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันสองมิติร่วมกับหน่วยความจำระยะสั้นแบบยาว โดยวิธีการเรียนรู้แบบไม่มีการถ่ายโอนการเรียนรู้ อีกทั้งได้ประเมินระยะเวลาในการเรียนรู้ ซึ่งพบว่าแบบจำลองคอนโวลูชันสองมิติสามารถใช้เวลาในการเรียนรู้ได้เร็วกว่าวิธีอื่น ๆ ที่นำมาทดสอบ

การจำแนกวิดีโอด้วยการเรียนรู้แบบกึ่งกำกับในงานวิจัยนี้ นำเสนอการจำแนกวิดีโอด้วยแบบจำลองแผนที่จัดระเบียบด้วยตนเอง ร่วมกับวิธีการจำแนกอื่นเพื่อจำแนกวิดีโอในขั้นตอนสุดท้าย ประกอบด้วยการลงคะแนน คำนวณค่าเอ็นโทรพี จำแนกด้วยการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม และการเรียนรู้ด้วยหน่วยความจำระยะสั้นแบบยาว อย่างไรก็ตามการเรียนรู้ด้วยแผนที่จัดระเบียบตัวเอง จำเป็นต้องระบุจำนวนภาพฉากที่จะเป็นตัวแทนของภาพฉากในแต่ละกลุ่มวิดีโอ งานวิจัยนี้จึงนำเสนอการวัดผลการเลือกจำนวนภาพฉากที่เหมาะสมด้วยการใช้แผนที่จัดระเบียบด้วยตนเองร่วมกับการลงคะแนน โดยพบว่าการใช้จำนวนภาพฉากจำนวน 500 ภาพฉากสำหรับจำแนกวิดีโอเหมาะสมที่สุดในงานวิจัยนี้ และได้ทดลองการจำแนกรายการโทรทัศน์ด้วยจำนวน 500 ภาพฉากกับวิธีจำแนกวิดีโอในขั้นตอนสุดท้ายอื่น ประกอบด้วย คำนวณค่าเอ็นโทรพี กาเรียนรู้ด้วย

แบบจำลองโครงข่ายประสาทเทียม และหน่วยความจำระยะสั้นแบบยาว พบว่าการจำแนกด้วยการเรียนรู้ของแผนที่จัดระเบียบด้วยตนเองร่วมกับการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียมในการจำแนกวิดีโอในขั้นตอนสุดท้าย มีประสิทธิภาพสูงสุดซึ่งสามารถจำแนกได้ที่ความถูกต้องร้อยละ 70.09 นอกจากนี้งานวิจัยนี้ยังคงเปรียบเทียบผลการทดลองกับการเรียนรู้ด้วยแบบจำลองคอนโวลูชันสามมิติ และแบบจำลองคอนโวลูชันสองมิติร่วมกับหน่วยความจำระยะสั้นแบบยาว โดยไม่มีการถ่ายโอนการเรียนรู้ และวิธีการเรียนรู้ด้วยแผนที่จัดระเบียบด้วยตนเองร่วมกับการเรียนรู้ของแบบจำลองโครงข่ายประสาทเทียม ใช้ระยะเวลาการเรียนรู้น้อยที่สุด เมื่อเทียบกับเทคนิคอื่น

อย่างไรก็ตามงานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพการเรียนรู้กับชุดข้อมูลพื้นฐานอื่น คือ Hollywood2 ซึ่งเป็นชุดข้อมูลการรู้จำภาพฉาก และการรู้จำการเคลื่อนไหว พบว่าแบบจำลองแผนที่จัดระเบียบด้วยตนเองร่วมกับการเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม สามารถจำแนกได้ด้วยความแม่นยำร้อยละ 93.7 ซึ่งมีประสิทธิภาพใกล้เคียงกับแบบจำลองพื้นฐานอื่น อาทิ SalCLSTM และ ACLNet

