

การเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับภาพดิจิทัลและวิดีโอ โดยใช้เทคนิคการเรียนรู้แบบเชิงลึก

Blurring Inappropriate Behavior Image for Digital Images and Videos Using Deep Learning Techniques

อติชา ไชยพนิชย์¹ ปภาดา โพธิ์¹ และ ชัยพิชิต คำพิมพ์^{1*}
Aticha Chaipanit¹, Papada Pothi¹ and Chaipichit Cumpim^{1*}

Received: 20 March 2023

Revised: 9 June 2023

Accepted: 4 July 2023

บทคัดย่อ

งานวิจัยนี้ได้ทำการพัฒนาระบบการเรียนรู้เชิงลึกสำหรับการเบลอภาพพฤติกรรมที่ไม่เหมาะสมเนื่องจากการเผยแพร่ภาพและคลิปกันอย่างกว้างขวางทำให้กระบวนการในการตัดต่อภาพหรือขั้นตอนในการปิดบังภาพหรือคลิปวิดีโอที่มีพฤติกรรมที่ไม่เหมาะสมทำได้ไม่ครอบคลุม เนื่องจากต้องใช้บุคลากรจำนวนมากในการตรวจสอบภาพและคลิปวิดีโอ ดังนั้นงานวิจัยนี้จึงได้พัฒนาระบบการเบลอภาพและคลิปวิดีโอ ซึ่งมีขั้นตอนการทำงาน 3 ส่วนคือ ส่วนที่ 1 คือ การฝึกสอนระบบการเรียนรู้แบบเชิงลึกโดยใช้โมเดล Mask R-CNN และส่วนที่ 2 คือขั้นตอนของการตรวจสอบพฤติกรรมที่ไม่เหมาะสมที่ปรากฏภายในภาพและคลิป จากนั้นเมื่อรู้ตำแหน่งของพฤติกรรมที่ไม่เหมาะสมภายในภาพและคลิปแล้วในขั้นตอนสุดท้ายจะทำการเบลอส่วนนั้นด้วยตัวกรองแบบเกาส์เซียน ผลการทดลองพบว่ามีความ Accuray Precision Recall และ F1 score ซึ่งคำนวณได้จาก Confusion Matrix เพื่อใช้ในการประเมินประสิทธิภาพของระบบในส่วนของการเบลอภาพพบว่ามีความ 0.82 0.88 0.91 และ 0.90 ตามลำดับ ส่วนของวิดีโอมีความ 0.88 0.93 0.92 และ 0.93 ตามลำดับเช่นกัน

คำสำคัญ: การเบลอภาพ การเบลอพฤติกรรมที่ไม่เหมาะสม Mask R-CNN ตัวกรองแบบเกาส์เซียน

¹ สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ นครปฐม ประเทศไทย 73170

¹ Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Rattanakosin, Nakhon Pathom, Thailand 73170

* ผู้รับผิดชอบประสานงาน (Corresponding author) e-mail: chaipichit.cum@rmutr.ac.th

ABSTRACT

This research has developed a deep learning system for blurring images of inappropriate behavior. Blurring inappropriate images can be challenging for employees, as these pictures and clips are often widely published. Therefore, this paper presents a system that consists of three steps: training deep learning, detecting inappropriate behavior, and blurring the detected regions. The first step involves training the deep learning system using the Mask R-CNN model. In the second step, inappropriate behavior is detected using Mask R-CNN. Finally, in the last step, the detected regions are blurred using a Gaussian filter. The results of the evaluation showed that the system's performance in blurring images was assessed using metrics such as Accuracy, Precision, Recall, and F1 score, calculated from the Confusion Matrix. The metrics for the blurred images were found to be 0.72, 0.8, 0.86, and 0.83, respectively. For videos, the metrics were 0.88, 0.93, 0.92, and 0.93, respectively.

Keywords: Blur image; Blur inappropriate behavior in image; Mask R-CNN; Gaussian Filter

บทนำ

ในปัจจุบันมีการเผยแพร่ข้อมูลภาพหรือคลิปวิดีโอออกสู่สื่อสังคมออนไลน์อย่างแพร่หลาย โดยข้อมูลบางส่วนที่ถูกเผยแพร่อาจมีส่วนของพฤติกรรมที่ไม่เหมาะสม เช่น ภาพเครื่องดื่มแอลกอฮอล์ ภาพการถืออาวุธมีด ปืน ภาพการสูบบุหรี่ สูบบุหรี่ยาเสพติด เป็นต้น ซึ่งการเผยแพร่ข้อมูลเหล่านี้โดยไม่คำนึงถึงพฤติกรรมที่ไม่เหมาะสมที่ปรากฏภายในภาพหรือคลิปวิดีโอ อาจทำให้เด็กหรือเยาวชนสามารถลอกเลียนแบบพฤติกรรมที่ปรากฏภายในสื่อเหล่านี้ได้โดยง่าย มีการกำหนดสิทธิ์การเข้าถึงข้อมูลเหล่านี้ในปัจจุบันแต่ก็ยังไม่สามารถที่จะทำการตัดต่อส่วนของข้อมูลที่ไม่เหมาะสมออกมาได้ ในปัจจุบันได้มีการพัฒนาระบบแบบเรียนรู้เชิงลึกขึ้นมาทำให้สามารถที่จะประยุกต์ระบบเหล่านี้เข้ากับการตัดต่อข้อมูลที่ไม่เหมาะสมภายในภาพหรือคลิปวิดีโอโดยระบบการเรียนรู้แบบเชิงลึกเหล่านี้จะกล่าวถึงในส่วนต่อไป

