

การศึกษาทดลองเกี่ยวกับการสร้างลายเซ็นโดยใช้เครือข่ายขัดแย้งกำเนิด

An Experimental Approach to Signature Generation Using Generative Adversarial Networks

คุณานันต์ กลิ่นจันทร์หอม, เดโช ศรีสวัสดิ์, ทรงคมกฤษ ไชยกาล,
ฐิติรัตน์ ศิริบวรรัตนกุล*

Kunanon Klinchanhom, Decho Srisavat, Songkomkrit Chaiyakan,
Thitirat Siriborvornratanakul*

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ กรุงเทพฯ ประเทศไทย
Graduate School of Applied Statistics, National Institute of Development
Administration, Bangkok, Thailand

* Corresponding author E-mail: thitirat@as.nida.ac.th

Received 14 March 2023; Revised 5 April 2025; Accepted 9 April 2025

บทคัดย่อ

ความเป็นมาและวัตถุประสงค์ : ลายเซ็นหรือลายมือชื่อมีความสำคัญอย่างมาก เนื่องจากใช้เป็นสัญลักษณ์ระบุตัวตนของบุคคลได้โดยทั่วไป การออกแบบลายเซ็นมักอาศัยคำแนะนำจากผู้เชี่ยวชาญหรือนักออกแบบที่มีความสามารถเฉพาะด้าน ขณะที่ในปัจจุบัน มีการนำเทคโนโลยีปัญญาประดิษฐ์มาใช้ในการตรวจจับลายเซ็นและการสร้างลายมือเขียน งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการสร้างลายเซ็นจากชื่อภาษาอังกฤษโดยใช้เครือข่ายขัดแย้งกำเนิด (Generative Adversarial Network: GAN)

วิธีดำเนินการวิจัย : งานวิจัยนี้ใช้ชุดข้อมูล IAM Handwriting โดยข้อมูลภาพลายมือเขียนถูกนำเข้าสู่กระบวนการเรียนรู้เชิงลึกโดยใช้ ScrabbleGAN จากนั้น นำลายเซ็นที่สร้างได้มาเปรียบเทียบกับผลลัพธ์จากตัวแบบที่ใช้ Long Short-Term Memory (LSTM) และ Transformer การประเมินผลดำเนินการผ่านการเปรียบเทียบโดยมนุษย์เพื่อพิจารณาความสมจริงของลายเซ็นที่ได้จากแต่ละตัวแบบ

ผลการวิจัย : ผลการทดลองแสดงให้เห็นว่า ScrabbleGAN สามารถสร้างลายเซ็นที่มีความสมจริงในระดับหนึ่ง แต่มีปัญหาในการกำจัดพื้นหลัง ซึ่งส่งผลกระทบต่อคุณภาพของลายเซ็นที่ได้ โดยเมื่อนำไปเปรียบเทียบกับตัวแบบ LSTM และ Transformer พบว่า ตัวแบบทั้งสองมีประสิทธิภาพในการลบพื้นหลังได้ดีกว่า นอกจากนี้ ความสมจริงของลายเซ็นที่สร้างโดย ScrabbleGAN ยังมีคุณภาพด้อยกว่าลายเซ็นที่สร้างจาก LSTM

สรุป : การใช้ ScrabbleGAN ในการสร้างลายเซ็นให้ผลลัพธ์ที่มีความสมจริงระดับหนึ่ง แต่ยังมีข้อจำกัดในด้านการกำจัดพื้นหลังและความถูกต้องของรูปแบบลายเซ็น เมื่อเปรียบเทียบกับเทคนิค LSTM และ Transformer ผลการวิจัยนี้ชี้ให้เห็นถึงความเป็นไปได้ของการใช้เครือข่ายขัดแย้งกำเนิดในการสร้างลายเซ็นจากชื่อภาษาอังกฤษ แต่ยังคงต้องปรับปรุงเพิ่มเติมเพื่อให้ได้ผลลัพธ์ที่น่าพึงพอใจมากขึ้น

การนำไปใช้ประโยชน์ในเชิงปฏิบัติ : ผลลัพธ์จากงานวิจัยนี้สามารถนำไปประยุกต์ใช้ในการพัฒนาระบบสร้างลายเซ็นอัตโนมัติ ซึ่งมีประโยชน์ในงานด้านการลงนามเอกสารอิเล็กทรอนิกส์ การออกแบบลายเซ็นเฉพาะบุคคล และการพัฒนาแอปพลิเคชันที่เกี่ยวข้องกับการสร้างลายมือเขียนด้วยปัญญาประดิษฐ์ นอกจากนี้ ผลการศึกษายังสามารถใช้เป็นแนวทางในการพัฒนาโมเดลสร้างลายเซ็นที่มีคุณภาพสูงขึ้นในอนาคต

คำสำคัญ : การสร้างลายเซ็น, การเรียนรู้เชิงลึก, เครือข่ายขัดแย้งกำเนิด

Abstract

Background and Objectives: Signature plays a crucial role in identity verification, as it serves as a unique representation of an individual. Traditionally, signature design relies on expert guidance from a skilled designer. Meanwhile, artificial intelligence (AI) has currently been predominantly utilized in signature verification and handwritten text generation. The present study therefore aimed to explore the generation of signatures from English names using Generative Adversarial Networks (GANs).

Methodology: The present research employed IAM Handwriting dataset. The dataset was processed through a deep learning framework utilizing ScrabbleGAN. The generated signatures were then compared with those produced by models based on Long Short-Term Memory (LSTM) and Transformer architectures. The evaluation was conducted through human assessment to determine the realism and quality of the generated signatures.

Main Results: The experimental results indicate that ScrabbleGAN was capable of generating relatively realistic signatures. However, it struggled with background removal, which affected the overall quality of the generated outputs. When compared ScrabbleGAN to LSTM and Transformer models, these latter approaches demonstrated superior performance in eliminating background noise. Additionally, the signatures generated by ScrabbleGAN were found to be less visually convincing than those produced by LSTM-based models.

