

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนใบหน้าด้วยการ
ทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก

Training Deep Neural Networks for Facial Landmarks Detection by Predicting
Principal Component Analysis's Coefficients

รัตยากร ไทยพันธ์ ชวัลรัตน์ ศรีนวลปาน วลัยภรณ์ ศรีเกลี้ยง จิรวัดน์ นาคสุวรรณ อวยพร ชูแก้ว
มัลลิกา รัตพันธ์ และ กฤตภาส สงศรีอินทร์*

Rattayagon Thaiphon, Chawanrat Srinounpan, Walaiporn Sornkliang, Jeerawat
Naksuwan, Auyporn Chukeaw, Munlika Rattaphun, & Kritaphat Songsri-in*

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

Computer Science Program, Faculty of Science and Technology,

Nakhon Si Thammarat Rajabhat University

Submitted 5/4/2022 ; Revised 17/5/2022 ; Accepted 26/5/2022

บทคัดย่อ

การระบุตำแหน่งจุดสังเกตบนใบหน้าเป็นขั้นตอนก่อนการประมวลผลที่สำคัญสำหรับการวิเคราะห์ภาพใบหน้าเพราะสามารถลดความหลากหลายและซับซ้อนของรูปภาพได้ งานวิจัยฉบับนี้จึงมีวัตถุประสงค์ในการปรับปรุงประสิทธิภาพการเรียนรู้โครงข่ายประสาทเทียมเชิงลึกสำหรับการตรวจจับจุดสังเกตบนใบหน้าในรูปภาพด้วยการการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักแทนที่การทำนายจุดสังเกตบนใบหน้าโดยตรง ซึ่งผลการทำนายค่าสัมประสิทธิ์นั้นสามารถแปลงกลับให้เป็นตำแหน่งจุดสังเกตได้จากแม่แบบของการวิเคราะห์องค์ประกอบหลัก จากผลการทดลองบนฐานข้อมูลมาตรฐาน 300W สรุปได้ว่าการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักนั้นมีประสิทธิภาพมากกว่าการทำนายตำแหน่งจุดสังเกตบนใบหน้าโดยตรง โดยขนาดที่เหมาะสมของค่าสัมประสิทธิ์อยู่ที่ 15 ซึ่งมีอัตราค่าเฉลี่ยความคลาดเคลื่อนน้อยที่สุดอยู่ที่ 5.723 เมื่อเปรียบเทียบกับอัตราค่าเฉลี่ยความคลาดเคลื่อนที่ 6.542 จากการทำนายจุดสังเกตบนใบหน้าโดยตรง

คำสำคัญ: การเรียนรู้เชิงลึก โครงข่ายประสาทเทียม การระบุตำแหน่งจุดสังเกตบนใบหน้า การวิเคราะห์องค์ประกอบหลัก

*ผู้ประสานงานหลัก (Corresponding Author)

E-mail: kritaphat_son@nstru.ac.th

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนใบหน้า
ด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก

Abstract

Facial landmarks detection is an important pre-processing step for facial analysis because it can reduce the diversity and complexity of images. The objective of this research aimed to improve the efficiency of training deep neural networks for facial landmarks detection by predicting Principal Component Analysis's coefficients instead of predicting facial landmarks directly. The predicted coefficients can then be converted back to facial landmarks using the Principal Component Analysis's templates. The experiments on the benchmark 300W dataset indicated that predicting coefficients of the Principal Component Analysis was more effective than predicting facial landmarks directly. We found that the optimal size of the coefficients was 15 and the minimum normalized mean error was 5.723 compared to 6.542 for direct facial landmarks prediction.

Keywords: deep learning, artificial neural networks, facial landmark detection, principal component analysis