ระบบการเรียนรู้เชิงลึกมีวิวัฒนาการมาจากโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network; CNN) ได้ถูกนำเสนอครั้งแรกในปี 1980 โดย Yann LeCun นักวิทยาศาสตร์คอมพิวเตอร์ชาวฝรั่งเศส โดยมีสถาปัตยกรรมต้นแบบจาก Neocognitron CNN นำมาใช้ในการแก้ปัญหา Classification ที่เกี่ยวข้องกับภาพ [1] ต่อมาได้มีการปรับปรุง CNN ให้มีประสิทธิภาพในการทำงานสูงขึ้นโดยได้พัฒนามาเป็น Regional Convolutional Neural Network (R-CNN) ซึ่งเป็นงานวิจัยที่ถูกนำเสนอโดย Ross Girshick [2] ในปี 2013 R-CNN เริ่มต้นด้วยการใช้วิธีการที่เรียกว่า Selective Search คือการแบ่งส่วนของภาพออกเป็น 2000 ส่วนเพื่อแยกขอบเขตที่น่าสนใจ (Region of Interest; ROI) ไปประมวลผลด้วย CNN แล้วนำมากำหนดประเภทของวัตถุภายใน ROI จากนั้นมีการปรับปรุง R-CNN ให้มีการทำงานได้เร็วยิ่งขึ้น ได้มีการพัฒนามาเป็น Fast Regional Convolutional Neural Network (Fast R-CNN) ถูกนำเสนอโดย Ross Girshick [3] เพื่อแก้ไขข้อบกพร่องบางประการของ R-CNN เนื่องจากว่าโมเดลที่กล่าวมาทั้ง 3 เป็นโมเดลที่ใช้ในการตรวจสอบวัตถุภายในภาพจึงได้มีการพัฒนาปรับปรุงโมเดลเหล่านี้

ให้สามารถแบ่งแยกแวกวัตถุที่ตรวจสอบได้ภายในภาพออกมาได้ สำหรับโมเดลที่นิยมใช้ในปัจจุบันคือ Mask Regional Convolutional Neural Network (Mask R-CNN) [4] ออกแบบมาเพื่อใช้สำหรับงานประเภท Instance Segmentation ซึ่งงานประเภทนี้จะเป็นการรวม Object detection และ Semantic Segmentation ไว้ด้วยกัน MobileNetV2 เป็นโมเดลที่ถูกออกแบบมาให้เหมาะกับการทำงานแบบเรียลไทม์ มีความเร็วในการประมวลผล และใช้ทรัพยากรในการประมวลผลน้อยสามารถใช้ได้ทั้งงานประเภท Classification, Detection, Embedding และ Segmentation สำหรับอีกโมเดลก็คือ You Only Look Once (YOLO) [5] เป็นโมเดลที่แตกต่างจากพวก R-CNN ในการแยกวัตถุ YOLO จะใช้ Bounding Box กับ Confidence มาใช้ในการแยกวัตถุจากภาพ สำหรับงานวิจัยที่กล่าวผ่านมามีการนำเอาไปประยุกต์ใช้ในงานวิจัยในหลายด้านโดยจะมีการนำเสนองานวิจัยเหล่านั้นในส่วนต่อไป