Conclusions: While ScrabbleGAN demonstrates potential in generating signatures from English names, its limitations in background removal and signature authenticity highlight the need for further refinements. The present study suggests that although GAN-based approaches can be utilized for signature generation, additional improvements are required to enhance the realism and consistency of the generated results.

Practical Application: The findings of the present study can contribute to the development of automated signature generation systems, which could be applied in digital document sign-

ing, personalized signature creation and AI-driven handwriting applications. Furthermore, the insights gained from the present research can serve as a foundation for improving signature synthesis models, ultimately leading to higher-quality and more reliable signature generation techniques.

Keywords: Signature Generation, Deep Learning, Generative Adversarial Networks

Introduction

“ลายเซ็น” หรือ “ลายมือชื่อ” เป็นการแสดงชื่อจริง ชื่อเล่น หรือนามแฝง ที่อาจจะประกอบไปด้วยสัญลักษณ์ต่าง ๆ ที่เป็นเอกลักษณ์แสดงถึงลักษณะเฉพาะของบุคคลนั้น ซึ่งลายเซ็นมีทั้งประเภทที่คล้ายกับลายมือ เขียน หรือแตกต่างจากลายมือเขียนโดยมีการตกแต่งสิ่งต่าง ๆ เพิ่มขึ้นตามความพึงพอใจของเจ้าของลายเซ็น และตามวาระโอกาสในการใช้งานลายเซ็นนั้น โดยลายเซ็นสามารถนำไปใช้อ้างอิงในการใช้ประโยชน์ต่าง ๆ ได้มากมาย [1] เช่น การอ้างอิงถึงสิทธิ์ในการทำธุรกรรมทางการเงิน ทางนิติกรรม เอกสารที่มีความสำคัญ เอกสารลับที่ไม่ต้องการลงนามที่ระบุชื่อชัดเจน หรือในด้านการบันเทิง เช่น การลงชื่อลงในภาพวาด การแจกลายเซ็นให้ผู้ที่เคารพ ผู้สนับสนุนของดารานักร้อง นักเขียน ผู้ที่มีชื่อเสียงในวงการใดวงการหนึ่ง จากที่กล่าวมาแสดงให้เห็นว่า ปัจจุบันลายเซ็นมีความสำคัญเป็นอย่างมากสำหรับผู้คนส่วนใหญ่ และในผู้ประกอบการอาชีพบางสายงานก็มีความจำเป็น ที่จะต้องมีลายเซ็นที่หลากหลายไว้สำหรับใช้ในแต่ละวาระโอกาสที่แตกต่างกัน ดังนั้น การสร้าง หรือการออกแบบลายเซ็นจึงมีความจำเป็นสำหรับผู้ที่เริ่มฝึกสร้างลายเซ็น และผู้ที่ต้องการสร้างลายเซ็นเพิ่มแต่ไม่มีแนวทางในการสร้างลายเซ็นใหม่ที่มีรูปแบบแตกต่างไปจากเดิม

ในปัจจุบัน พบว่า มีตัวช่วยในการสร้างลายเซ็นหลายรูปแบบ เช่น วิดีโอสอนการสร้างลายเซ็น ผู้ช่วยในการออกแบบลายเซ็นให้สอดคล้องกับความเชื่อที่ส่งผลต่อสภาพจิตใจและความมั่นใจของผู้ใช้ลายเซ็น (อาทิ <https://askastrologer.com/articles/signature-for-luck/>) การสร้างลายเซ็นออนไลน์โดยการแปลงภาพลายเซ็นที่เขียนด้วยการใช้โปรแกรมคอมพิวเตอร์ (อาทิ <https://signaturely.com/online-signature/>) ในส่วนของการสร้างลายเซ็นด้วยปัญญาประดิษฐ์ (Artificial Intelligence : AI) นั้น นิยมทำในรูปแบบของระบบสร้างลายมือเขียน (Handwriting generator) โดยใช้ทั้งการเรียนรู้ของเครื่องจักร (Machine Learning : ML) และการเรียนรู้เชิงลึก (Deep Learning : DL) ในส่วนของการเรียนรู้เชิงลึกที่ใช้โครงข่ายประสาทเทียมนั้น ก็มีการใช้ตัวแบบที่หลากหลาย อาทิ เครือข่ายประสาทกำเริบหรือเครือข่ายประสาทเทียมที่เกิดซ้ำ (Recurrent Neural Network: RNN) และเครือข่ายขัดแย้งกำเนิด (Generative Adversarial Networks: GAN)

ในงานวิจัยชิ้นนี้คณะผู้วิจัยมีความสนใจในการทดลองและพัฒนาตัวแบบสำหรับสร้างลายเซ็นโดยใช้ ScrabbleGAN [2] เป็นหลัก ผลลัพธ์ที่ได้จะถูกนำมาเปรียบเทียบกับผลลัพธ์จากตัวแบบ LSTM [3] และ Transformer [4] จุดเด่นของงานวิจัยนี้คือการนำตัวแบบที่ใช้ในงานสร้างและตรวจสอบลายมือเขียนโดยเฉพาะ มาประยุกต์ใช้เพื่อให้ตัวแบบมีความสามารถที่ไม่จำกัดอยู่แค่เพียงการแปลงภาพลายเซ็นจากลายมือ

เขียนให้กลายเป็นลายเส้นดิจิทัล แต่รวมถึงมีความสามารถในการปรับแต่งลายเส้นดังกล่าวให้ดูแตกต่าง หรือมีรูปแบบที่แปลกใหม่ซับซ้อนขึ้นกว่าลายมือเขียนตั้งต้นได้