บทนำ

การวิเคราะห์ใบหน้าในรูปภาพและวิดีโอเป็นขั้นตอนที่สำคัญทางคอมพิวเตอร์วิทัศน์ซึ่งมีการนำไปใช้ประโยชน์ในเชิงพาณิชย์อย่างหลากหลาย เช่น การติดตามใบหน้า การระบุตัวตน การจำแนกอารมณ์ และการตกแต่งรูปภาพใบหน้า ด้วยประโยชน์ของการวิเคราะห์ใบหน้าเหล่านี้จึงทำให้มีนักวิจัยจำนวนมากได้ศึกษาและพัฒนาแบบจำลองเพื่อสกัดข้อมูลที่สำคัญเหล่านี้จากสื่อดิจิทัล เช่น รูปภาพหรือวิดีโอโดยอัตโนมัติ [1-5] การวิเคราะห์ใบหน้าเหล่านี้มักมีการใช้การตรวจจับตำแหน่งใบหน้าและจุดสังเกตเป็นขั้นตอนก่อนการประมวลผลเพื่อลดความหลากหลายและซับซ้อนของรูปภาพ เนื่องจากการหมุน การปรับตำแหน่ง และการปรับขนาด ซึ่งสามารถปรับปรุงประสิทธิภาพของการวิเคราะห์รูปภาพในลำดับต่อ ๆ ไปได้ [6-9] แบบจำลองที่เคยเป็นที่นิยมในอดีตสำหรับการระบุตำแหน่งจุดสังเกตบนใบหน้า เช่น strongly supervised deformable part-based models (DPMs) [1], active appearance models (AAMs) [10] และ constrained local models (CLMs) [11] ใช้การวิเคราะห์ทางสถิติเพื่อศึกษาความสัมพันธ์ระหว่างส่วนต่าง ๆ ของจุดสังเกต และจำลองลักษณะของจุดสังเกตด้วยการแปลงคุณลักษณะที่ไม่แปรผันตามมาตราส่วน (scale-invariant feature transform: SIFT) [12] หรือ ฮิสโตแกรมของทิศทางค่าเกรเดียนท์ (histogram of oriented gradients: HOG) [13] ซึ่งสามารถสกัดลักษณะของรูปภาพได้ดีกว่าการจับความเข้มของสีโดยตรง แต่เนื่องจากข้อมูลรูปภาพและตำแหน่งของจุดสังเกตที่ถูกรวบรวมมีจำนวนมากขึ้น วิธีการเรียนรู้เชิงลึก เช่น โครงข่ายประสาทเทียมคอนโวลูชัน จึงเริ่มเป็นที่นิยมเพราะสามารถใช้ประโยชน์จากจำนวนข้อมูลที่เพิ่มขึ้นจนทำให้มีความแม่นยำในการทำนายผลที่ดีกว่า [6,7] อย่างไรก็ตามผลการทำนายของโครงข่ายประสาทเทียมเหล่านี้ยังมีความคลาดเคลื่อนของการทำนายอย่างมีนัยสำคัญ ณ ตำแหน่งของใบหน้าที่ถูกบดบัง หรือ มีการแสดงออกของสีหน้าที่แปลกไปจากปกติ เช่น อารมณ์โกรธ หรือ ตกใจ ดังนั้นงานวิจัยฉบับนี้จึงนำเสนอการปรับปรุงประสิทธิภาพของโครงข่ายประสาทเทียมด้วยการผสมผสานการเรียนรู้เชิงลึกเข้ากับการวิเคราะห์ทางสถิติของโครงข่ายใบหน้าด้วยการวิเคราะห์องค์ประกอบหลัก [14] เพื่อให้การระบุตำแหน่งจุดสังเกตของโครงข่ายประสาทเทียมมีความเสถียรและมีความแม่นยำมากขึ้น

วัตถุประสงค์

เพื่อปรับปรุงประสิทธิภาพสำหรับปัญหาการระบุจุดสังเกตบนใบหน้าในรูปภาพด้วยการทำนายค่าสัมประสิทธิ์จากการวิเคราะห์องค์ประกอบหลักในการเรียนรู้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

วิธีดำเนินการวิจัย

ปัญหาระบุตำแหน่งจุดสังเกตบนใบหน้า

ปัญหาระบุตำแหน่งจุดสังเกตบนใบหน้าเป็นขั้นตอนหลังจากการระบุตำแหน่งของใบหน้าในแต่ละรูปภาพ โดยที่ภาพใบหน้าแต่ละภาพจะถูกระบุตำแหน่งของจุดสังเกตต่าง ๆ ซึ่งแต่ละฐานข้อมูลอาจมีการใช้จำนวนของจุดสังเกตที่สนใจไม่เท่ากัน เช่น ภาพที่ 1 จำนวนจุดสังเกตของใบหน้าแต่ละใบหน้าจากฐานข้อมูล 300W [15,16] มีจุดสังเกตทั้งสิ้น 68 จุด ซึ่งแต่ละจุดบ่งบอกตำแหน่งของส่วนต่าง ๆ ที่สำคัญบนใบหน้า เช่น คิ้ว ตา จมูก และ ปาก เป็นต้น



ภาพที่ 1 การระบุตำแหน่งจุดสังเกตบนใบหน้า ซึ่งระบุจุดสังเกตทั้งหมด 68 จุดจากฐานข้อมูล 300W [15,16]

เนื่องจากขนาดใบหน้าแต่ละในภาพอาจจะมีขนาดไม่เท่ากัน ภาพใบหน้าในฐานข้อมูลสำหรับปัญหาการระบุตำแหน่งจุดสังเกตบนใบหน้าจึงมักถูกปรับขนาดให้อยู่ในระดับเดียวกันเพื่อให้ข้อมูลนำเข้าสำหรับแบบจำลองมีขนาดเท่ากัน ซึ่งมีความสะดวกต่อการคำนวณของแบบจำลอง