สำหรับงานวิจัยที่ประยุกต์ใช้ระบบการเรียนรู้แบบเชิงลึกเข้ามามีคือ การอ่านเลขสายรถประจำทาง จากภาพ [6] ซึ่งเป็นงานวิจัยที่นำการเรียนรู้เชิงลึกโมเดล Fast R-CNN มาเพื่อทำนายเลขสายรถประจำทาง เพื่อช่วยเหลือผู้พิการสายตาเลือนลาง โดยใช้เทคนิคการประมวลผลภาพและใช้ข้อมูลจีพีเอสของป้ายรถ ประจำทาง งานวิจัยนี้ได้ทำการทดสอบการอ่านค่าจริง ในงานวิจัยนี้ได้ทำการทดลองอ่านเลขสายรถประจำ ทางได้ถูกต้อง 62 เปอร์เซ็นต์ งานวิจัยต่อมา [7] เป็นการนำ Deep CNN มาจำแนกพรรณไม้เพื่อเปรียบเทียบ Convolutional Neural Network คือ LeNet5, AlexNet, GoogLeNet และ VGGNet โดยใช้ชุดข้อมูล PNE และ 102 Flower ที่เป็นภาพที่อยู่ในสิ่งแวดล้อมธรรมชาติและ Folio เป็นภาพที่ถ่ายในห้องทดลอง พบว่า GoogLeNet มีประสิทธิภาพสูงสุดในการจำแนกพรรณไม้ที่อยู่ในแวดล้อมทางธรรมชาติงานวิจัยต่อมา [8] ได้นำเอากระบวนการเรียนรู้แบบเชิงลึกมาใช้ในการทำนายปริมาณน้ำฝนในประเทศไทย โดยเป็นการทำนาย ล่วงหน้าว่าฝนจะตกในพื้นที่หรือไม่จากนั้นจึงใช้การทำนายปริมาณน้ำฝนด้วย regression model ผลลัพธ์ ของงานวิจัยนี้พบว่ามีความคลาดเคลื่อนต่ำสุดงานวิจัยของณัฐณิชา ยงยิ่ง [9] เป็นงานวิจัยที่นำระบบการ เรียนรู้เชิงลึกมาจำแนกข้อมูลภาพถ่ายว่าเป็น ถนนหนทาง ต้นไม้ อาคาร หรือสิ่งอื่นๆ เพื่อนำปรับปรุงข้อมูล ถนนในโปรแกรม OpenStreetMap ในพื้นที่เขตชนบท ผลการทดลองของงานวิจัยนี้ได้จำแนกข้อมูลถนน และวิเคราะห์ข้อมูลได้อย่างถูกต้องแม่นยำและรวดเร็ว งานวิจัยต่อมาการตรวจหาต้นไม้เป็นโรคโดยอัตโนมัติ ด้วยภาพถ่ายมุมสูงจากโดรน [10] ซึ่งมีการทำงาน 2 ขั้นตอน ขั้นตอนแรกคือขั้นตอนการตรวจสอบต้นไม้ ภายใต้อากาศโดยใช้เทคนิค Mask R-CNN ขั้นตอนสุดท้ายคือการนำเอาภาพของต้นไม้ที่ตรวจสอบได้ไปเปรียบเทียบกับฐานข้อมูล ImageNet ของโมเดล VGG16 เพื่อยืนยันว่าภาพถ่ายต้นไม้ที่ตรวจสอบได้เป็นโรคหรือไม่ งานวิจัยต่อมา [11] เป็นการศึกษาเปรียบเทียบประสิทธิภาพของระบบการเรียนรู้แบบเชิงลึก 5 โมเดล คือ Fast R-CNN_NAS2, Faster R-CNN Inception_ResNetV2 Atrous, SSD ResNet50 FPN, SSD MobileNetV1 FPN และ Faster R-CNN ResNet101 โดยนำทั้ง 5 โมเดลมาใช้ในการจำแนกพื้นผิวถนน ผล การตรวจสอบพบว่าโมเดล SSD ResNet50 FPN สามารถจำแนกพื้นผิวถนนได้ดีที่สุด ต่อมาการตรวจหา COVID-19 โดยการใช้การเรียนรู้เชิงลึกกับภาพซีทีสแกน [12] เป็นการนำการเรียนรู้เชิงลึก Mask R-CNN มา ตรวจจับความผิดปกติของเนื้อเยื่อปอดที่เป็นลายเส้นเฉพาะของไวรัส COVID-19 เพื่อแยกส่วนบริเวณของ เนื้อเยื่อปอดที่ได้รับความเสียหาย งานวิจัยนี้ได้นำเอากระบวนการเรียนรู้แบบเชิงลึกโมเดล Mask R-CNN มาใช้ ในการจำแนกภาพถ่ายรังสีเอ็กซเรย์ของเนื้อเยื่อทรวงอก ผลจากการทดลองของงานวิจัยนี้พบว่าสามารถทำนาย ได้ถูกต้อง 89 เปอร์เซ็นต์แล้วใช้เวลาในการทำนายอยู่ที่ 9.71 วินาที งานวิจัยต่อมาเป็นการประยุกต์ใช้ระบบ การเรียนรู้แบบเชิงลึกสำหรับการจำแนกเมล็ดพันธุ์ข้าวไทย โดยใช้ภาพถ่ายของเมล็ดพันธุ์ข้าวด้วยวิธีการแบบ Mask R-CNN และ Transfer Learning [13] ต่อมาเป็นการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกสำหรับการ

รู้จำโรคขอบใบแห้งและโรคไหม้ของข้าว (2021) [14] งานวิจัยนี้เป็นการประยุกต์ใช้ระบบการเรียนรู้แบบเชิงลึกในการจำแนกโรคขอบใบแห้งและโรคไหม้ของข้าวเพื่อจะได้สามารถหาวิธีป้องกันโรคทั้งสองได้อย่างทันที่ หากสามารถวิเคราะห์หรือจำแนกลักษณะของโรคทั้งสองได้อย่างรวดเร็วและถูกต้องแม่นยำ โดยงานวิจัยนี้ได้ใช้โมเดล YOLOV3 ผลการทดลองจากการนำไปใช้งานพบว่างานวิจัยนี้มีประสิทธิภาพสูงกว่าผลที่ได้จากงานวิจัยที่ผ่านมา งานวิจัยต่อมากล่าวถึงการใช้ระบบการตรวจสอบบุคคลด้วยระบบการเรียนรู้แบบเชิงลึก [15] เป็นการพัฒนาระบบติดตามบุคคลแบบอัตโนมัติโดยใช้เลเซอร์พอยเตอร์ในการชี้ไปยังตัวบุคคลที่เคลื่อนที่ผ่านหน้ากล้องเว็บแคมและงานวิจัยสุดท้าย [16] งานวิจัยนี้เป็นการทดลองเปรียบเทียบระบบการเรียนรู้แบบเชิงลึก 4 โมเดลคือ Lenet-5 VGG16 ResNet-50 และโมเดลที่นำเสนอในงานวิจัยนี้สำหรับทำจำแนกโรคของมะนาวจากภาพถ่ายใบมะนาว จากการทดลองพบว่าโมเดลที่นำเสนอมีค่าความถูกต้องแม่นยำไม่ต่างจากโมเดล ResNet-50 แต่ก็ยังไม่มีความสูงกว่า 2 โมเดลที่เหลือ จากงานวิจัยที่กล่าวผ่านมาเป็นการใช้ระบบการเรียนรู้แบบเชิงลึก ซึ่งงานวิจัยที่นำเสนอในเอกสารนี้ได้นำเอาโมเดล Mask R-CNN มาประยุกต์ใช้ในงานวิจัยสำหรับแนวคิดของงานวิจัยนี้จะได้กล่าวในส่วนต่อไป