Related Works

งานวิจัยนี้เกี่ยวข้องกับตัวแบบสำหรับสร้างลายเส้นจากชื่อ โดยรับอินพุตเป็นข้อความตัวอักษรแล้วแสดงผลออกมาเป็นรูปภาพของลายมือที่เขียนข้อความตัวอักษรดังกล่าว ทั้งนี้ ในบริบทของการสร้างลายเส้นลายมือเขียนที่สร้างขึ้นมักจะถูกสร้างให้แตกต่างจากต้นฉบับเล็กน้อยแต่ยังคงเค้าโครงของลายมือนั้นเอาไว้ได้ ตัวแบบการเรียนรู้เชิงลึกที่นิยมใช้กันในการสร้างลายเส้นนั้นมีหลายตัว ในงานวิจัย Luhman [5] เลือกที่จะใช้เครือข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ชื่อ MobileNetV2 ในการเรียนรู้ข้อมูลสไตร์การเขียนจากภาพลายมือชื่อก่อน จากนั้นจึงใช้แบบจำลองแบบแพร่ (Diffusion model) และเครือข่ายประสาทเทียมแบบป้อนไปข้างหน้า (Feedforward neural network: FNN) เพื่อสร้างลายเส้นผลลัพธ์ที่เป็นลายมือชื่อออกมา ทั้งนี้ตัวแบบของ [5] สามารถสร้างลายมือเขียนที่มีความเชื่อมโยงเชิงเส้นของรูปการเขียนจากผู้เขียนสองคนที่แตกต่างกันได้ ลายมือเขียนที่ได้มีความสมจริงในเรื่องการลงน้ำหนักลายเส้นตัวอักษรใกล้เคียงกับลายมือเขียนจริงของมนุษย์

แม้ตัวแบบประเภท CNN จะสามารถถูกนำมาใช้เพื่อสร้างภาพของลายเส้นหรือภาพของลายมือเขียนของมนุษย์ได้ แต่หากพิจารณาถึงลักษณะเด่นของของลายเส้นซึ่งต้องการความสมจริงในแง่ของความต่อเนื่องเชื่อมโยงของลายเส้นแล้ว ตัวแบบประเภทที่มีความสามารถในการเรียนรู้ข้อมูลต่อเนื่อง (อาทิ RNN, LSTM, Transformer) จึงเป็นตัวเลือกประเภทซึ่งมักถูกนำมาใช้งานในบริบทนี้ ตัวอย่างเช่นงานวิจัยของ Grave [3] ใช้ตัวแบบ LSTM ในการสร้างลายมือเขียนที่เหมือนจริงได้หลายรูปแบบ จุดเด่นของงานวิจัยนี้คือเทคนิคที่ใช้มีความซับซ้อนไม่มาก และในงานวิจัยของ Bhunia และคณะ [4] ก็ได้นำเสนอตัวแบบชื่อ Handwriting Transformer (HWT) ซึ่งภายในเป็นโครงสร้างแบบ Encoder-Decoder ที่ใช้เทคนิค Attention ผลจากการปรับแต่งโครงสร้างสถาปัตยกรรมของ [4] ทำให้ลายมือเขียนที่สร้างออกมามีสไตร์ที่หลากหลาย ครอบคลุมถึงลายมือจากการเขียนพู่กันที่มีความซับซ้อนและเขียนได้ยาก จุดเด่นของงานวิจัย [4] คือสามารถสร้างลายมือเขียนได้หลายรูปแบบ ทั้งแบบที่ซับซ้อนและไม่ซับซ้อน

ในกรณีที่ตัวอักษรแต่ละตัวมีลักษณะ Disconnected กล่าวคือ ไม่มีการเชื่อมต่อระหว่างอักขระภายในคำเดียวกัน ลักษณะดังกล่าวจะเพิ่มความซับซ้อนให้กับระบบสร้างลายเส้น เนื่องจากต้องมีการออกแบบวิธีการเชื่อมโยงเส้น (Stroke) ลายเส้นใหม่ให้ต่อเนื่องในตัวอักษรเดียวกันและแยกจากกันในตัวอักษรคนละตัว รวมถึงต้องกำหนดรูปแบบใหม่ในการเชื่อมโยงตัวอักษรเข้าด้วยกันเพื่อให้ลายเส้นดูสมจริงในภาษาต่าง ๆ งานวิจัยที่ทำการสร้างลายเส้นในลักษณะนี้ ได้แก่ งานวิจัยของ Tan และคณะ [6] ที่สร้างลายมือเขียนภาษาจีนโดยการรับอินพุตที่เป็นการเขียนจากปากกา และใช้ตัวแบบ FNN ร่วมกับ Multipoint regression และงานวิจัยของ Mustapha และคณะ [7] ที่สร้างลายมือเขียนภาษาอาหรับผ่าน Conditional deep convolutional generative adversarial networks (CDCGAN) โดยใช้คลาสของตัวอักษรอาหรับ 28 รูปแบบเป็นเงื่อนไข

(Condition) ที่ใช้ควบคุมตัวแบบ GAN ซึ่งภายในประกอบด้วย 2 ส่วน ส่วนแรกคือ Generator (G) ซึ่งใช้เวกเตอร์ที่เข้ารหัสแบบ One-hot ขนาด 28 มิติ และเวกเตอร์ Gaussian noise ขนาด 100 มิติเป็นข้อมูลอินพุตนำเข้า ส่วนที่สองคือ Discriminator (D) ที่ใช้ Softmax classifier ในการพิจารณาแยกแยะลายมือชื่อที่ได้รับเป็นอินพุต