ฐานข้อมูล 300W

งานวิจัยนี้ทดสอบการระบุตำแหน่งจุดสังเกตบนใบหน้าบนชุดข้อมูลมาตรฐาน 300W [15,16] ซึ่งฐานข้อมูลนี้กำหนดจุดสังเกตบนใบหน้าทั้งหมด 68 จุด ชุดข้อมูลสำหรับการเรียนรู้มีภาพใบหน้าจำนวน 3,148 ภาพ และชุดทดสอบสำหรับการทดสอบมีภาพใบหน้าจำนวน 600 ภาพ โดยภาพใบหน้าในฐานข้อมูลรวบรวมมาจากสถานการณ์จริง จึงทำให้ข้อมูลภาพใบหน้ามีความหลากหลายทั้งในแง่ของท่าทาง อารมณ์ แสง เงา และเชื้อชาติ ดังภาพที่ 1

โครงข่ายประสาทเทียมคอนโวลูชัน

งานวิจัยนี้ได้ใช้โครงข่ายประสาทเทียมคอนโวลูชันแบบง่ายจากแม่แบบของ Feng และคณะ [17] ดังแสดงในภาพที่ 2 ซึ่งมีการปรับขนาดของภาพใบหน้าให้มีขนาด $64 \times 64 \times 3$ พิกเซล ผลการทำนายของโครงข่ายมีขนาด 68×2 จุด ซึ่งบ่งบอกถึงตำแหน่งแนวนอนและแนวตั้งของจุดสังเกตบนใบหน้าทั้ง 68 จุด โครงข่ายประสาทเทียมมีจำนวนชั้นทั้งหมด 7 ชั้น โดยแบ่งเป็นแบบคอนโวลูชัน 5 ชั้น และแบบเชื่อมต่อทั้งหมด (fully connected) 2 ชั้น ลักษณะของตัวแปรเสริม ขนาดของข้อมูลนำเข้าและข้อมูลส่งออกในแต่ละชั้นแสดงดังตารางที่ 1 โดยชั้นที่มีลักษณะเป็นแบบคอนโวลูชันจะมีการลดขนาดของข้อมูลให้เล็กลงครั้งหนึ่งด้วยการใช้ชั้นลดขนาดด้วยค่าที่มากที่สุด (max pooling) นอกจากนี้ระหว่างการเปลี่ยนผ่านของชั้นคอนโวลูชันชั้นสุดท้ายและชั้นเชื่อมต่อทั้งหมด ขนาดของข้อมูลจะถูกปรับจาก 3 มิติ เป็น 1 มิติ ($2 \times 2 \times 512 = 1,024$)



ภาพที่ 2 ลักษณะของโครงข่ายประสาทเทียมคอนโวลูชันแบบง่ายจากแม่แบบของ Feng และคณะ [17]

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนใบหน้าด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก

ตารางที่ 1 ขนาดของตัวแปรเสริม ข้อมูลนำเข้า และข้อมูลส่งออกในแต่ละชั้น

ชั้น	ลักษณะของชั้น	ขนาดตัวแปรเสริม	ขนาดข้อมูลนำเข้า	ขนาดข้อมูลส่งออก
1	คอนโวลูชัน	$3 \times 3 \times 3 \times 32$	$64 \times 64 \times 3$	$32 \times 32 \times 32$
2	คอนโวลูชัน	$3 \times 3 \times 32 \times 64$	$32 \times 32 \times 32$	$32 \times 32 \times 32$
3	คอนโวลูชัน	$3 \times 3 \times 64 \times 128$	$32 \times 32 \times 32$	$32 \times 32 \times 32$
4	คอนโวลูชัน	$3 \times 3 \times 128 \times 256$	$32 \times 32 \times 32$	$4 \times 4 \times 256$
5	คอนโวลูชัน	$3 \times 3 \times 256 \times 512$	$4 \times 4 \times 256$	$2 \times 2 \times 512$
6	เชื่อมต่อทั้งหมด	2048×1024	2048	1024
7	เชื่อมต่อทั้งหมด	1024×136	1024	136

ฟังก์ชันสูญเสีย

โดยทั่วไปโครงข่ายประสาทเทียมมักมีการใช้ผลต่างกำลังสอง (mean square error: MSE) หรือผลต่างสัมบูรณ์ (mean absolute error: MAE) เป็นฟังก์ชันสูญเสียในการเรียนรู้แบบจำลอง ซึ่งมีรูปแบบดังสมการที่ 1 และ 2

$$\text{ผลต่างสัมบูรณ์จุดสังเกต } (y, \hat{y}) = |y - \hat{y}| \quad (1)$$

$$\text{ผลต่างกำลังสองจุดสังเกต } (y, \hat{y}) = (y - \hat{y})^2 \quad (2)$$

กำหนดให้ y คือ ค่าของป้ายกำกับของตำแหน่งจุดสังเกตบนไบนารี
 \hat{y} คือ ผลการทำนายของแบบจำลอง

