

การจำแนกภาพสลิปโอนเงินและอ่านอักขระด้วยแสง Image Classification of Transfer Slips and Optical Character Recognition

ชญาณนท์ อีสสาอาด¹ สัจจาภรณ์ ไวจรรยา^{1*} ณัฐโชติ พรหมฤทธิ์¹ และ มนุ มากมณี²
Chayanon Issaard¹ Sajjaporn Waijanya^{1*} Nuttachot Promrit¹ and Manu Markmanee²

¹ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร จ.นครปฐม ประเทศไทย

²บริษัท บราเธอร์ แอนด์ บราเธอร์ จำกัด จ.นนทบุรี ประเทศไทย

¹Department of Computing, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand

²Brother and Brother Company Limited, Nonthaburi, Thailand

วันที่ส่งบทความ : 23 มกราคม 2567 วันที่แก้ไขบทความ : 6 ตุลาคม 2568 วันที่ตอบรับบทความ : 8 ตุลาคม 2568

Received: 23 January 2024, Revised: 6 October 2025, Accepted: 8 October 2025

บทคัดย่อ

บทความนี้นำเสนอการจำแนกภาพสลิปโอนเงินและการอ่านอักขระด้วยแสง (Optical Character Recognition: OCR) โดยใช้เทคนิคการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ ซึ่งใช้สถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ที่ออกแบบเฉพาะ สำหรับการจำแนกภาพสลิปโอนเงิน โดยประกอบด้วยชั้น Convolutional 2D จำนวน 3 ชั้น ร่วมกับการใช้ Max pooling เพื่อป้องกันการเกิด Overfitting ในการจำแนกภาพสลิปโอนเงิน (Classification) จากนั้นจึงใช้เทคนิคการแบ่งส่วนภาพ (Image segmentation) ด้วย YOLOv5 สำหรับการตรวจสอบชื่อของธนาคาร และใช้เทคนิคการอ่านอักขระด้วยแสงด้วยเทคโนโลยี Tesseract OCR ในการอ่านข้อมูลบนสลิปโอนเงิน ในการทดสอบได้แบ่งชุดข้อมูลแบบสุ่มสำหรับการฝึกฝน (Training set) คิดเป็น 80% และชุดข้อมูลทดสอบ (Test set) 20% โดย 20% ของชุดข้อมูลฝึกฝนถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบความถูกต้อง (Validation set) แบบจำลองได้รับการฝึกฝนโดยกำหนดค่า Batch size เท่ากับ 32 และกำหนด Epochs จำนวน 50 รอบ พบว่าการจำแนกภาพสลิปโอนเงินที่ดีที่สุดได้ค่าความถูกต้องสูงสุดเท่ากับ 99% หลังจากนั้น นำภาพที่ได้จากการจำแนกมาตรวจหาชื่อของธนาคาร โดยใช้การแบ่งส่วนภาพ ก่อนนำภาพไปอ่านอักขระด้วยแสง และแสดงผลข้อมูลในรูปแบบที่มนุษย์อ่านเข้าใจได้ และส่งออกในรูปแบบ JSON

คำสำคัญ : โครงข่ายประสาทเทียมแบบคอนโวลูชัน การอ่านอักขระด้วยแสง การประมวลผลภาพ การแบ่งส่วนของภาพ

*ที่อยู่ติดต่อ E-mail address: waijanya_s@silpakorn.edu

<https://doi.org/10.55003/scikmitl.2025.261945>

Abstract

This article presents the classification of transfer slip images and Optical Character Recognition (OCR) use of automatic learning techniques that mimic the human neural network. The approach utilizes a Convolutional Neural Network (CNN) algorithm specifically designed for classifying transfer slip images, comprising three 2D convolutional layers combined with Max Pooling to prevent overfitting during the classification process. Subsequently, image segmentation is performed using YOLOv5 to identify the bank names, followed by OCR using Tesseract OCR technology to read the information on the transfer slips. The dataset was randomly split into a training set (80%) and test set (20%). Within the training set, 20% of the data was used as a validation set. The model was trained using a batch size of 32 and 50 epochs, achieving a classification accuracy of 99%. After classification, the segmented images were used to identify the bank names, and once identified, the images were subjected to OCR to extract the text. The results were displayed in a human-readable format and exported in JSON format.

Keywords: Convolutional Neural Network (CNN), Optical Character Recognition (OCR), Image Processing, Image Segmentation

1. บทนำ

ทุกวันนี้การใช้ธนาคารออนไลน์ (Online banking) และโมบายแบงก์กิ้ง (Mobile banking) ได้รับความนิยมในการใช้จ่ายมากขึ้น ทำให้ไม่จำเป็นต้องพกพาเงินสดเพื่อความสะดวกในการใช้จ่าย สังเกตได้จากการที่ร้านค้าเริ่มมีการแสดงคิวอาร์โค้ด (QR code) ของบัญชีไว้ที่หน้าร้าน เพื่อให้ผู้ซื้อชำระเงินผ่านทาง Mobile banking เมื่อชำระเสร็จแล้ว ผู้ซื้อจะได้สลิปโอนเงินเป็นหลักฐาน ซึ่งผู้ขายมักขอถ่ายภาพสลิปการโอนเงินเพื่อบันทึกข้อมูลและยืนยันยอดเงิน อย่างไรก็ตาม การจัดเก็บข้อมูลสลิปโอนเงินนั้นมีจำนวนมาก และต้องใช้มนุษย์ในการจัดเก็บข้อมูล ทำให้เกิดข้อผิดพลาดในการทำงานในด้านความถูกต้องของข้อมูลและทำให้กระบวนการล่าช้า อีกทั้งยังใช้คนจำนวนมากเพื่อให้ได้ข้อมูลที่รวดเร็วขึ้น ซึ่งยังส่งผลให้มีต้นทุนด้านค่าใช้จ่ายในการจ้างคนสูงตามมา