หากพิจารณาลึกลงในส่วนของงานวิจัยที่ใช้ GAN ในการสร้างลายเซ็นหรือสร้างลายมือเขียน งานวิจัยของ Ji และ Chen [8] เสนอ Handwriting Generative Adversarial Network Framework (HWGANs) สำหรับสังเคราะห์ลายมือเขียน โดยตัวแบบนี้เป็น GAN ที่มี Discriminator ซึ่งใช้ตัวแบบ CNN-LSTM ในการสกัดคุณลักษณะเด่น (Feature extraction) ฝานความสามารถในการวิเคราะห์ภาพของ CNN และความสามารถในการวิเคราะห์ความต่อเนื่องของข้อมูลของ LSTM เข้าด้วยกัน ผลลัพธ์ลายมือเขียนที่สร้างจาก HWGANs ถูกเปรียบเทียบกับผลลัพธ์จาก Alex's Generator ซึ่งเป็นตัวแบบที่ใช้เพียง RNN สำหรับสร้างลายเซ็นเพียงอย่างเดียว (กล่าวคือ Alex's Generator มีเฉพาะส่วนที่ทำหน้าที่ Generator เท่านั้น ไม่มีส่วนของตัวแบบที่ทำหน้าที่ Discriminator เหมือน HWGANs) ผลการทดลองแสดงให้เห็นว่า HWGANs สามารถสร้างลายมือเขียนที่มีความคมชัด ดูเป็นธรรมชาติ และเรียบร้อยกว่า นอกจากนี้ ในบางกรณียังให้ผลลัพธ์ที่มีสไตล์ที่หลากหลายมากกว่า Alex's Generator ด้วย นอกจากนี้งานวิจัยของ Fogel และคณะ [2] ชื่อ ScrabbleGAN ได้ต่อยอดตัวแบบ GAN ในบริบทของการสร้างลายมือเขียนให้ซับซ้อนยิ่งขึ้น โดยใช้เทคนิคการเรียนรู้แบบกึ่งมีผู้สอน (Semi-supervised learning) และ Attention decoder ในการสังเคราะห์ลายมือเขียนภาษาอังกฤษทำให้สามารถปรับแต่งผลลัพธ์ได้มากขึ้นกว่าเทคนิคในอดีต จุดเด่นของ ScrabbleGAN คือลายมือเขียนที่ออกมามีความหลากหลายในคุณลักษณะด้านต่าง ๆ เช่น การวัดตัวของตัวอักษร ความหนาบางของเส้นตัวอักษรการเชื่อมเส้นระหว่างตัวอักษร

จากการศึกษางานวิจัยที่เกี่ยวข้อง ในงานวิจัยชิ้นนี้คณะผู้วิจัยจึงเลือกตัวแบบ ScrabbleGAN [2] เป็นตัวแบบหลักในการสร้างลายเซ็น เนื่องจากเป็นตัวแบบที่ให้ผลลัพธ์ลายมือเขียนได้หลากหลายรูปแบบ สามารถปรับแต่งให้มีความเหมาะสมและใกล้เคียงกับลายเซ็นที่เป็นลายมือต้นฉบับ และเพื่อให้เห็นประสิทธิภาพในเชิงเปรียบเทียบด้วย คณะผู้วิจัยจึงเลือกตัวแบบ LSTM [3] และ Transformer [4] มาใช้เป็นตัวแบบสำหรับเปรียบเทียบผลเพิ่มเติมด้วย

Research Methodology

ในงานวิจัยชิ้นนี้ คณะผู้วิจัยได้ทดลองตัวแบบการเรียนรู้เชิงลึกหลักสำหรับสร้างภาพลายเซ็นคือ ScrabbleGAN [2] และทำการทดลองเชิงเปรียบเทียบกับตัวแบบอีก 2 ตัว รวมเป็นตัวแบบการเรียนรู้เชิงลึกทั้งหมด 3 ตัวในงานวิจัยชิ้นนี้ หลักการทำงานของตัวแบบทั้งสาม คือ ตัวแบบรับอินพุตข้อความซึ่งเป็นชื่อซึ่งอยู่ในรูปแบบของอักขระ (Text หรือ String) และสร้างเอาต์พุตคือรูปภาพลายเซ็นของอินพุตข้อความชื่อดังกล่าว

กระบวนการเตรียมชุดข้อมูล

งานวิจัยนี้ใช้ชุดข้อมูลรูปภาพของ IAM Handwriting [9] ในคลังข้อมูล Lancaster-Oslo/Bergen Corpus ซึ่งเป็นชุดข้อมูลรูปภาพลายมือเขียนภาษาอังกฤษจำนวน 13,353 ภาพ จากลายมือผู้เขียนทั้งหมด 657 คน แบ่งได้เป็นลายมือเขียนทั้งหมด 1,539 หน้า จำนวนคำทั้งหมด 115,320 คำ รูปภาพในชุดข้อมูลนี้ ถูกคณะผู้วิจัยสุ่มขึ้นมาเป็นจำนวนร้อยละ 10 (ประมาณ 1,335 ภาพ) เพื่อนำมาใช้ในการทดลองฝึกสอนตัวแบบสร้างลายเซ็น

ตัวแบบหลัก: ตัวแบบ ScrabbleGAN

เนื่องจาก ScrabbleGAN [2] เป็นตัวแบบประเภท GAN ดังนั้นโครงสร้างภายในจึงประกอบด้วยตัวแบบย่อย 2 ส่วน ได้แก่ ตัวแบบส่วน Generator (G) สำหรับสร้างข้อมูล และตัวแบบส่วน Discriminator (D) สำหรับตรวจสอบข้อมูล ทั้งนี้ตัวแบบทั้งสองส่วนถูกออกแบบให้แข่งขันกัน โดย G พยายามสร้างภาพปลอมให้เหมือนจริงที่สุด ในขณะที่ D พยายามแยกแยะระหว่างภาพจริงและภาพปลอมที่สร้างโดย G ออกจากกันให้ได้มากที่สุด อย่างไรก็ตาม สถาปัตยกรรมของ ScrabbleGAN ดังแสดงใน Figure 1 นั้น เป็นการผสมผสานระหว่างสถาปัตยกรรมของ BigGAN [10] และ Convolutional RNN (CRNN) โดย Generator ทำหน้าที่สร้างภาพลายเซ็นขึ้นจากเวกเตอร์ค่ารบกวน (Noise vector) ภาพลายเซ็นดังกล่าวประกอบขึ้นจากตัวอักษรแต่ละตัวในอินพุตข้อความ ผลลัพธ์ภาพที่สร้างจาก Generator นอกจากจะถูกส่งต่อไปให้ Discriminator ตามปกติของตัวแบบ GAN แล้ว ยังมีการส่งภาพผลลัพธ์ดังกล่าวไปยังส่วนใหม่ที่เรียกว่า Recognizer ด้วย