การวิเคราะห์องค์ประกอบหลัก

การวิเคราะห์องค์ประกอบหลัก (principal component analysis: PCA) เป็นวิธีการทางสถิติที่ใช้วิเคราะห์ข้อมูลหลายตัวแปรเพื่อหาองค์ประกอบหลักซึ่งอธิบายการกระจายตัวของข้อมูลได้มากที่สุด การวิเคราะห์องค์ประกอบหลักมักถูกนำมาใช้ลดขนาดของข้อมูลหลายตัวแปรให้มีขนาดเล็กลงโดยยังคงสภาพความเป็นข้อมูลเดิม โดยข้อมูลเดิมสามารถเปลี่ยนให้มีขนาดลดลงได้ด้วยสมการที่ 3 [14]

$$p = U (y - \bar{y}) \quad (3)$$

กำหนดให้ p คือ ค่าสัมประสิทธิ์ของข้อมูลที่ถูกทำให้มีขนาดเล็กลง

y คือ ค่าของข้อมูลหลายตัวแปรตั้งต้น

\bar{y} คือ ค่าเฉลี่ยของข้อมูลหลายตัวแปร "y"

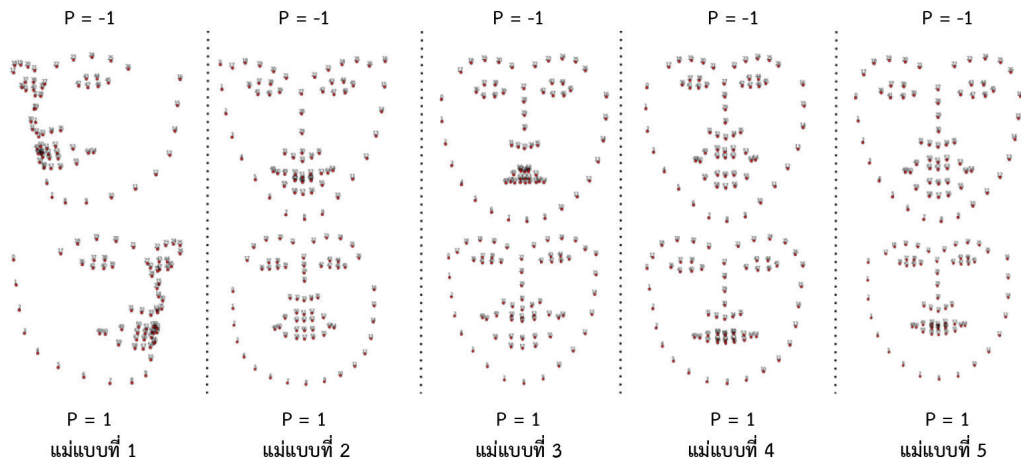
U คือ ค่าของแม่แบบซึ่งได้มาจากการแก้ปัญหาการวิเคราะห์องค์ประกอบหลัก

การแปลงค่าสัมประสิทธิ์ซึ่งเป็นข้อมูลขนาดเล็กให้มีขนาดเท่ากับข้อมูลเดิมสามารถแปลงได้โดยสมการที่ 4 [14]

$$y = \bar{y} + Up \quad (4)$$

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนไบนารีด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก

ภาพตัวอย่างของจุดสังเกตบนใบหน้าที่สร้างจากสมการที่ 4 ซึ่งเป็นผลบวกของจุดสังเกตเฉลี่ย \bar{y} กับผลคูณของค่าสัมประสิทธิ์ p (-1 และ 1) กับแม่แบบ U ที่ได้มาจากการแก้ปัญหาการวิเคราะห์องค์ประกอบหลักของจุดสังเกตบนใบหน้าจากฐานข้อมูล 300W แสดงดังภาพที่ 3 ซึ่งแสดงปัจจัยต่าง ๆ ที่บ่งบอกความเป็นจุดสังเกตตามลำดับความสำคัญ เช่น แม่แบบที่ 1 มีความสัมพันธ์กับการเอียงของใบหน้า แม่แบบที่ 2 แสดงมุมก้มมุมเงยของใบหน้า และแม่แบบที่ 3 แสดงการยิ้มของใบหน้า เป็นต้น



ภาพที่ 3 ตัวอย่างการแปลงค่าสัมประสิทธิ์ $p = -1$ (แถวที่ 1) และ $p = 1$ (แถวที่ 2) จากการวิเคราะห์องค์ประกอบหลักของแม่แบบ (U) 5 อันดับแรก