สำหรับวิธีการที่นำเสนอในงานวิจัยนี้ได้มีการนำเอาระบบการเรียนรู้แบบเชิงลึกมาใช้ในการตรวจสอบพฤติกรรมหรือท่าทางที่ไม่เหมาะสมของบุคคลภายในภาพ โดยการฝึกสอนระบบให้เรียนรู้ภาพพฤติกรรมที่ไม่เหมาะสมโดยใช้โมเดล Mask R-CNN เมื่อตรวจสอบได้แล้วก็จะทำการเบลอบริเวณที่ถูก Mask ด้วยตัวกรองแบบเกาส์เซียน เนื่องจากภาพที่ถูกตรวจสอบพบด้วยโมเดลระบบการเรียนรู้แบบเชิงลึกโดยทั่วไป บริเวณที่ถูก Detect จะถูกรอบด้วยสี่เหลี่ยมผืนผ้าซึ่งจะรวมเอาบริเวณของภาพที่ไม่ได้เกี่ยวข้องกับการแสดงพฤติกรรมที่ไม่เหมาะสมรวมเข้าไปด้วย ดังนั้นเมื่อทำการเบลอบริเวณของภาพที่เกิดการเบลอก็มีขนาดใหญ่เกินไป จากที่กล่าวผ่านมามีได้นำเอาโมเดล Mask R-CNN มาใช้เพราะโมเดลนี้เมื่อตรวจสอบส่วนที่ไม่เหมาะสมภายในภาพก็จะทำการระบุบริเวณของที่ไม่เหมาะสมตามพื้นที่ที่ปรากฏขึ้นจริงภายในภาพ ส่วนการเบลอบริเวณที่ได้เลือกใช้ตัวกรองแบบเกาส์เซียน เนื่องจากตัวกรองนี้จะมียุคที่เมื่อนำไปกรองภาพจะทำให้ภาพมีลักษณะที่เบลอซึ่งระดับของการเบลอขึ้นอยู่กับค่าความแปรปรวนที่กำหนดให้กับตัวกรอง ดังนั้นในงานวิจัยนี้จึงได้นำเอาตัวกรองนี้มาใช้ในการเบลอบางส่วนของภาพที่ต้องการปิดบังไว้ สำหรับรายละเอียดของวิธีการนี้จะได้กล่าวถึงในส่วนต่อไป

วัสดุอุปกรณ์ที่ใช้ในงานวิจัยนี้ประกอบด้วย ชุดข้อมูลภาพที่มีส่วนของพฤติกรรมที่ไม่เหมาะสมจำนวน 7,550 ภาพ โดยค่าต่ำสุดของความกว้างในแนวนอนหรือความสูงของภาพควรมีค่าอยู่ระหว่าง 800 ถึง 1024 พิกเซล ซึ่งเป็นภาพสี RGB และภาพขาวดำ ซึ่งแบ่งเป็นภาพการสูบบุหรี่ 1480 ภาพ ภาพการสูบบุหรี่ไฟฟ้า 980 ภาพ ภาพการสูบบุหรี่ 800 ภาพ ภาพการถือปืน 1350 ภาพ ภาพการถือมีด 1455 ภาพ และการถือเครื่องดื่มแอลกอฮอล์ 1485 ภาพ ที่มาของภาพได้มาจากการสืบค้นจาก Search Engine โดยเลือกเอาเฉพาะภาพที่อยู่ในข้อกำหนดสัญญาอนุญาตครีเอทีฟคอมมอนส์ (Creative Commons License) โดยจะกำหนดส่วนของพฤติกรรมที่ไม่เหมาะสมภายในภาพและกำหนดคลาสให้กับชุดข้อมูลภาพทั้งหมด ผ่านทางเว็บไซต์ www.makesense.ai โดยในระบบนี้จะมีเพียงคลาสเดียว คือ คลาสของพฤติกรรมที่ไม่เหมาะสม (Inappropriate Class) แต่ภายในคลาสได้ใช้ภาพที่ประกอบไปด้วย การสูบบุหรี่ การสูบบุหรี่ไฟฟ้า การถืออาวุธปืน การถือมีด การสูบบุหรี่ และการถือเครื่องดื่มแอลกอฮอล์ ซึ่งจะถูกกำหนดบริเวณที่มีการปรากฏของพฤติกรรมภายในภาพ จากนั้นจะถูกนำไปใช้ในการฝึกสอนระบบการเรียนรู้แบบเชิงลึก Mask R-CNN และได้ใช้ตัวกรองแบบเกาส์เซียนในการเบลอบริเวณของพฤติกรรมที่ไม่เหมาะสม

สำหรับซอฟต์แวร์ที่ใช้ในการพัฒนาระบบนี้เป็นซอฟต์แวร์แบบ Open Source ได้แก่ Python version 3, NumPy, OpenCV, TensorflowPyTorch, Pillow, Scikit-image และ Keras