แนวคิดของ ScrabbleGAN คือการใช้ Discriminator เพื่อโน้มนำให้ Generator เรียนรู้ที่จะสร้างภาพลายเซ็นที่ดูสมจริงมากที่สุด (Real-looking image) ผ่านการใช้ฟังก์ชันสูญเสียแบบ Adversarial (Adversarial loss) ตามหลักการปกติของตัวแบบ GAN อย่างไรก็ตาม เนื่องจาก Discriminator ถูกฝึกสอนมาให้มองภาพทั้งภาพและตัดสินใจว่าภาพนั้นเป็นภาพจริงหรือปลอม ตัว Discriminator จึงขาดซึ่งความสามารถในการพิจารณาตัวอักษรเป็นรายตัวและไม่สามารถตัดสินใจว่าภาพ ๆ หนึ่งนั้นประกอบด้วยตัวอักษรที่ถูกต้องตามลายเซ็นที่ผู้ใช้ต้องการหรือไม่ เพื่อเติมเต็มความสามารถในส่วนนี้ ScrabbleGAN จึงเพิ่มตัวแบบส่วนที่เรียกว่า Recognizer ขึ้นมาสำหรับทำหน้าที่รู้จำ (Recognize) ว่าภาพที่ Generator สร้างมานั้น ประกอบด้วยลำดับของตัวอักษรที่อ่านรู้เรื่อง (Readable) และถูกต้องตามอินพุตที่ผู้ใช้กำหนดมาหรือไม่ โดยส่วนของ Recognizer จะถูกฝึกสอนด้วยฟังก์ชันสูญเสียที่เรียกว่า Connectionist temporal classification (CTC) ทั้งนี้ เมื่อใช้ฟังก์ชันสูญเสียทั้งสองส่วนร่วมกันในการฝึกสอน ก็จะช่วยโน้มนำให้ Generator สามารถสร้างภาพที่ทั้งสวยงามสมจริงและอ่านรู้เรื่องถูกต้องตามตัวอักษรลายเซ็นที่ต้องการได้

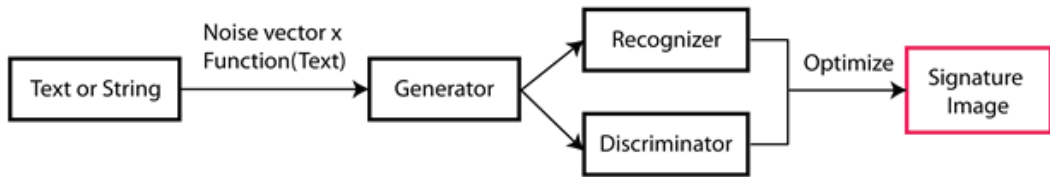


Figure 1 Architecture of ScrabbleGAN

ตัวแบบเปรียบเทียบตัวที่ 1: ตัวแบบ LSTM

LSTM ถือเป็นเครือข่ายประเภท RNN แบบหนึ่ง que เพิ่มเติมหน่วยความจำ (Memory cell) และกลไกประตู (Gate mechanism) ภายในทำให้สามารถจัดการกับข้อมูลต่อเนื่องที่มีความยาวได้ดีขึ้นกว่า RNN แบบดั้งเดิม ในงานวิจัยนี้คณะผู้วิจัยเลือกตัวแบบ LSTM จากงานวิจัย Graves [3] ที่เสนอการนำ LSTM มาต่อยอดใช้สำหรับสร้างลายเซ็น โดยตัวแบบ LSTM ของ Graves [3] สามารถใช้ได้ทั้งกับข้อมูลฝึกสอนที่เป็นข้อมูลการเขียนปากกาแบบเรียลไทม์ และข้อมูลตัวอักษรข้อความธรรมดาที่ไม่มีเรื่องของเวลาเกี่ยวข้องกับดังแสดงใน Figure 2 (ในบริบทของอินพุตที่เป็นข้อความตัวอักษร ซึ่งไม่มีเรื่องของเวลาในการจรดปากกาเพื่อเขียนลายเซ็นมาเกี่ยวข้อง) จะเห็นว่าจุดเด่นของตัวแบบนี้ที่ต่างจากตัวแบบ LSTM ปกติ คือ การมี Skip connection ที่เชื่อมต่อเลเยอร์อินพุตไปถึงเลเยอร์ซ่อน (Hidden layer) ทุกเลเยอร์โดยตรง ลักษณะนี้จะช่วยบรรเทาปัญหา Gradient ที่ไม่เสถียรในระหว่างการฝึกสอนตัวแบบได้ อีกหนึ่งจุดเด่นของตัวแบบ LSTM จาก Graves [3] คือการมีเลเยอร์หน้าต่าง (Window layer) ซึ่งทำหน้าที่รวมผลลัพธ์การสร้างตัวอักษรแต่ละตัวเข้าด้วยกันให้กลายเป็นลายเซ็นผลลัพธ์เพียงอันเดียว อ้างอิงจาก Graves [3] ตัวแบบนี้ถูกฝึกสอนด้วย Root Mean Square Propagation (RMSProp) และในการใช้งานเพื่อสร้างลายเซ็นก็ไม่จำเป็นต้องมีการเตรียมข้อมูล (Data preprocessing) เพิ่มเติม แตกต่างจากงานวิจัยอื่นที่มักต้องมีการเตรียมข้อมูลหลายขั้นตอน

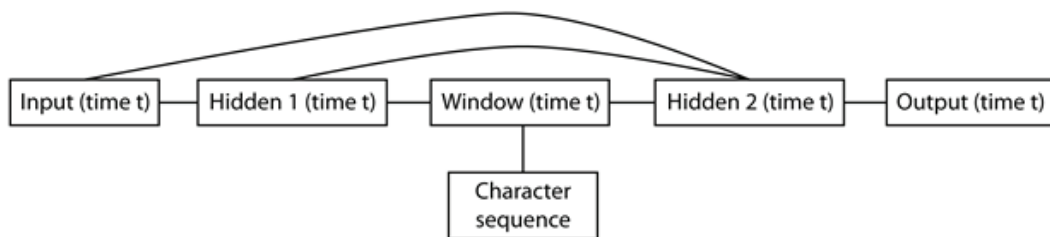


Figure 2 Architecture of LSTM

ตัวแบบเปรียบเทียบตัวที่ 2: ตัวแบบ Transformer

Transformer [11] เป็นสถาปัตยกรรมตัวแบบที่กำเนิดขึ้นจากข้อจำกัดของสถาปัตยกรรมยุคแรกอย่าง RNN และ LSTM ที่ไม่สามารถจัดการกับข้อความที่มีความยาวมาก ๆ ได้อย่างมีประสิทธิภาพและประสิทธิผล กลไกสำคัญของ Transformer คือ Self-attention ที่ทั้งช่วยเร่งความเร็วในการประมวลผลข้อมูลที่มีความยาวมาก ๆ ได้