5.2 ข้อเสนอแนะ

จากการทดลองของแบบจำลองคอนโวลูชันสองมิติ พบว่ากระบวนการเลือกภาพฉากก่อนการลดจำนวนภาพฉากด้วยตัวกรองมัธยฐานไม่มีการพิจารณาความสัมพันธ์ของภาพฉาก จึงทำให้มีภาพฉากที่ไม่เกี่ยวข้อง ส่งผลให้ภาพที่เกิดจากตัวกรองมัธยฐานมีความผิดเพี้ยนในบางภาพ

การทดลองของแผนที่จัดระเบียบด้วยตนเองร่วมกับการจำแนกในขั้นสุดท้ายด้วยวิธีอื่น พบว่าการใช้ข้อมูลของกลุ่มข้อมูลไม่มีการพิจารณาคุณลักษณะที่แท้จริง หรือตัวแทนคุณลักษณะของกลุ่มข้อมูลนั้น ๆ จากภาพฉากที่ถูกจำแนกการเพิ่มแบบจำลองเพื่อดึงคุณลักษณะของกลุ่มข้อมูลจากภาพฉาก จะสามารถช่วยเพิ่มประสิทธิภาพในการเรียนรู้ได้ อีกทั้งยังคาดว่าการเพิ่มประสิทธิภาพในการเรียนรู้กับชุดข้อมูลขนาดใหญ่

บรรณานุกรม

1. Zha, S., et al., *Exploiting image-trained CNN architectures for unconstrained video classification.* arXiv preprint arXiv:1503.04144, 2015.
2. Tran, D., et al. *Learning spatiotemporal features with 3d convolutional networks*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
3. Wu, Z., et al., *Modeling Spatial-Temporal Clues in a Hybrid Deep Learning Framework for Video Classification*, in *Proceedings of the 23rd ACM international conference on Multimedia - MM '15*. 2015. p. 461-470.
4. Kohonen, T., *The self-organizing map.* Proceedings of the IEEE, 1990. **78**(9): p. 1464-1480.
5. Deng, J., et al., *ImageNet: A large-scale hierarchical image database*, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009. p. 248-255.
6. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
7. Canziani, A., A. Paszke, and E. Culurciello, *An analysis of deep neural network models for practical applications.* arXiv preprint arXiv:1605.07678, 2016.
8. Paszke, A., et al., *Enet: A deep neural network architecture for real-time semantic segmentation.* arXiv preprint arXiv:1606.02147, 2016.
9. Karpathy, A., et al. *Large-scale video classification with convolutional neural networks*. in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.
10. Sun, J., J. Wang, and T.C. Yeh, *Video understanding: from video classification to captioning*. 2019, Stanford University. http://cs231n. stanford. edu/reports/2017/pdfs/709. pdf~....
11. Poms, A., et al., *Scanner: Efficient video analysis at scale.* ACM Transactions on Graphics (TOG), 2018. **37**(4): p. 138.
12. Dhanachandra, N., K. Manglem, and Y.J. Chanu, *Image segmentation using K-means clustering algorithm and subtractive clustering algorithm.* Procedia

Computer Science, 2015. **54**: p. 764-771.

13. Zhu, W., J. Lu, and J. Zhou, *Nonlinear subspace clustering for image clustering.* Pattern Recognition Letters, 2018. **107**: p. 131-136.
14. Mo, H., et al., *Color segmentation of multi-colored fabrics using self-organizing-map based clustering algorithm.* Textile Research Journal, 2017. **87**(3): p. 369-380.
15. Soomro, K., A.R. Zamir, and M. Shah, *A dataset of 101 human action classes from videos in the wild.* Center for Research in Computer Vision, 2012.
16. Abu-El-Haija, S., et al., *Youtube-8m: A large-scale video classification benchmark.* arXiv preprint arXiv:1609.08675, 2016.

ประวัติผู้เขียน

ชื่อ-สกุล	Itthisak Phueaksri
วัน เดือน ปี เกิด	15 December 1989
สถานที่เกิด	Nonthaburi
วุฒิการศึกษา	Department of Computer Engineering, Chulalongkorn University
ที่อยู่ปัจจุบัน	135/276 (16) Soi 57 Yak 6 Reawadee Road Tambon Tarad Kwan Amphoe Muang Nonthaburi