วิธีดำเนินการวิจัย

วิธีการที่นำเสนอในงานวิจัยนี้ประกอบด้วย 3 ส่วน คือ ส่วนที่ 1 การฝึกสอนระบบการเรียนรู้แบบเชิงลึก ส่วนที่ 2 การตรวจสอบพฤติกรรมที่ไม่เหมาะสมภายในภาพและวิดีโอด้วย Mask R-CNN และองค์ความรู้ที่ได้จากการฝึกสอนระบบการเรียนรู้แบบเชิงลึก และส่วนที่ 3 การเบลอสวนของพฤติกรรมที่ไม่เหมาะสมภายในภาพดิจิทัลและวิดีโอด้วยตัวกรองแบบเกาส์เซียน

1. การฝึกสอนระบบการเรียนรู้แบบเชิงลึก

การฝึกสอนกลุ่มข้อมูล จะเริ่มด้วยการเตรียมชุดข้อมูลภาพที่มีส่วนของบริเวณพฤติกรรมที่ไม่เหมาะสม โดยจัดเตรียมชุดข้อมูลภาพทั้งหมด 7,550 ภาพจะถูกแบ่งสำหรับ Training 80 เปอร์เซ็นต์ และสำหรับ Validation 20 เปอร์เซ็นต์ เมื่อจัดเตรียมชุดข้อมูลเรียบร้อยแล้ว ในขั้นตอนต่อไปจะนำชุดข้อมูลที่ได้จัดเตรียมไว้ อัปโหลดขึ้นไปบนเว็บไซต์ www.makesens.ai เพื่อทำการกำหนดส่วนของบริเวณพฤติกรรมที่ไม่เหมาะสมภายในภาพ และระบุหมายเลขกลุ่มให้กับทุกภาพเพื่อสร้างไฟล์ annotations ซึ่งเป็นไฟล์ .json หลังจากนั้นจะทำการนำข้อมูลที่จัดเตรียมไว้และไฟล์ annotations นำมาอัปโหลดขึ้น Google Drive เพื่อที่จะใช้ฝึกสอนใน Google Colab ซึ่งจะใช้ชุดคำสั่ง Mask R-CNN ในการฝึกสอน โดยจำเป็นต้องติดตั้งเครื่องมือที่ใช้ในการฝึกสอน ดังนี้ TensorFlow, h5py และ matplotlib ทำการ import Mask R-CNN เพื่อเรียกใช้ config และ model ในการฝึกสอน เมื่อเสร็จสิ้นการฝึกสอนจะได้ไฟล์ weight เป็นนามสกุล .h5 เพื่อนำไปใช้กับระบบในการตรวจสอบพฤติกรรมที่ไม่เหมาะสมในภาพและวิดีโอต่อไป สำหรับขั้นตอนวิธีการฝึกสอนระบบการเรียนรู้แบบเชิงลึกได้แสดงใน Figure 1

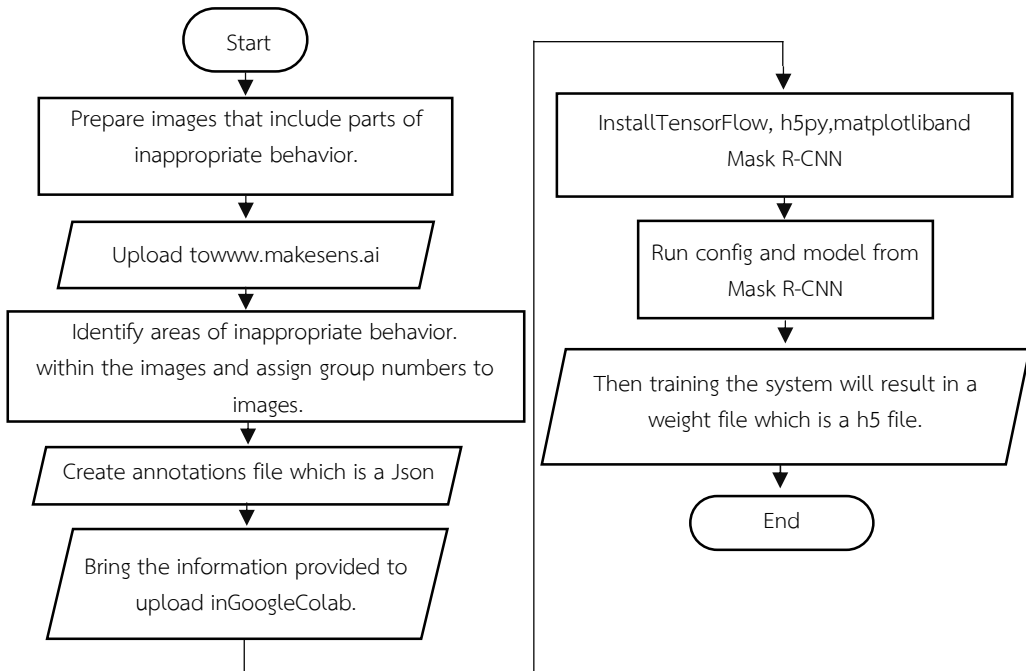


Figure 1 Training deep learning systems

2. การตรวจสอบพฤติกรรมที่ไม่เหมาะสมภายในภาพและวิดีโอด้วย Mask R-CNN และองค์ความรู้ที่ได้จากการฝึกสอนระบบการเรียนรู้แบบเชิงลึก