และช่วยให้ความสัมพันธ์ระหว่างข้อมูลในคนละช่วงเวลาไม่ตกหล่นหายไปในการประมวลผลด้วย แม้ Transformer ตัวต้นฉบับที่นำเสนอโดย Vaswani และคณะ [11] จะมีข้อจำกัดในเรื่องของปริมาณหน่วยประมวลผลและปริมาณเวลาประมวลผลที่ต้องใช้ถึง $O(N^2)$ เมื่อ N คือความยาวของอินพุต แต่ก็กล่าวได้ว่า Transformer [11] คือสถาปัตยกรรมที่ปฏิวัติวงการการประมวลผลข้อความ (Natural Language Processing: NLP) และถูกใช้งานอย่างต่อเนื่องจนถึงยุคปัจจุบันซึ่งเป็นยุคของปัญญาประดิษฐ์แบบรู้สร้าง (Generative AI) และ ChatGPT สำหรับการสร้างภาพลายเซ็นในงานวิจัยนี้ คณะผู้วิจัยเลือกตัวแบบ Handwriting Transformer [4] ซึ่งเป็น Transformer ที่ถูกปรับแต่งให้เหมาะสมกับการสร้างลายเซ็นโดยเฉพาะ ทั้งนี้คณะผู้วิจัยไม่ได้ทำการฝึกสอนใด ๆ เพิ่มเติมกับตัวแบบนี้ แต่เลือกที่จะใช้ตัวแบบที่ถูกฝึกสอน (Pre-trained) มาเรียบร้อยแล้วโดย Bhunia และคณะ [4] ใน <https://github.com/ankanbhunia/Handwriting-Transformers> (ดำเนินการผ่านไลบรารี PyTorch 1.4)

กระบวนการปรับค่าพารามิเตอร์ของตัวแบบ

การปรับค่าพารามิเตอร์ (Parameters tuning) ในงานวิจัยนี้ ผู้จัดทำได้ปรับค่าที่คาดว่าจะดีที่สุดของพารามิเตอร์ในตัวแบบนั้น ๆ โดยเป็นการทดลองปรับค่าโดยสุ่มซึ่งอ้างอิงค่าเริ่มต้นจากค่าที่เคยถูกทดลองไว้ในงานวิจัยอื่นในอดีต สำหรับตัวแบบ ScrabbleGAN ข้อมูลการปรับแต่งพารามิเตอร์ของคณะผู้วิจัยแสดงดัง Table 1 ในส่วนของตัวแบบ LSTM (ดำเนินการผ่านไลบรารี TensorFlow 1) คณะผู้วิจัยทำการปรับพารามิเตอร์ดังนี้ sequence length = 256, batch size = 64, epochs = 30, window mixtures = 10, output mixtures = 20, LSTM layers = 3 layers, และ 400 unit per layer

Table 1 Hyperparameter tuning values of the ScrabbleGAN modelre of LSTM

Loss and optimizer (Adam optimizer)		Shared spec (G and D)	
G, D, R Optimizer	0.001	epochs	5,10,15,20
Learning rate		batch_size	64
Optimizer beta 1	0.0	latent_dim	128
Optimizer beta 2	0.999	embed_y	(32, 8192)
Loss function	Not saturating	num_gen	16
disc_iters	1	kernel_reg	@spectral_norm
Input			
input dimension	(32, 160, 1)	number classes	52
buffer size	80377	bucket size	10
character vector	'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ'		

Experimental Results

ผลการสร้างภาพลายเส้นของตัวแบบทั้ง 3 ของงานวิจัยนี้แสดงดัง Table 2 จากผลการทดลองพบว่า ภาพลายเส้นที่ได้จากตัวแบบ ScrabbleGAN นั้นมีความสมจริงระดับหนึ่ง แต่มีการกำจัดพื้นหลังที่ไม่สมบูรณ์ เมื่อเทียบกับตัวแบบอื่นอีกสองตัว ซึ่งอาจเกิดจากโครงสร้างของ ScrabbleGAN ที่มีความซับซ้อนมากกว่า ตัวอื่นจึงอาจต้องการจำนวนข้อมูลฝึกสอนมากเพื่อให้แบบจำลองสามารถเรียนรู้ได้ถึงจุดที่สามารถสร้างลายเส้นได้อย่างสะอาดและสมบูรณ์ ในส่วนของตัวแบบ LSTM นั้นให้ลายเส้นที่มีเส้นราบเรียบตลอดทุกช่วงของตัวอักษร ซึ่งแม้จะทำให้ลายเส้นดูสะอาดตาและคมชัดเหมาะสำหรับการเริ่มต้นออกแบบลายเส้น แต่ลักษณะของเส้นที่ราบเรียบเกินไปทำให้ลายเส้นดูเป็นสังเคราะห์มากกว่าลายมือจริง ในส่วนของตัวแบบ Transformer นั้นให้ลายเส้นที่คล้ายกับลายมือเขียนแบบที่ไม่ใช่ลายมือตัวบรรจง โดยมีความแตกต่างของน้ำหนักเส้นในช่วงต่าง ๆ ของตัวอักษร ซึ่งอาจเป็นผลมาจากกลไก Self-Attention ที่ให้ความสำคัญกับลักษณะเฉพาะของแต่ละส่วนของลายเส้น ทำให้เกิดความหลากหลายในน้ำหนักของเส้นมากกว่าตัวแบบ LSTM ที่เน้นความต่อเนื่องของข้อมูลเป็นหลัก