เพื่อแก้ปัญหาดังกล่าว ในปัจจุบันจึงมีการใช้เทคนิคการอ่านอักขระด้วยแสง (Optical Character Recognition: OCR) มาประยุกต์ใช้ในการเก็บข้อมูลอย่างแพร่หลาย เนื่องจากการอ่านอักขระด้วยแสงนั้นเป็นเครื่องมือที่สามารถแปลงข้อมูลจากหลากหลายแหล่ง ไม่ว่าจะเป็นข้อมูลที่มาจากภาพ เอกสาร สิ่งพิมพ์ อิเล็กทรอนิกส์ หรือเอกสารที่สแกนเข้ามาในรูปแบบการพิมพ์ลายมือ ให้อยู่ในรูปแบบดิจิทัลที่ง่ายต่อการนำไปวิเคราะห์หรือปรับปรุงและนำไปใช้ได้ทันที โดยเทคนิคการอ่านอักขระด้วยแสงนั้นในปัจจุบันได้รับการพัฒนาให้รองรับได้หลากหลายภาษามากขึ้นรวมถึงภาษาไทยด้วย

การสร้างปัญญาประดิษฐ์ (Artificial Intelligence) เพื่อเพิ่มประสิทธิภาพให้เทคนิคการอ่านอักขระด้วยแสงนั้น มีการพัฒนาอย่างต่อเนื่องเพื่อเพิ่มความแม่นยำและความถูกต้องของการตรวจจับตัวอักษรในหลากหลายภาษาพร้อมกัน ดังเห็นได้จากงานวิจัยที่เกี่ยวข้องกับการอ่านอักขระด้วยแสงพร้อมกันหลาย ภาษา (Park et al., 2020) และความสามารถในการอ่านอักขระด้วยแสงในภาษาไทยและภาษาอังกฤษ (Tanprasert et al., 1999) โดยใช้ Self-Organizing Feature Map (SOM) ซึ่งเป็นเครือข่ายประสาทเทียมแบบที่ไม่ต้องใช้การกำกับดูแล มาช่วยจัดระเบียบข้อมูลรับเข้าอย่างเป็นธรรมชาติ โดย SOM จะสร้างแผนที่คุณลักษณะเชิงพื้นที่จากข้อมูลที่ได้รับ สำหรับขั้นตอนแรกในการจำแนกตัวอักษรมีการใช้ Kohonen's Map ในการแบ่งกลุ่มตัวอักษรเบื้องต้น เพื่อให้สามารถจับลักษณะและโครงสร้างของตัวอักษรที่มีความคล้ายคลึงกันได้อย่างชัดเจน จากนั้นจึงนำกระบวนการแพร่กระจายย้อนกลับเพื่อมาปรับปรุงความแม่นยำในการจำแนกตัวอักษรอย่างละเอียดและเพิ่มประสิทธิภาพของระบบโดยรวม พร้อมทั้งพัฒนาเทคโนโลยี OCR ให้สามารถรองรับการรู้จำตัวอักษรภาษาไทยได้อย่างมีประสิทธิภาพมากขึ้น (Iamsamang & Kongkachandra, 2013) และใช้เทคนิคการแยกตัวอักษรโดยใช้การวิเคราะห์ถดถอยเชิงเส้นและการเรียนรู้ด้วยต้นไม้ตัดสินใจ สำหรับการตรวจสอบระดับการเขียนในภาษาไทยที่มีอยู่ 4 ระดับ เพื่อหาตำแหน่งของอักษรที่สัมผัสกัน โดยเริ่มจากการวิเคราะห์ตัวอักษรที่สัมผัสกันแนวดิ่งและแนวนอน จากนั้นการสร้างแบบจำลองตัวอักษรด้วยต้นไม้ตัดสินใจ และในปัจจุบันได้มีการประยุกต์ใช้การอ่านอักขระด้วยแสงในการตรวจจับคำพูดในภาพอีกด้วย (Lestari & Mulyana, 2022)

จากปัญหาที่พบว่ามี การเก็บข้อมูลที่ล่าช้าและมีความคลาดเคลื่อนในการเก็บข้อมูล ได้มีงานวิจัยหลากหลายชิ้นได้เสนอแนวทางแก้ไขโดยใช้เทคนิคต่าง ๆ เช่น Convolutional Neural Network (CNN) เพื่อจำแนกประเภทใบแจ้งหนี้ (Arslnid & Uymaz, 2022) ทั้งนี้เมื่อสามารถจำแนกประเภทของสลิปใบเงินได้แล้ว ผู้วิจัยยังมีความต้องการที่จะระบุที่มาของธนาคารผู้โอนจึงได้ศึกษาเพิ่มเติมในเรื่องการแบ่งสัดส่วนของภาพ (Image segmentation) (Rezkiani et al., 2022) ซึ่งจำแนกประเภทของเครื่องหมายบนเอกสารด้วยอัลกอริทึม You Only Look Once (YOLO) โดยใช้ Darknet Framework ซึ่งจากงานวิจัยพบว่า YOLOv4 ที่ดีที่สุด สามารถตรวจจับและจดจำโลโก้ด้วยค่าเฉลี่ยความแม่นยำ (mean Average Precision: mAP) ที่ 93.73% จึงได้มีการพัฒนาต่อยอดโดยนำผลการศึกษาดังกล่าวมาประยุกต์ใช้ร่วมกับ YOLOv5 พร้อมปรับแต่งการแบ่งสัดส่วนของภาพให้เหมาะสมกับลักษณะเฉพาะของข้อมูลที่ใช้ ซึ่งมีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพในการจำแนกประเภทและตรวจจับวัตถุในภาพให้ดียิ่งขึ้น หลังจากจำแนกประเภทสลิปและชื่อธนาคารได้แล้ว มีการใช้เทคนิค OCR เข้ามาาร่วมด้วยหลังจากการจำแนก (Vaidya et al., 2019) เพื่อการทำแบบฟอร์มการกรอกข้อมูลอัตโนมัติสำหรับธนาคารเพื่อเก็บข้อมูล โดยใช้ CNN ร่วมกับ OCR และยังมีงานวิจัยที่เกี่ยวข้องกับการอ่านตัวอักษรที่เขียนด้วยลายมือจึงเกิดการฝึกฝนให้สามารถเรียนรู้จำตัวอักษรที่เขียนด้วยลายมือ (Memon et al., 2020) และหากต้องการให้ภาพที่จะใช้การอ่านอักขระด้วยแสง มีความชัดเจนมากขึ้นและง่ายต่อการเรียนรู้จำ จึงมีการใช้การแปลงภาพสีเป็นภาพระดับเทา (Grayscale) และเพื่อป้องกันการลดคอนทราสต์ของภาพ ความชัด และเงา จึงได้เสนอการใช้ อัลกอริทึมใหม่ที่มีการประมาณค่าสี (Red Green Blue: RGB) การลดและการเพิ่ม Chrominance โดยในงานวิจัยนี้ได้ยืนยันว่าได้รักษาคุณสมบัติเด่นของภาพสี เช่น คอนทราสต์ความคมชัดเงาและโครงสร้างภาพ (Saravanan, 2010) และยังมีการใช้เทคนิคการประมวลผลภาพเข้ามาช่วยเพื่อเป็นการเตรียมภาพให้พร้อม