ในขั้นตอนนี้จะนำไฟล์ weight ที่ได้จากขั้นตอนที่ 1 มาทำการตรวจสอบพฤติกรรมที่ไม่เหมาะสมภายในภาพ ดังแสดงใน Figure 2 (a) โดยใช้โมเดลการเรียนรู้เชิงลึกแบบ Mask R-CNN เมื่อตรวจสอบได้ก็จะทำการกำหนดบริเวณนั้นให้เป็นส่วนของ



Figure 2 Original image (a) Image that marks the location of inappropriate areas (b)

หลังจากตรวจสอบเจอส่วนของภาพพฤติกรรมที่ไม่เหมาะสมแล้วจะนำตำแหน่งของบริเวณที่ตรวจสอบได้ไปใช้งานร่วมกับการเบลอด้วยตัวกรองแบบเกาส์เซียนในขั้นตอนต่อไป สำหรับขั้นตอนการตรวจสอบพฤติกรรมที่ไม่เหมาะสมในภาพและวิดีโอด้วย Mask R-CNN แสดงดัง Figure 3

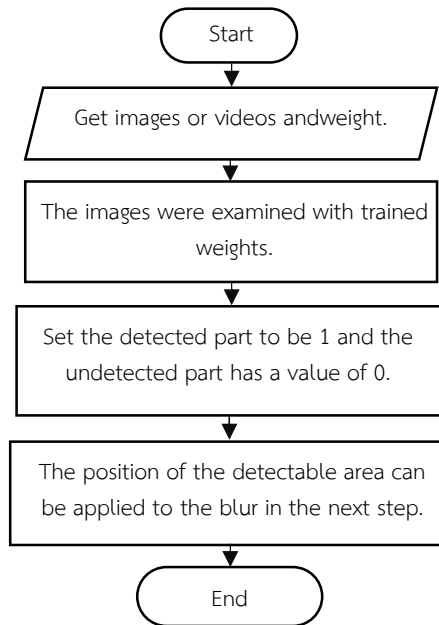


Figure 3 Detecting unsuitable behavior step in digital images and video by Mask R-CNN

3. การเบลอส่วนของพฤติกรรมที่ไม่เหมาะสมภายในภาพดิจิทัลและวิดีโอด้วยตัวกรองแบบเกาส์เซียน ในส่วนนี้ตำแหน่งของพฤติกรรมที่ไม่เหมาะสมที่ตรวจสอบได้จากข้อ 2 จะถูกทำให้เบลอด้วยวิธีการ Convolution โดยใช้ตัวกรองแบบเกาส์เซียนดังสมการที่ 1 และ 2 ตามลำดับ

$$(I * G_{\sigma})(i, j) = \sum_{k=1}^m \sum_{l=1}^n I(i - k, j - l) G_{\sigma}(k, l) \quad (1)$$

โดยที่ G_{σ} คือ

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y}{2\sigma^2}} \quad (2)$$

- x คือระยะห่างจากจุดกำเนิดในแกนแนวนอน
 y คือระยะห่างจากจุดกำเนิดในแกนตั้ง
 σ คือค่าเบี่ยงเบนมาตรฐาน โดยยิ่งค่าเบี่ยงเบนมาตรฐานมากก็จะเบลอมาก โดยแสดงดัง Figure 4

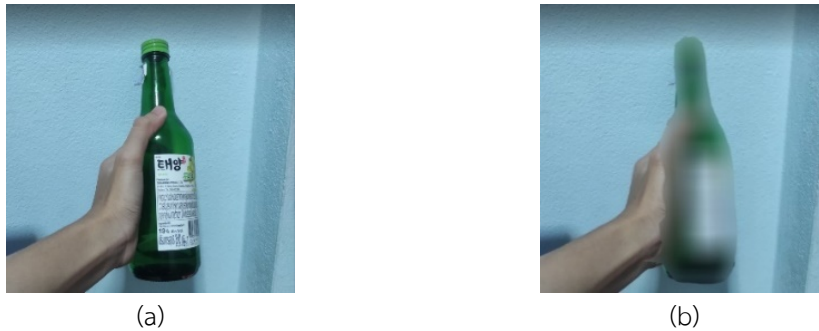


Figure 4 (a) Gaussian filter with $\sigma = 1$ and (b) $\sigma = 99$

จากนั้นจะนำผลลัพธ์ที่ได้ไปทำการซ้อนทับกับภาพหรือวิดีโอเดิมตามสมการที่ 3

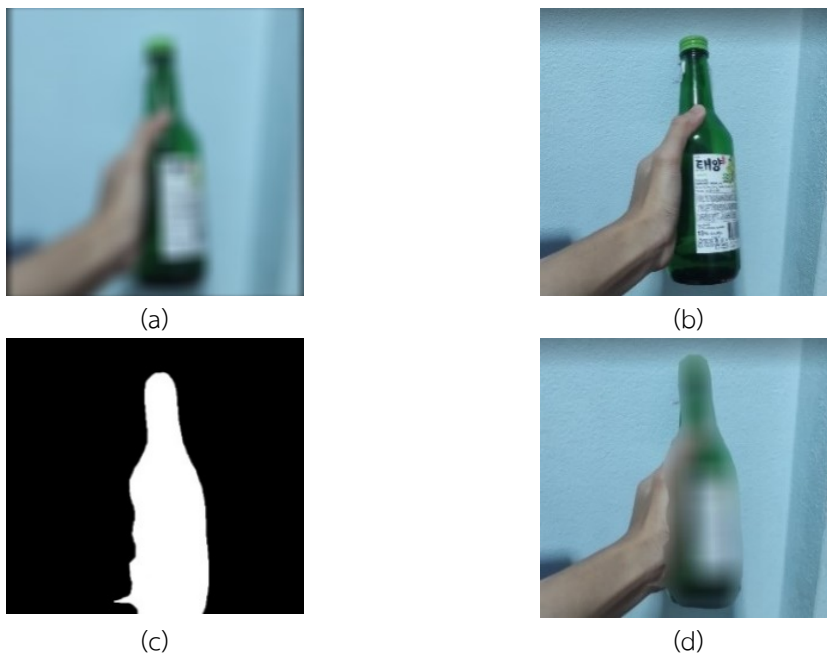


Figure 5 (a) Filtered image (b) Unfiltered image (c) opacity value image and (d) Image overlay results