โดยสรุป ผลลัพธ์จากตัวแบบหลักของงานวิจัยนี้ ซึ่งก็คือ ScrabbleGAN ที่ได้รับการฝึกสอนเพิ่มเติมด้วยชุดข้อมูลของงานวิจัยนี้ ยังคงให้ผลลัพธ์ที่ไม่ดีนักในปัจจุบัน โดยเฉพาะปัญหาพื้นหลังของลายเส้นที่เต็มไปด้วยสัญญาณรบกวนและขาดความสะอาดตา ปัญหาดังกล่าวคณะผู้วิจัยคาดว่าจะเกิดจากข้อจำกัดด้านทรัพยากรในการฝึกสอน ซึ่งส่งผลให้คณะผู้วิจัยไม่สามารถใช้ข้อมูลทั้งหมดจากชุดข้อมูล IAM Handwriting ได้อย่างเต็มที่ อีกทั้งยังไม่สามารถเพิ่มชุดข้อมูลลายมือและลายเส้นจากแหล่งอื่น ๆ เพื่อเพิ่มความหลากหลายหรือปรับแต่งพารามิเตอร์ที่ต้องใช้ทรัพยากรประมวลผลสูงได้ หากมีเวลาและทรัพยากรเพียงพอสำหรับการฝึกสอนและปรับแต่งตัวแบบให้รองรับข้อมูลที่หลากหลายและมีปริมาณมากขึ้น โดยเฉพาะการรวมชุดข้อมูลที่ครอบคลุมลายเส้นและลายมือที่แตกต่างกัน คณะผู้วิจัยคาดว่าผลลัพธ์จาก ScrabbleGAN จะมีคุณภาพที่ดีขึ้นได้อย่างมีนัยสำคัญ

เพื่อพัฒนาระบบสร้างลายเส้นให้ดียิ่งขึ้นในอนาคต คณะผู้วิจัยมีข้อเสนอแนะ 4 ประการดังต่อไปนี้

1. ยกระดับคุณภาพและความหลากหลายของลายเส้น

ควรมีการรวบรวมชุดข้อมูลลายเส้นที่ออกแบบโดยผู้เชี่ยวชาญด้านการออกแบบลายเส้น ตลอดจนบุคคลที่มีลายมือเป็นเอกลักษณ์และเคยใช้ลายเส้นหลากหลายรูปแบบ ข้อมูลเหล่านี้จะช่วยเพิ่มความหลากหลายและมิติทางศิลปะให้กับลายเส้นที่ตัวแบบสร้างขึ้น

2. สำรวจเทคนิคขั้นสูงเพื่อเพิ่มประสิทธิภาพการสร้างลายเส้น

ควรศึกษาและทดลองเทคนิคการสร้างภาพและการสังเคราะห์ลายมือเพิ่มเติม เช่น ทดลองใช้ GAN รูปแบบอื่นที่แตกต่างจาก ScribbleGAN, ใช้ Diffusion Models แทน GAN ในการสร้างภาพ หรือประยุกต์ใช้ Vision-Language Models (VLM) เพื่อให้ระบบสามารถสร้างภาพลายเส้นตามคำสั่งข้อความ (Prompt) ได้อย่างแม่นยำ

3. เพิ่มความน่าเชื่อถือด้วยการวัดผลเชิงปริมาณ

ณ ปัจจุบัน ผลลัพธ์ของงานวิจัยยังเป็นการวิเคราะห์เชิงคุณภาพที่มีจำนวนตัวอย่างจำกัด ส่งผลให้การเปรียบเทียบกับงานวิจัยอื่นทำได้ยาก ในอนาคต ควรมีการทดลองวัดผลเชิงปริมาณโดยใช้ตัวชี้วัดมาตรฐาน เช่น Frechet Inception Distance (FID), Geometric-Score Metrics, และ GQM Scores เพื่อประเมินประสิทธิภาพของตัวแบบอย่างเป็นระบบ

4. ประเมินผลผ่านมนุษย์เพื่อความสมจริงและการใช้งานที่เหมาะสม

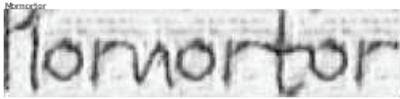
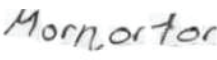
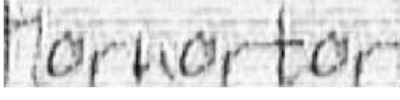
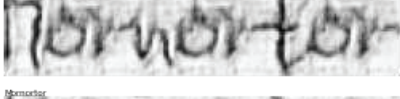
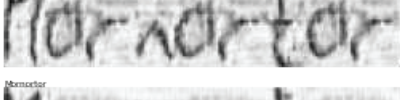
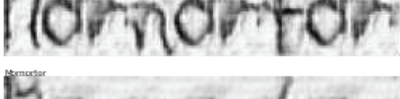
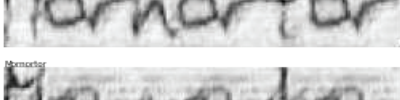
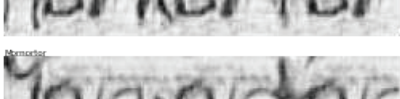
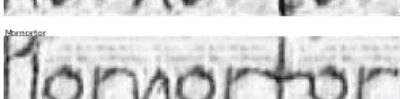
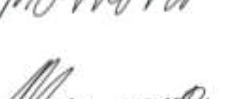
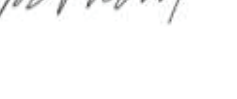
ควรดำเนินการทดลองเชิงคุณภาพโดยนำลายเซ็นที่สร้างจากตัวแบบแต่ละตัวไปให้ผู้เข้าร่วมการทดลองที่เป็นมนุษย์ประเมิน เพื่อวิเคราะห์ในแง่มุมต่าง ๆ ได้แก่ ความสมจริงของลายเซ็น, ความชัดเจนในการอ่าน, ความหลากหลายของรูปแบบ, และ ระดับความคิดสร้างสรรค์ในการออกแบบลายเซ็น ทั้งนี้ จะช่วยให้นักวิจัยสามารถปรับปรุงตัวแบบให้ตอบโจทย์การใช้งานได้อย่างแท้จริง