ก่อนที่จะนำไปอ่านอักขระด้วยแสงด้วยการแปลงภาพให้เป็นภาพสีเทา และใช้ Threshold เข้ามาช่วยในการดึงค่าลักษณะของตัวอักษรและการลบสัญญาณรบกวนของภาพออก โดยการลดขนาดภาพเป็นมาตรฐาน Total variation และตรวจสอบลักษณะของ Salt and Pepper Noise (SPN) เพื่อตรวจจับจุดภาพที่เกิดสัญญาณรบกวนและปรับให้เป็นมาตรฐานให้แบบ Total variation เพื่อกู้คืนจุดภาพที่มีสัญญาณรบกวน (Hien et al., 2022)

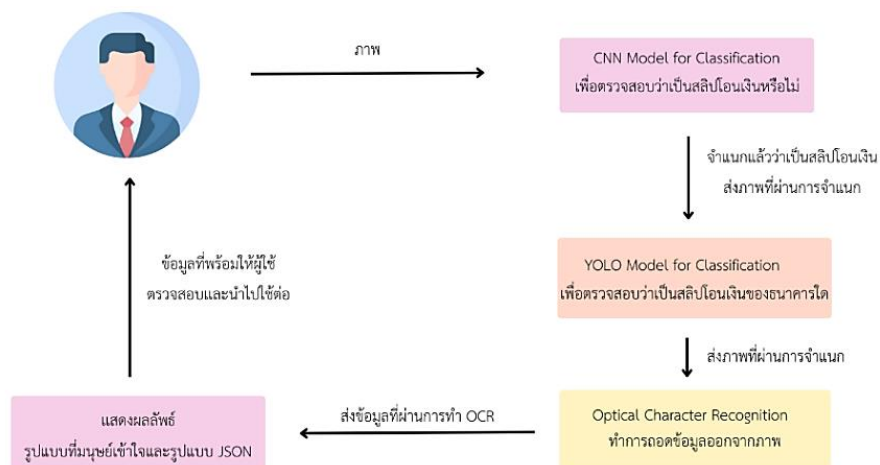
จากปัญหาข้างต้นที่เกิดปัญหาในด้านความล่าช้าและความผิดพลาด ผู้วิจัยจึงได้พัฒนาแนวทางใหม่ในการจำแนกภาพสลิปโอนเงินและการอ่านอักขระด้วยแสง เพื่อแก้ไขข้อผิดพลาดและลดระยะเวลาในการดำเนินการ โดยในงานวิจัยนี้ได้ใช้เทคโนโลยี CNN ในการจำแนกภาพสลิปโอนเงิน และใช้ YOLOv5 สำหรับการแบ่งส่วนภาพและตรวจจับชื้อธนาคาร ซึ่งเมื่อได้ชื้อธนาคารแล้วจะนำไปผ่านกระบวนการ OCR ด้วย Tesseract เพื่อถอดข้อมูลตัวอักษรที่ปรากฏบนสลิปโอนเงิน กระบวนการทั้งหมดนี้มีวัตถุประสงค์เพื่อเพิ่มความสะดวก ความรวดเร็ว และความถูกต้องในการจัดเก็บข้อมูลให้แก่ผู้ใช้งานอย่างมีประสิทธิภาพ

2.วิธีดำเนินการวิจัย

ส่วนนี้จะอธิบายภาพรวมการทำงานของระบบทั้งหมด ครอบคลุมตั้งแต่ขั้นตอนการจัดเก็บข้อมูลสำหรับการฝึกฝนทั้งสองโมเดล การอธิบายโครงสร้างของโมเดลที่ใช้ การปรับปรุงคุณภาพของภาพก่อนนำไปใช้ในกระบวนการถอดข้อความ เครื่องมือที่ใช้ในการถอดข้อความออกจากภาพ ไปจนถึงการนำผลลัพธ์สุดท้ายไปใช้งานต่อ

2.1 ภาพการทำงานโดยรวม

ระบบนี้มีจุดประสงค์เพื่อจำแนกสลิปโอนเงินโดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) และการจำแนกชื้อของธนาคารโดยใช้ YOLOv5 รวมถึงการใช้เทคโนโลยีการอ่านอักขระด้วยแสงเพื่อถอดข้อมูลจากสลิปโอนเงิน โดยภาพรวมการทำงานของระบบแสดงได้ดังรูปที่ 1