$$C = \alpha F + (1 - \alpha)B \quad (3)$$

โดยที่ C คือภาพผลลัพธ์

B คือภาพที่ยังไม่เบลอ

F คือภาพที่ถูกเบลอ

α คือค่าความโปร่งแสงมีค่าอยู่ระหว่าง 0 และ 1 ถ้ามีค่าเป็น 1 แสดงว่าตำแหน่งนั้นเป็นส่วนของพฤติกรรมที่ไม่เหมาะสมและมีค่าเป็น 0 แสดงว่าตำแหน่งนั้นเป็นพฤติกรรมปกติ สำหรับตัวอย่างของการซ้อนภาพแสดงดัง Figure 5

สำหรับการเบลอภาพที่ไม่เหมาะสมในไฟล์วิดีโอนั้นจะทำการแยกเฟรมภาพทุกเฟรมภายในไฟล์วิดีโอออกมาจากนั้นนำแต่ละเฟรมภาพไปตรวจสอบหาบริเวณของส่วนพฤติกรรมที่ไม่เหมาะสมภายในแต่ละเฟรมภาพ ถ้าตรวจสอบพบในเฟรมภาพใดก็จะนำเฟรมภาพนั้นไปเบลอดังขั้นตอนที่กล่าวผ่านมา สุดท้ายจะนำเฟรมภาพที่ผ่านการตรวจสอบและถูกเบลอแล้วกลับไปสร้างเป็นไฟล์วิดีโออีกครั้ง

ผลการวิจัย

สำหรับการพัฒนาระบบการเรียนรู้แบบเชิงลึกของงานวิจัยนี้ได้พัฒนาโดยใช้ภาษา Python ใช้ไลบรารี Open CV, Matplotlib, NumPy, Torch Vision, Pillow, Scikit-Image, Tensor flow, Keras และใช้โมเดลการเรียนรู้แบบเชิงลึก Mask R-CNN โดยใช้ภาพบริเวณส่วนที่ปรากฏพฤติกรรมที่ไม่เหมาะสม 444 ภาพ ภาพเนื่องจากภาพหนึ่งภาพอาจจะปรากฏบริเวณส่วนของพฤติกรรมที่ไม่เหมาะสมได้มากกว่า 1 บริเวณ โดยจำนวนของบริเวณของพฤติกรรมที่ไม่เหมาะสมที่ปรากฏจริงที่นำมาใช้ในการทดสอบจึงมีทั้งสิ้น 509 บริเวณ และภาพที่ไม่ปรากฏบริเวณส่วนของพฤติกรรมที่ไม่เหมาะสมจำนวน 60 ภาพและใช้วิดีโอที่ปรากฏพฤติกรรมที่ไม่เหมาะสมจำนวน 100 วิดีโอ โดยการวัดประสิทธิภาพของระบบการเรียนรู้แบบเชิงลึกจะมีการนำเอาค่า Intersection over Union(IoU) และ Confusion Matrix มาใช้ในการวัดประสิทธิภาพของโมเดล

1. Intersection Over Union (IoU)

IoU เป็นการวัดประสิทธิภาพของระบบการเรียนรู้แบบเชิงลึกที่นิยมในด้านการตรวจจับวัตถุ (Object Detection) โดยเป็นการหาอัตราส่วนซ้อนทับกันของภาพ Ground truth ซึ่งเป็นภาพพื้นที่ของข้อมูลที่เป็นจริง และภาพผลการตรวจสอบด้วยโมเดลการเรียนรู้แบบเชิงลึก ซึ่งจะเป็นผลลัพธ์ของการใช้โมเดลตรวจสอบโดยคำนวณได้ดัง Figure 6

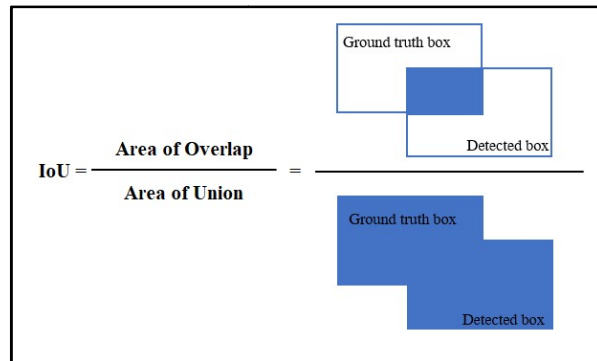


Figure 6 IoU Calculation

โดยการคำนวณค่า IoU ได้ใช้ภาพ Ground truth 30 ภาพและภาพผลการตรวจสอบด้วยโมเดลการเรียนรู้แบบเชิงลึกมาทำการหาอัตราส่วนการซ้อนทับกันแสดงดัง Figure 7 โดยกรอบสีเหลืองคือ Ground truth และกรอบสีน้ำเงินคือส่วนที่โมเดลตรวจสอบได้ และแสดงค่า IoU ของแต่ละภาพดัง Table 1