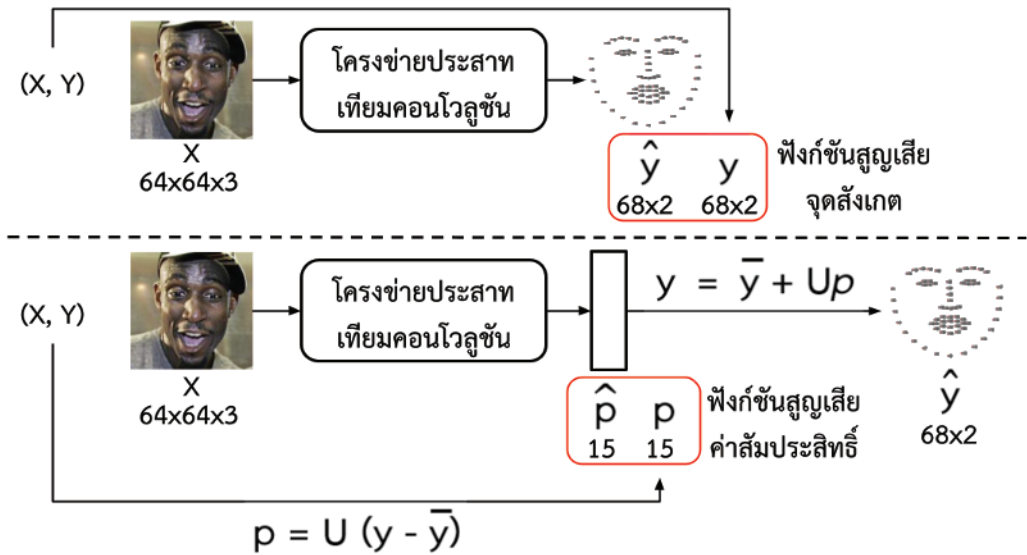
การลดขนาดของข้อมูลด้วยการวิเคราะห์องค์ประกอบหลักนอกจากจะทำให้ได้ข้อมูลที่มีขนาดเล็กลงแล้วยังช่วยลดสัญญาณรบกวนรวมถึงข้อมูลที่ไม่มีความสำคัญในการบ่งบอกความเป็นข้อมูลอีกด้วย ดังนั้นผู้วิจัยจึงประเมินว่าการนำค่าสัมประสิทธิ์ของการวิเคราะห์องค์ประกอบหลักมาใช้เรียนรู้โครงข่ายประสาทเทียมน่าจะช่วยให้แบบจำลองสามารถจัดการกับข้อมูลที่มีลักษณะของภาพหรือจุดสังเกตที่ไม่สมบูรณ์หรือผิดแปลกได้ดีขึ้น

การใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักเป็นฟังก์ชันสูญเสีย

ลักษณะทางกายภาพแบบสองมิติของจุดสังเกตบนใบหน้าทั้ง 68 จุดอธิบายได้ด้วยการวิเคราะห์องค์ประกอบหลัก โดยขั้นตอนการใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักเป็นฟังก์ชันสูญเสียประกอบด้วย

1. การปรับรูปแบบของข้อมูลจุดสังเกตบนใบหน้าให้เป็น 1 มิติ ($68 \times 2 = 136$)
2. การเรียนรู้แม่แบบการวิเคราะห์องค์ประกอบหลัก U จากจุดสังเกตบนใบหน้าในชุดข้อมูลเรียนรู้
3. การแปลงขนาดของข้อมูลจุดสังเกตบนใบหน้า (y) เป็นค่าสัมประสิทธิ์ (p) ด้วยสมการที่ 3

หากกำหนดให้ลักษณะของชุดข้อมูลเดิมเป็น (X, Y) โดยที่ X คือภาพใบหน้าทั้งหมดในฐานข้อมูล และ Y คือ ตำแหน่งของจุดสังเกตในแต่ละภาพ แบบจำลองโครงข่ายประสาทเทียมทั่วไปจะทำนายค่าของจุดสังเกต \hat{y} โดยตรง ดังภาพที่ 3 ในขณะที่การใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักเป็นฟังก์ชันสูญเสียจะทำนายค่าสัมประสิทธิ์ \hat{p} ก่อน ซึ่งตำแหน่งของจุดสังเกตจะถูกแปลงจากค่าสัมประสิทธิ์ \hat{p} เป็นค่าของจุดสังเกต \hat{y} โดยสมการที่ 4 ดังภาพที่ 4



ภาพที่ 4 แผนผังเปรียบเทียบการเรียนรู้โครงข่ายประสาทเทียมซึ่งทำนายจุดสังเกตโดยตรง (บน) และ ทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก (ล่าง)

การใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักเป็นฟังก์ชันสูญเสียในการเรียนรู้โครงข่ายประสาทเทียมทำได้สองวิธี โดยใช้สมการที่ 5 และ 6

$$\text{ผลต่างสัมบูรณ์ค่าสัมประสิทธิ์} (p, \hat{p}) = |p - \hat{p}| \quad (5)$$

$$\text{ผลต่างกำลังสองค่าสัมประสิทธิ์} (p, \hat{p}) = (p - \hat{p})^2 \quad (6)$$

กำหนดให้ p คือ ค่าสัมประสิทธิ์ซึ่งมีขนาดเล็กของจุดสังเกตบนใบหน้า

\hat{p} คือ ค่าทำนายค่าสัมประสิทธิ์จากโครงข่ายประสาทเทียม

การวัดประสิทธิภาพแบบจำลอง

เนื่องจากขนาดของใบหน้าในแต่ละรูปภาพมีความแตกต่างกันมาก การเปรียบเทียบประสิทธิภาพของการระบุตำแหน่งใบหน้าจึงใช้อัตราค่าเฉลี่ยความคลาดเคลื่อน (Normalized Mean Error, NME) ดังสมการที่ 7 [6]

$$\text{อัตราค่าเฉลี่ยความคลาดเคลื่อน} (y, \hat{y}) = \frac{|y - \hat{y}|}{E} \quad (7)$$

กำหนดให้ y คือ ค่าของป้ายกำกับของตำแหน่งจุดสังเกตบนใบหน้า

\hat{y} คือ ผลการทำนายของแบบจำลอง

E คือ ค่าระยะห่างระหว่างขอบตาด้านบน

ซึ่งแบบจำลองที่มีประสิทธิภาพในการระบุตำแหน่งจุดสังเกตบนใบหน้าที่ดีกว่าจะให้ค่าอัตราค่าเฉลี่ยความคลาดเคลื่อนที่ต่ำกว่า

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนใบหน้าด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลัก

ผลการวิจัย

การระบุขนาดของค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลัก

ขนาดของค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลักส่งผลต่อการเรียนรู้ของโครงข่ายประสาทเทียม ดังตารางที่ 2 พบว่าเมื่อขนาดของค่าสัมประสิทธิ์เพิ่มขึ้นอัตราค่าเฉลี่ยความคลาดเคลื่อนจะลดลงจนมีค่าน้อยที่สุด เมื่อขนาดของค่าสัมประสิทธิ์มีค่าเป็น 15 หลังจากนั้นค่าความคลาดเคลื่อนจะมีค่าเพิ่มขึ้นเรื่อย ๆ ตามขนาดของค่าสัมประสิทธิ์ที่เพิ่มขึ้น ซึ่งสามารถสรุปได้ว่าหากค่าสัมประสิทธิ์มีขนาดเล็กเกินไปการแปลงค่าสัมประสิทธิ์กลับไปเป็นตำแหน่งจุดสังเกตบนใบหน้าจะมีรายละเอียดของตำแหน่งที่น้อยเกินไป ทำให้เกิดความคลาดเคลื่อนของการทำนายมากขึ้น ในทางตรงกันข้ามหากค่าสัมประสิทธิ์นั้นมีขนาดใหญ่เกินไปโครงข่ายประสาทเทียมอาจเรียนรู้ข้อมูลที่จำเป็นต่อการทำนายตำแหน่งจุดสังเกตซึ่งสามารถทำให้ความแม่นยำในการระบุตำแหน่งลดลงได้

ตารางที่ 2 ประสิทธิภาพของโครงข่ายประสาทเทียมที่เรียนรู้ด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลักด้วยขนาดต่าง ๆ

ขนาดค่าสัมประสิทธิ์	อัตราค่าเฉลี่ยความคลาดเคลื่อน	
	ผลต่างสัมบูรณ์ค่าสัมประสิทธิ์	ผลต่างกำลังสองค่าสัมประสิทธิ์
5	8.267	8.307
10	7.012	7.488
15	5.723	5.943
20	6.208	6.220
25	7.474	7.579
30	8.073	8.768

การเปรียบเทียบประสิทธิภาพของการใช้ค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลัก

ผู้วิจัยกำหนดการเรียนรู้โครงข่ายประสาทเทียมด้วยฟังก์ชันสูญเสียในรูปแบบต่าง ๆ 4 รูปแบบ และเปรียบเทียบประสิทธิภาพของแบบจำลองดังตารางที่ 3

ตารางที่ 3 เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมที่ทำนายตำแหน่งจุดสังเกตบนใบหน้าโดยตรงและค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลัก

ฟังก์ชันสูญเสีย	อัตราค่าเฉลี่ยความคลาดเคลื่อน
ผลต่างสัมบูรณ์จุดสังเกต	6.542
ผลต่างกำลังสองจุดสังเกต	7.028
ผลต่างสัมบูรณ์ค่าสัมประสิทธิ์	5.723
ผลต่างกำลังสองค่าสัมประสิทธิ์	5.943

จากตารางที่ 3 การใช้ค่าผลต่างสัมบูรณ์ให้ค่าอัตราค่าเฉลี่ยความคลาดเคลื่อนที่น้อยกว่าการใช้ค่าผลต่างกำลังทั้งสำหรับการทำนายจุดสังเกตโดยตรงและการทำนายค่าสัมประสิทธิ์ สำหรับการเปรียบเทียบการทำนายจุดสังเกตโดยตรงและการทำนายค่าสัมประสิทธิ์เราจะเห็นได้ว่าการทำนายค่าสัมประสิทธิ์สามารถลดค่าอัตราค่าเฉลี่ยความคลาดเคลื่อนได้ดีกว่าทั้งสำหรับการใช้ค่าสัมบูรณ์และค่าผลต่างกำลังสอง

การเรียนรู้โครงข่ายประสาทเทียมเชิงลึกเพื่อการระบุตำแหน่งจุดสังเกตบนใบหน้าด้วยการทำนายค่าสัมประสิทธิ์การวิเคราะห์หองค์ประกอบหลัก

อภิปรายผลการวิจัย

งานวิจัยนี้ได้ทดลองการใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักในการเรียนรู้โครงข่ายประสาทเทียมคอนโวลูชันเพื่อระบุจุดสังเกตบนใบหน้าบนฐานข้อมูล 300W จากการทดลองพบว่าการใช้ฟังก์ชันผลต่างสัมบูรณ์มีประสิทธิภาพที่ดีกว่าการใช้ฟังก์ชันผลต่างกำลังสองซึ่งสอดคล้องกับงานวิจัยของ Feng และคณะ [17,18] และ Wang และคณะ [19] ที่ได้แสดงให้เห็นว่าฟังก์ชันผลต่างกำลังสองที่มักเป็นที่นิยมสำหรับการแก้ปัญหาค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักนั้นไม่เหมาะสมต่อการทำนายตำแหน่งจุดสังเกตโดยเฉพาะอย่างยิ่งในรูปภาพที่ใบหน้าถูกบิดบัง หรือ มีตำแหน่งของจุดสังเกตที่ผิดแปลกจากปกติ นอกจากนี้การทดลองยังสรุปได้ว่าการทำนายค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักนั้นมีประสิทธิภาพมากกว่าการทำนายจุดสังเกตโดยตรง ทั้งนี้เนื่องจากการแปลงค่าสัมประสิทธิ์กลับไปเป็นตำแหน่งจุดสังเกตบนใบหน้าด้วยการคูณกับแม่แบบนั้นให้ตำแหน่งจุดสังเกตบนใบหน้าชัดเจนและเสถียร ในทางตรงกันข้ามการทำนายจุดสังเกตโดยตรงมักมีความผิดพลาดในการทำนายตำแหน่งจุดสังเกตที่ยาก ผลการวิจัยนี้มีความสอดคล้องกับงานวิจัยของ Liu และคณะ [20] และ Goulart และคณะ [21] ซึ่งได้แสดงให้เห็นว่าการปรับใช้การวิเคราะห์องค์ประกอบเป็นส่วนหนึ่งของระบบสามารถลดขนาดของแบบจำลองและช่วยเพิ่มความเสถียรต่อสิ่งรบกวน ซึ่งส่งผลให้ระบบมีประสิทธิภาพในการทำนายโดยที่ดีขึ้น

ในอนาคตงานวิจัยนี้สามารถนำไปต่อยอดด้วยการนำไปปรับใช้กับการระบุตำแหน่งของจุดสังเกตในชุดข้อมูลประเภทอื่น เช่น การระบุตำแหน่งรอยง่ามของคน และการระบุตำแหน่งองค์ประกอบของรถยนต์ เป็นต้น