รูปที่ 1. การทำงานโดยรวมของระบบ

รูปที่ 1 แสดงภาพรวมการทำงานของระบบ ซึ่งเริ่มต้นจากการรับภาพจากผู้ใช้ แล้วใช้ CNN เพื่อจำแนกว่าเป็นภาพสลิปโอนเงินหรือไม่ จากนั้น ระบบจะตรวจสอบว่าสลิปดังกล่าวเป็นของธนาคารใด ขั้นตอนสุดท้าย คือการนำภาพไปถอดข้อมูลด้วย OCR และแสดงผลลัพธ์ 2 รูปแบบ ได้แก่ 1) รูปแบบที่มนุษย์เข้าใจ และ 2) รูปแบบ JSON เพื่อให้ง่ายต่อการนำข้อมูลไปใช้งานต่อ

2.2 การเก็บข้อมูล

การทดลองนี้เก็บข้อมูลแบบสุ่ม โดยได้รับข้อมูลภาพจากบริษัท Brother and Brother Company Limited ซึ่งประกอบด้วยข้อมูล 2 ส่วนคือ ภาพที่ไม่ใช่สลิปโอนเงิน (Not transfer slip) จำนวน 1,500 ภาพ และภาพสลิปโอนเงิน (Transfer slip) จำนวน 1,500 ภาพ

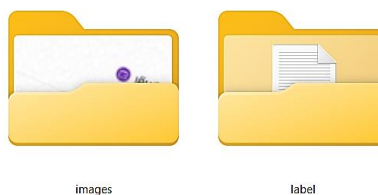
2.3 การเตรียมข้อมูลสำหรับโมเดล

การเตรียมข้อมูลสำหรับใช้ในโมเดลใช้ทั้งหมด 3,000 ภาพ โดยแบ่งเป็นภาพสลิปโอน (Slip) 1,500 ภาพ และภาพที่ไม่ใช่สลิปโอนเงิน (Notslip) 1,500 ภาพ ดังรูปที่ 2



รูปที่ 2. การแบ่งข้อมูล

การเตรียมข้อมูลสำหรับใช้ในโมเดลใช้ทั้งหมด 2,064 ไฟล์ โดยแบ่งเป็นภาพทั้งหมด (images) 1,032 ภาพ และไฟล์ป้ายกำกับ (label) ทั้งหมด 1,032 ไฟล์ ดังรูปที่ 3



รูปที่ 3. ข้อมูลที่ใช้ในการเทรนโมเดล YOLO

มีการชุดข้อมูลสำหรับการฝึกฝน (Train set) 80% และชุดข้อมูลทดสอบ (Test set) 20% โดย 20% ของชุดข้อมูลฝึกฝนถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบความถูกต้อง (Validation set) และปรับโมเดลตามผลการทดสอบโดยในแต่ละ Class แบ่งเป็นภาพที่ไม่ใช่สลิปโอนเงิน 1,500 ภาพ และภาพสลิปโอนเงิน 1,500 ภาพ โดยตัวอย่างภาพของสลิปโอนเงินแสดงดังรูปที่ 4 และตัวอย่างภาพที่ไม่ใช่สลิปโอนเงินแสดงดังรูปที่ 5



รูปที่ 4. ตัวอย่างภาพของสลิปโอนเงิน



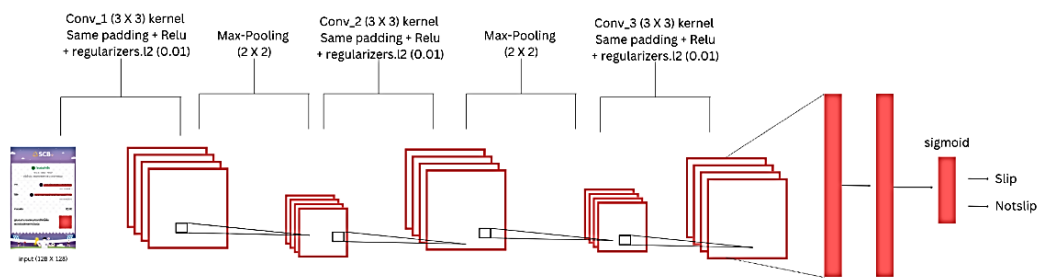
รูปที่ 5. ตัวอย่างภาพที่ไม่ใช่สลิปโอนเงิน

2.4 โมเดลจำแนกประเภทสลิปโอนเงิน

โมเดลที่ใช้ในการจำแนกสลิปโอนเงินเป็นสถาปัตยกรรม CNN ที่ออกแบบขึ้นใหม่ โดยมีโครงสร้างที่ประกอบด้วยชั้น Convolutional 2D จำนวน 3 ชั้น ซึ่งแต่ละชั้นจะตามด้วยชั้น Max Pooling เพื่อลดการ Overfitting โดยทดลองนำเข้าภาพ จำนวน 2 ชุด ในการนิยามโมเดลมีการกำหนด Input shape = (128,128,3) และ Output layer เป็น 2 โหนด คือ Slip และ Notslip ค่า Regularizers เท่ากับ 0.01 และ Dropout เท่ากับ 0.25 ใช้ฟังก์ชัน Activation Function แบบ ReLU และ Output layer ใช้ Activation Function แบบ Sigmoid ส่วนของ Optimizer ใช้เป็น Adam และมีการกำหนด Learning Rate เท่ากับ 0.001 ค่า Batch size เท่ากับ 32 และกำหนด Epochs เท่ากับจำนวน 50 รอบ เพื่อให้โมเดลมีโอกาสเรียนรู้ข้อมูลได้อย่างเพียงพอ