Conclusion

งานวิจัยนี้นำเสนอการใช้เทคนิคการเรียนรู้เชิงลึกสำหรับสังเคราะห์รูปภาพลายเซ็นที่เป็นลายมือมนุษย์ โดยใช้เพียงอินพุตข้อความของชื่อที่ต้องการสร้างลายเซ็น ตัวแบบ 3 ตัวที่ถูกทดลองเปรียบเทียบกันในงานนี้ได้แก่ ตัวแบบ ScrabbleGAN ซึ่งเป็นตัวแทนของสถาปัตยกรรม GAN ตัวแบบ LSTM ซึ่งเป็นตัวแทนของสถาปัตยกรรม RNN และตัวแบบ Transformer ซึ่งเป็นตัวแทนของสถาปัตยกรรม Transformer ผลการศึกษพบว่าตัวแบบการเรียนรู้เชิงลึกทั้งสาม ต่างก็สามารถสร้างลายเซ็นที่เป็นลายมือเขียนตามอินพุตของชื่อภาษาอังกฤษใด ๆ ที่ได้รับมาได้ แต่เนื่องจากทรัพยากรการฝึกสอน ScrabbleGAN ที่จำกัดของคณะผู้วิจัยเอง ทำให้ลายเซ็นที่ได้จาก ScrabbleGAN ยังมีข้อจำกัดด้านคุณภาพของภาพที่เห็นได้ชัดเมื่อเทียบกับอีก 2 ตัวแบบ เช่น การมีสัญญาณรบกวนในพื้นหลัง

สำหรับการพัฒนาต่อยอดเพื่อปรับปรุงคุณภาพของภาพลายเซ็นให้ดียิ่งขึ้นนั้น นอกจากการเลือกใช้ชุดข้อมูลที่มีจำนวนมากขึ้นครอบคลุมลายเซ็นแบบต่าง ๆ และการปรับแต่งพารามิเตอร์ฝึกสอนของแบบจำลอง GAN เพื่อหาค่าที่เหมาะสมที่สุดแล้ว ยังรวมถึงการสำรวจความเป็นไปได้ของสถาปัตยกรรม Generative AI ประเภทอื่น อาทิ สถาปัตยกรรมแบบเปิดสาธารณะในกลุ่ม Stable Diffusion [12, 13] สำหรับสร้างภาพลายเซ็นจากอินพุตที่เป็นสัญญาณรบกวนและข้อความที่ต้องการสร้างลายเซ็น และแบบจำลองในกลุ่ม Vision-Language model [14] ซึ่งเป็นลักษณะของแชทบอตที่มีความสามารถในการทำความเข้าใจในรูปภาพและบางแบบจำลองก็รวมความสามารถในการสร้างรูปภาพได้ด้วย

Table 2 Generated signature images from the three models for the name “Mornortor”

ScrabbleGAN [2]	LSTM [3]	Transformer [4]
		
		
		
		
		
		
		
		
		
		
		

References

1. Diaz, M., Ferrer, M.A., Impedovo, D., Malik, M., Pirlo, G. and Plamondon, R. 2019. A perspective analysis of handwritten signature technology. *ACM Computing Surveys*, 51, 1-39. <https://doi.org/10.1145/3274658>
2. Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S. and Litman, R. 2020. ScrabbleGAN: Semi-supervised varying length handwritten text generation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4324–4333. <https://doi.org/10.1109/CVPR42600.2020.00438>
3. Graves, A. 2014. Generating sequences with recurrent neural networks. arXiv:1308.0850, 43. <https://doi.org/10.48550/arXiv.1308.0850>
4. Bhunia, K., Khan, S., Cholakkal, H., Anwer, R.M., Khan, F.S. and Shah, M. 2021. Handwriting transformers. *2021 IEEE/CVF International Conference on Computer Vision*, 1066-1074. <https://doi.org/10.1109/ICCV48922.2021.00112>
5. Luhman, T. and Luhman, E. 2020. Diffusion models for handwriting generation. arXiv: 2011.06704, 17. <https://doi.org/10.48550/arXiv.2011.06704>
6. Tan, B.R., Yin, F., Wu, Y.C. and Liu, C.L. 2017. Chinese handwriting generation by neural network based style transformation. *Lecture Notes in Computer Science*, 10666, 408–419. https://doi.org/10.1007/978-3-319-71607-7_36.
7. Mustapha, I.B., Hasan, S., Nabus, H. and Shamsuddin, S.M. 2022. Conditional deep convolutional generative adversarial networks for isolated handwritten Arabic character generation. *Arabian Journal for Science and Engineering*, 47, 1309-1320. <https://doi.org/10.1007/s13369-021-05796-0>
8. Ji, B. and Chen, T. 2020. Generative adversarial network for handwritten text. arXiv: 1907.11845, 12. <https://doi.org/10.48550/arXiv.1907.11845>
9. Marti, U.V. and Bunke, H. 2002. The IAM-database: An English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5, 39-46. <https://doi.org/10.1007/s100320200071>
10. Brock, A., Donahue, J. and Simonyan, K. 2019. Large scale GAN training for high fidelity natural image synthesis. *International Conference on Learning Representations*, 6-9 May 2019, New Orleans, LA, USA, 1-35.

11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., Kaiser, L. and Polosukhin, I. 2017. Attention is all you need. *Conference on Neural Information Processing Systems*, 4-9 December 2017, Long Beach, CA, USA, 1-11.
12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18-24 June 2022, New Orleans, Louisiana, 10674-10685. <https://doi.org/10.1109/CVPR52688.2022.01042>.
13. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J. and Rombach, R. 2024. SDXL: Improving latent diffusion models for high-resolution image synthesis. *International Conference on Learning Representations*, 7-11 May 2024, Vienna, Austria, 13.
14. Zhang, J., Huang, J., Jin, S. and Lu, S. 2024. Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 5625-5644. <https://doi.org/10.1109/TPAMI.2024.3369699>