สรุปผลการวิจัย

งานวิจัยนี้ได้ศึกษาแนวทางในการปรับใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักในการเรียนรู้โครงข่ายประสาทเทียมคอนโวลูชันสำหรับการแก้ปัญหาระบุจุดสังเกตบนใบหน้า จากผลการทดลองบนฐานข้อมูล 300W สรุปได้ว่าการใช้ผลต่างสัมบูรณ์ของค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักนั้นมีประสิทธิภาพในการเรียนรู้โครงข่ายประสาทเทียมมากกว่าการใช้ผลต่างสัมบูรณ์ของจุดสังเกตบนใบหน้าโดยตรง โดยขนาดที่เหมาะสมของค่าสัมประสิทธิ์อยู่ที่ 15 ซึ่งให้อัตราค่าเฉลี่ยความคลาดเคลื่อนอยู่ที่ 5.723 เมื่อเปรียบเทียบกับผลต่างสัมบูรณ์ของจุดสังเกตที่ให้อัตราค่าเฉลี่ยความคลาดเคลื่อนที่ 6.542 ผลการทดลองยืนยันว่าการใช้ค่าสัมประสิทธิ์การวิเคราะห์องค์ประกอบหลักสามารถทำให้โครงข่ายประสาทเทียมมีความเสถียรและมีประสิทธิภาพในการระบุตำแหน่งจุดสังเกตในรูปภาพที่ใบหน้าถูกบิดบังได้ดีขึ้น

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนทุนจากโครงการสนับสนุนทุนนักวิจัยใหม่ วท. ประจำปี พ.ศ. 2564 ขอขอบคุณคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครศรีธรรมราช ที่ให้การสนับสนุนสถานที่และอุปกรณ์สำหรับการทำวิจัย

เอกสารอ้างอิง

- [1] Jacquet, M., & Champod, C. (2020). Automated face recognition in forensic science: Review and perspectives. *Forensic science international*, 307, 110124.
- [2] Wu, X., He, R., Sun, Z., & Tan, T. (2018). A light CNN for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13, 2884- 2896.
- [3] Ko, B. (2018). A brief review of facial emotion recognition based on visual information. *Sensors*, 18(2), 401.

- [4] Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., & Cohen-Or, D. (2021). Encoding in style: a stylegan encoder for image-to-image translation (pp. 2287-2296). In *Conference on Computer Vision and Pattern Recognition (CVPR)*. USA.
- [5] Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation (pp 8789-8797). In *Conference on Computer Vision and Pattern Recognition (CVPR)*. USA.
- [6] Li, C., Wang, R., Li, J., & Fei, L. (2020). Face detection based on YOLOv3 (pp 277-284). In *International Conference on Intelligent Computing, Communication and Devices (ICCD)*. CHINA.
- [7] เพชระ อินทนนท์, และ สุรศักดิ์ มั่งสิงห์. (2563). การทดสอบสมรรถนะการตรวจจับของโมเดลการตรวจจับจุดสังเกตบนใบหน้า. *วารสารวิชาการชายน์เทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต*, 4(2), 13-22.
- [8] Li, X., Lai, S., & Qian, X. (2021). DBCFace: Towards PURE convolutional neural network face detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(4), 1792-1804.
- [9] Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks (pp 4396-4405). In *Conference on Computer Vision and Pattern Recognition (CVPR)*. USA.
- [10] Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6), 681-685.
- [11] Cristinacce, D. & Cootes, T. F. (2006). Feature detection and tracking with constrained local models (pp 95.1-95.10). In *the British Machine Vision Conference (BMVC)*. UK.
- [12] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- [13] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection (pp 886-893). In *Conference on Computer Vision and Pattern Recognition (CVPR)*. USA.
- [14] Jolliffe, I. T. (2002). *Principal Component Analysis (2nd edition)*. New York: Springer Science & Business Media.
- [15] Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge (pp 397-403). In *International Conference on Computer Vision Workshop (ICCV-W)*. Australia.
- [16] Sagonas, C., Antonakos, E., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2016). 300 Faces in-the-wild challenge: database and results. *Image and Vision Computing*, 47, 3-18.
- [17] Feng, Z., Kittler, J., Awais, M., Huber, P., & Wu, X. (2018). Wing loss for robust facial landmark localisation with convolutional neural networks (pp 2235-2245). In *Conference on Computer Vision and Pattern Recognition (CVPR)*. USA.
- [18] Feng, Z. H., Kittler, J., Awais, M., & Wu, X. J. (2020). Rectified wing loss for efficient and robust facial landmark localisation with convolutional neural networks. *International Journal of Computer Vision*, 128(8), 2126-2145.
- [19] Wang, X., Bo, L., & Fuxin, L. (2019). Adaptive wing loss for robust face alignment via heatmap regression. (pp 6971-6981). In *International Conference on Computer Vision (ICCV)*. Korea.

- [20] Liu, Y., Cao, Y., Li, Y., Liu, M., Song, R., Wang, Y., Zhigang X., & Ma, X. (2016). Facial expression recognition with PCA and LBP features extracting from active facial patches. (pp. 368-373). In *IEEE International Conference on Real-time Computing and Robotics (RCAR)*. Cambodia.
- [21] Goulart, C., Valadão, C., Delisle-Rodriguez, D., Funayama, D., Favarato, A., Baldo, G., Vinicius, B., Eliete, C., & Bastos-Filho, T. (2019). Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction. *Sensors*, 19(13), 2844-2868.