ในการทดลองกำหนดจำนวนรอบการฝึกนี้ช่วยให้โมเดลสามารถจับลักษณะสำคัญของข้อมูลได้อย่างสมบูรณ์ ทั้งนี้ จำนวน Epochs ที่กำหนดได้รับการพิจารณาให้เหมาะสมกับข้อมูลภาพที่มีอยู่ในแต่ละคลาส ซึ่งแบ่งออกเป็น 2 คลาส (Slip และ Notslip) คลาสละ 1,500 ภาพ เพื่อหลีกเลี่ยงปัญหา Underfitting ที่โมเดลอาจเรียนรู้ข้อมูลได้ไม่เพียงพอ และลดโอกาสเกิด Overfitting ที่โมเดลอาจเรียนรู้ข้อมูลเฉพาะเจาะจงมากเกินไป ร่วมกับการใช้เทคนิค Dropout และการปรับค่า Regularizers เพื่อเพิ่มประสิทธิภาพของโมเดลในการทำนายข้อมูลใหม่ และเมื่อกำหนดค่าให้โมเดลเริ่มการจำแนกแล้ว ผลลัพธ์จากการนำภาพเข้าป้อนนั้นได้คำตอบออกมาเป็น “สลิปโอนเงิน” หรือ “ไม่ใช่สลิปโอนเงิน”

โมเดลจำแนกประเภทสลิปโอนเงินมีขั้นตอนการทำงานแสดงได้ดังรูปที่ 6 โดยโมเดลสามารถเรียนรู้และได้ค่าความถูกต้องสูงสุดที่ 99% ซึ่งแสดงให้เห็นว่า โมเดลมีประสิทธิภาพในการจำแนกสลิปโอนเงินได้อย่างแม่นยำ

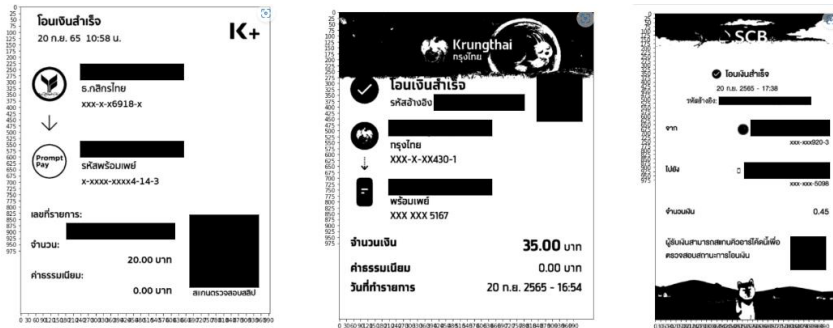


รูปที่ 6. โมเดลจำแนกประเภทสลิปโอนเงิน

2.5 การสร้างโมเดลจำแนกชื่อธนาคาร

สร้างโมเดลจำแนกชื่อธนาคาร โดยใช้ Image segmentation และโมเดล YOLOv5 เพื่อแยกและระบุตำแหน่งของเครื่องหมาย (Logo) ของธนาคารบนสลิป โดย YOLOv5 เป็นโมเดลการเรียนรู้เชิงลึกสำหรับการตรวจจับวัตถุบนภาพ ซึ่งสถาปัตยกรรมประกอบด้วย Backbone Network ในการสกัดคุณลักษณะ Neck Network เพื่อรวมคุณลักษณะจากหลายระดับ และ Head Network ที่ทำหน้าที่ทำนายตำแหน่งและจำแนกประเภท โดยมีหลักการทำงานดังนี้ การแบ่งภาพแบบกริด การตรวจจับวัตถุบนภาพ โดยแบ่งเป็นการทำนายกรอบภาพ คำนวณความน่าจะเป็นของคลาสและการกำหนดค่าความเชื่อมั่น และสุดท้ายจะประมวลผลภาพเพื่อกำจัดกรอบซ้ำซ้อน โดยใช้เทคนิค Non-Maximum Suppression (NMS) ทางผู้จัดทำได้นำโมเดลมาฝึกฝนใหม่เพื่อให้สามารถจำแนก Logo ออกเป็น 6 ประเภท ได้แก่ ธนาคารกรุงเทพ (bank_bkk) ธนาคารกรุงไทย (bank_krungthai) ธนาคารออมสิน (bank_omsin) ธนาคารกสิกรไทย (bank_kplus) ธนาคารไทยพาณิชย์ (bank_scb) และธนาคารทหารไทยธนชาต (bank_ttb) ซึ่ง YOLOv5 ทำงานโดยแบ่งภาพเป็นกรอบ ทำนายกรอบขอบเขต (Bounding boxes) และคำนวณความน่าจะเป็นของแต่ละคลาส ด้วยความเร็วและความแม่นยำที่สูง ทำให้สามารถประมวลผลภาพได้อย่างมีประสิทธิภาพ

ตัวอย่างการ Segment ภาพสลิปโอนเงิน แสดงดังรูปที่ 7 ซึ่งเป็นภาพการ Segment สลิปโอนเงินของธนาคารทหารไทยธนชาต



รูปที่ 9. ผลลัพธ์ของการปรับปรุงคุณภาพของภาพ

2.7 การถอดข้อมูลออกจากภาพ

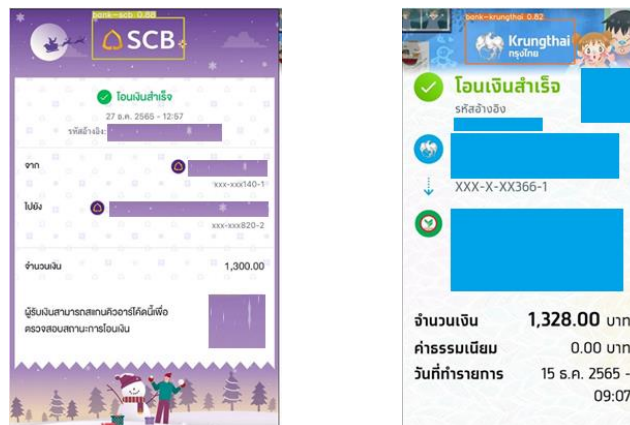
หลังจากการปรับปรุงภาพให้มีความเหมาะสมสำหรับการถอดข้อมูลแล้ว จากนั้นนำภาพมาถอดข้อมูลด้วย Tesseract OCR เพื่อแปลงข้อมูลจากภาพเป็นข้อความ จากนั้นผลลัพธ์ที่ได้จะถูกแปลงให้อยู่ในรูปแบบ JSON เพื่อให้สะดวกต่อการนำไปใช้ต่อ

3. ผลการวิจัยและอภิปรายผล

3.1 ผลการทดสอบประสิทธิภาพในส่วนของการจำแนกชื่อธนาคาร

การทดสอบนี้มีจุดประสงค์เพื่อจำแนกชื่อธนาคาร โดยการใช้ภาพของสลิปโอนเงินของแต่ละธนาคาร มีการรับภาพเข้ามาและแบ่งภาพตามเครื่องหมายการของธนาคาร ซึ่งในการฝึกฝนโมเดล YOLOv5 ใช้ภาพในการทดสอบทั้งสิ้น 760 ภาพ จำนวนรอบในการฝึกฝน 100 รอบ ผลจากการทดสอบพบว่าโมเดลสามารถทำนายได้ถูกต้องสูงสุดอยู่ที่ 99.27%

ตัวอย่างการจำแนกภาพสลิปแสดงดังรูปที่ 10 โดยสามารถตรวจจับเครื่องหมายของธนาคารไทยพาณิชย์ได้ความแม่นยำที่ 88% และเครื่องหมายของธนาคารกรุงไทยมีความแม่นยำอยู่ที่ 82%



รูปที่ 10. การ Segmentation โดยทำผ่านโมเดล YOLOv5 และทำนายชื่อของธนาคารได้อย่างแม่นยำ

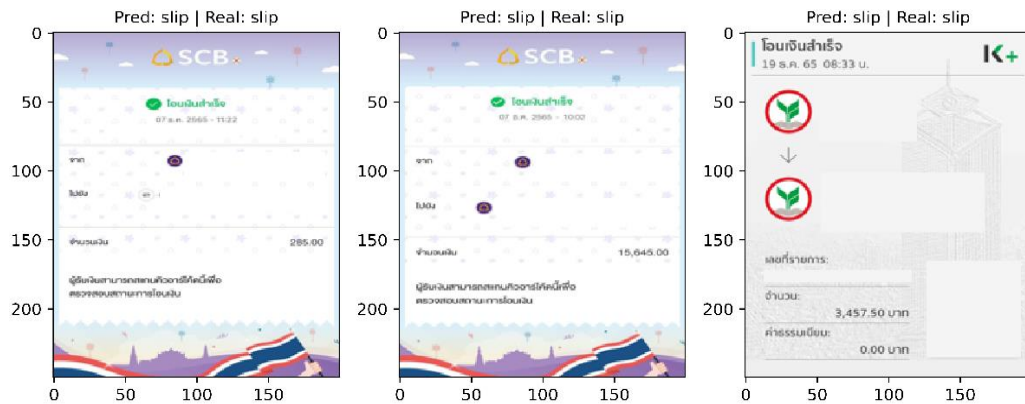
3.2 ผลการทดสอบประสิทธิภาพของการจำแนกภาพที่เป็นสลิปโอนเงิน

การทดสอบประสิทธิภาพและวิเคราะห์ผลความถูกต้องของโมเดลการจำแนกภาพสลิปโอนเงิน โดยใช้ชุดข้อมูลในการฝึกสอนคือ ภาพที่เป็นสลิปโอนเงินและภาพที่ไม่ใช่สลิปโอนเงิน และทดสอบการวัดประสิทธิภาพได้ผล Test set accuracy เท่ากับ 99% และ loss เท่ากับ 3.56%

3.2.1 ตัวอย่างผลการจำแนกภาพที่เป็นสลิปโอนเงิน

ตัวอย่างของการจำแนกภาพสลิปโอนเงินจะแบ่งออกเป็น 2 ประเภท คือ

1) ตัวอย่างผลลัพธ์ที่ทำนายได้ถูกต้อง แสดงดังรูปที่ 11 จากการทดสอบพบว่า ภาพที่ไม่เดลทำนายได้ถูกต้องนั้น เป็นภาพสลิปของธนาคารไทยพาณิชย์ และธนาคารกรุงไทย



รูปที่ 11. ตัวอย่างที่ทำนายได้ถูกต้อง

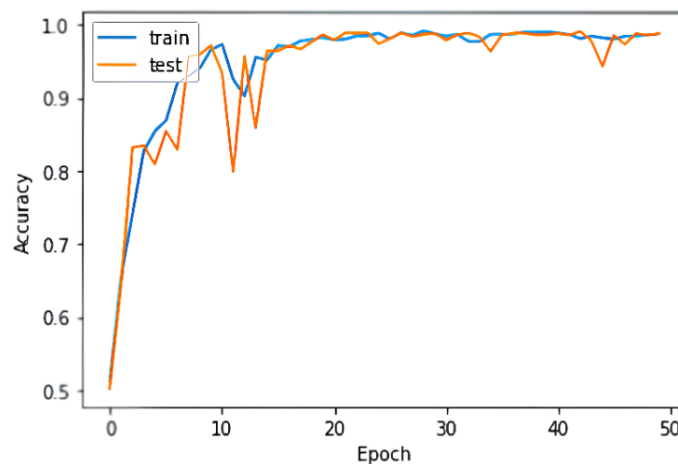
2) ตัวอย่างที่ทำนายได้ไม่ถูกต้อง แสดงดังรูปที่ 12 จากการทดสอบพบว่า ภาพที่ไม่เดลทำนายผิดอาจเนื่องจากในภาพนั้นมีตัวอักษรภาษาไทยหรืออื่น ๆ อยู่ ทำให้โมเดลจำแนกเป็นสลิปโอนเงิน



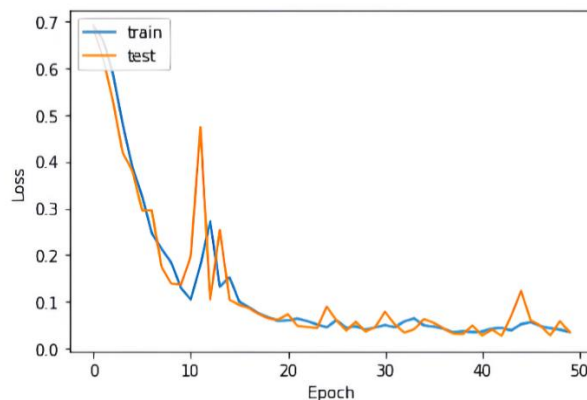
รูปที่ 12. ตัวอย่างที่ทำนายได้ไม่ถูกต้อง

3.2.2 ผลการประเมินประสิทธิภาพโมเดลจากกราฟ

การประเมินประสิทธิภาพของโมเดลจากกราฟ แบ่งออกเป็น 2 ส่วน ได้แก่ กราฟแสดงความแม่นยำ ดังรูปที่ 13 และกราฟแสดงค่าความสูญเสียดังรูปที่ 14 เพื่อวิเคราะห์ว่าโมเดลสามารถเรียนรู้ได้ดีหรือไม่ โดยใช้ข้อมูลจากการเรียนรู้แต่ละรอบมาสร้างกราฟเพื่อเปรียบเทียบผลลัพธ์

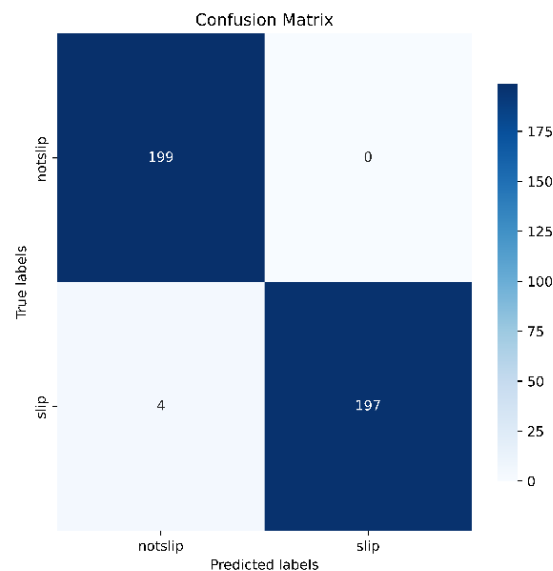


รูปที่ 13. ตัวอย่างที่ทำนายได้ไม่ถูกต้อง



รูปที่ 14. ภาพกราฟเปรียบเทียบ Accuracy และ Loss ของโมเดล

จากรูปที่ 13 และ 14 แสดงกราฟเปรียบเทียบค่า Accuracy และ Loss ของโมเดลที่ได้ฝึกฝน ซึ่งจะเห็นได้ว่ามีค่า Accuracy ที่ค่อนข้างสูง และเกิดการ Overfitting ที่ค่อนข้างต่ำ อีกทั้งหากพิจารณาจากตาราง Confusion Matrix ของโมเดลการจำแนกภาพที่เป็นสลิปโอนเงิน ดังรูปที่ 15



รูปที่ 15. Confusion Matrix ของโมเดลการจำแนกภาพที่เป็นสลิบโอนเงิน

จากรูปที่ 15 เป็นการแสดง Confusion Matrix ของโมเดลการจำแนกภาพที่เป็นสลิบโอนเงินโดยแสดงถึงประสิทธิภาพของโมเดลเพราะมีการทำนายผิดเพียง 4 ครั้ง นอกจากนี้ยังมีการคำนวณค่า Precision Recall F1-score และ Support การจำแนกประเภท ดังแสดงในตารางที่ 1 และตารางที่ 2 แสดงค่า Precision Recall F1 Score และ Support ประสิทธิภาพของโมเดล

ตารางที่ 1. ค่า Precision, Recall, F1-score, Support การจำแนกประเภท

Class	Precision	Recall	F1-score	Support
0 (Notslip)	0.9803	1.000	0.9900	199
1 (Slip)	1.0000	0.9801	0.9899	201

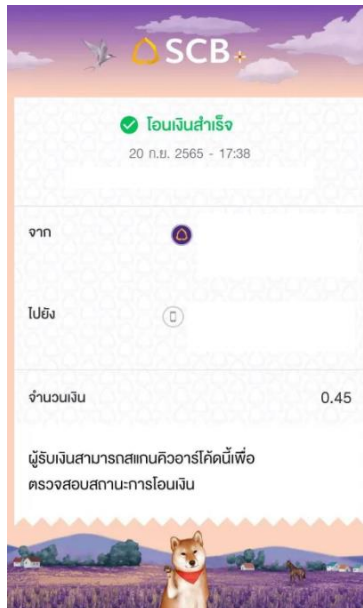
ตารางที่ 2. ค่า Precision, Recall, F1-score, Support ประสิทธิภาพโมเดล

	Precision	Recall	F1-score	Support
Accuracy	-	-	0.9900	400
Macro avg	0.9901	0.9900	0.9900	400
Weighted avg	0.9902	0.9900	0.9900	400

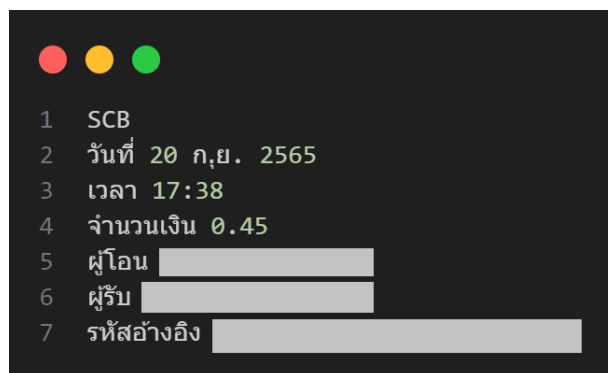
3.3 ผลลัพธ์การทดสอบประสิทธิภาพของการถอดข้อความออกจากภาพ

ในการทดลองนี้ ได้นำภาพที่ผ่านกระบวนการประมวลผลแล้ว ดังตัวอย่างในรูปที่ 16 จากนั้นนำมาทดสอบการถอดข้อความ ผลลัพธ์ที่ได้จากการถอดข้อความจะแสดงใน 2 รูปแบบ คือ 1) รูปแบบที่มนุษย์

สามารถอ่านและทำความเข้าใจได้ ดังรูปที่ 17 และ 2) รูปแบบ JSON ดังตัวอย่างในรูปที่ 18 ซึ่งง่ายต่อการนำข้อมูลไปใช้งานต่อ เช่น การจัดเก็บข้อมูล หรือการยืนยันยอดเงินโอน



รูปที่ 16. ตัวอย่างภาพสลิปโอนเงินที่นำมาทดสอบการใช้ OCR



รูปที่ 17. ผลลัพธ์ของการใช้ OCR ถอดข้อมูลในรูปแบบที่มนุษย์สามารถอ่านแล้วเข้าใจได้



รูปที่ 18. ผลลัพธ์ในรูปแบบ JSON

4. สรุปผลการทดลอง

งานวิจัยนี้เป็นการพัฒนาระบบอ่านสลิปโอนเงินด้วยเทคโนโลยีการอ่านอักขระด้วยแสง (OCR) โดยมีวัตถุประสงค์หลักเพื่อจำแนกภาพสลิปโอนเงินออกจากภาพประเภทอื่น ๆ ในขั้นตอนแรก ระบบจะใช้ อัลกอริทึม CNN ร่วมกับ Regularization ด้วยเทคนิค Dropout เพื่อจำแนกภาพ ซึ่งให้ค่าความถูกต้องสูงสุดเท่ากับ 99% หลังจากนั้น ภาพสลิปจะถูกนำไปใช้ในการจำแนกหาชื่อของธนาคารด้วยโมเดล YOLOv5 ก่อนเข้าสู่กระบวนการประมวลผลด้วยภาพ ซึ่งประกอบด้วย การกำจัดสัญญาณรบกวน การแปลงภาพให้เป็นภาพระดับสีเทา และการแปลงภาพระดับสีเทากลายเป็นไบนารี จากนั้นนำภาพที่ได้ไปเข้าสู่กระบวนการอ่านอักขระด้วยแสง และส่งมอบข้อมูลที่สกัดได้ให้แก่ผู้ใช้และระบบหลังบ้านต่อไป

เอกสารอ้างอิง (References)

- Arsln, Ö., & Uymaz, S. A. (2022). Classification of invoice images by using convolutional neural networks. *Journal of Advanced Research in Natural and Applied Sciences*, 8(1), 8-25.
- Hien, N. N., Thanh, D. N. H., Erkan, U., & Tavares, J. M. R. S. (2022). Image noise removal method based on thresholding and regularization techniques. *IEEE Access*, 10, 71584-71595. <https://doi.org/10.1109/ACCESS.2022.3188315>
- Iamsamang, W., & Kongkachandra, R. (2013). Automatic Thai handwritten characters segmentation based on linear regression analysis and decision tree learning. *Thai Journal of Science and Technology*, 2, 167-174. (in Thai)
- Lestari, I. N. T., & Mulyana, D. I. (2022). Implementation of OCR (optical character recognition) using Tesseract in detecting character in quotes text images. *Journal of Applied Engineering and Technological Science*, 4(1), 58-63. <https://doi.org/10.37385/jaets.v4i1.905>
- Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). *IEEE Access*, 8, 142642-142668. <https://doi.org/10.1109/ACCESS.2020.3012542>
- Park, J., Lee, E., Kim, Y., Kang, I., Koo, H. I., & Cho, N. I. (2020). Multi-lingual optical character recognition system using the reinforcement learning of character segmenter. *IEEE Xplore*, 8, 174437-174446. <https://doi.org/10.1109/ACCESS.2020.3025769>
- Rezkiani, K., Nurtanio, I., & Syafaruddin. (2022). Logo detection using You Only Look Once (YOLO) method. *Proceedings of the 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)* (pp. 227-232). IEEE. <https://doi.org/10.1109/ICE3IS56585.2022.10010121>

- Saravanan, C. (2010). Color image to grayscale image conversion. *Proceedings of the 2010 Second International Conference on Computer Engineering and Applications* (pp. 196-199). IEEE. <https://doi.org/10.1109/ICCEA.2010.192>
- Tanprasert, C., Sinthupinyo, W., Dubey, P., & Tanprasert, T. (1999). Improved mixed Thai & English OCR using two-step neural net classification. *NECTEC Technical Journal*, 1, 41-46.
- Vaidya, M., Rule, P. K., Kumar, H., Jain, A., & Kamble, A. R. (2019). Automating data entry forms for banks using OCR and CNN. *International Journal for Applied Sciences and Engineering Technology*, 7, 890-893.