Figure 7 sample image of IoU

Table 1 Result of calculating the IoU

Image	IoU
Test(1)	0.74
Test(2)	0.86
Test(3)	0.80
Test(4)	0.72
Test(5)	0.88
Test(6)	0.61
Test(7)	0.89
Test(8)	0.34
Test(9)	0.71
Test(10)	0.82
Test(11)	0.89
Test(12)	0.43
Test(13)	0.79
Test(14)	0.43
Test(15)	0.35
Test(16)	0
Test(17)	0.87
Test(18)	0.76
Test(19)	0.480
Test(20)	0.62
Test(21)	0.75

Table 1 Result of calculating the IoU (continue)

Image	IoU
Test(22)	0.68
Test(23)	0.58
Test(24)	0.34
Test(25)	0.87
Test(26)	0.68
Test(27)	0.64
Test(28)	0.70
Test(29)	0.30
Test(30)	0.88

จาก Table1 นำค่า IoU ที่ได้มาคำนวณ Average IoU ได้เท่ากับ 0.65

2. Confusion Matrix

Confusion Matrix คือตารางที่ใช้วัดความสามารถของระบบการเรียนรู้แบบเชิงลึกที่ผ่านการฝึกสอน โดยค่าที่ได้จากตารางนี้จะมีทั้งหมด 4 ค่า คือ

True Positive (TP) คือค่าที่เป็นจริงและระบบทำนายว่ามีภาพพฤติกรรมที่ไม่เหมาะสม

True Negative (TN) คือค่าที่ไม่จริงและระบบทำนายว่าไม่มีภาพพฤติกรรมที่ไม่เหมาะสม

False Positive (FP) คือค่าที่ไม่จริงแต่ระบบทำนายว่ามีภาพพฤติกรรมที่ไม่เหมาะสม

False Negative (FN) คือค่าที่เป็นจริงแต่ระบบทำนายว่าไม่มีภาพพฤติกรรมที่ไม่เหมาะสม

โดยค่าทั้ง 4 นี้จะนำมาใช้คำนวณหาตัววัดประสิทธิภาพของระบบการเรียนรู้แบบเชิงลึก 4 ค่าคือ Accuracy, Precision, Recall และ F1 Score โดยที่ Accuracy เป็นความแม่นยำของระบบ คำนวณได้จากสมการที่ 4

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

สำหรับ Precision เป็นการวัดความเที่ยงตรงของระบบ คำนวณได้จากสมการที่ 5

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

สำหรับ Recall เป็นการวัดความถูกต้องของระบบ คำนวณได้จากสมการที่ 6

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

และสำหรับ F1 Score เป็นค่าที่แสดงประสิทธิภาพของระบบ คำนวณได้จากสมการที่ 7

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

โดยจะนำตัวชี้วัดเหล่านี้ไปคำนวณเพื่อประเมินประสิทธิภาพของระบบในขั้นตอนต่อไป

2.1 ผลการทดลองการเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับภาพดิจิทัล

จากการทดสอบการเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับภาพดิจิทัลทั้งหมด 504 ภาพ แบ่งเป็น ภาพพฤติกรรมการดื่มแอลกอฮอล์ 103 ภาพ ภาพการถือมีด 50 ภาพ ภาพการสูบบุหรี่ไฟฟ้า 34 ภาพ ภาพการสูบบุหรี่ 105 ภาพ ภาพการถืออาวุธปืน 102 ภาพ ภาพการสูบกัญชา 50 ภาพ และภาพที่ไม่มีพฤติกรรมที่ไม่เหมาะสม 60 ภาพ โดยผลที่ได้จาก Confusion Matrix ดังแสดงใน Figure 8

		Actual Values	
		Positive	Negative
Predicted Values	Positive	450	59
	Negative	43	17

Figure 8 The Confusion Matrix of blurring inappropriate behavior in digital images

จาก Figure 8 เมื่อพิจารณา Confusion Matrix จะได้ค่าของ TP เท่ากับ 450 ค่าของ TN เท่ากับ 17 ค่าของ FP เท่ากับ 59 และค่าของ FN เท่ากับ 43 จากนั้นทำการคำนวณหาค่า Accuracy, Precision, Recall และ F1 Score โดยแสดงดัง Table 2

Research Article

Journal of Advanced Development in Engineering and Science

Vol. 14 • No. 39 • January – April 2024

Table 2 Experimental results of inappropriate blurring behavior for digital images

Accuracy	Precision	Recall	F1 Score
0.82	0.88	0.91	0.90

เมื่อทำการวัดประสิทธิภาพของระบบการเรียนรู้แบบเชิงลึกพบว่ามีความ Accuray เท่ากับ 0.82 Precision เท่ากับ 0.88 Recall เท่ากับ 0.91 และ F1 Score 0.9 ซึ่งทั้ง 4 ค่ามีค่าเกิน 0.8 แสดงให้เห็นว่าระบบการเรียนรู้แบบเชิงลึกของงานวิจัยนี้มีความแม่นยำในการทำนายสูง

ตัวอย่างภาพของผลการทดลองการเบลอพฤติกรรมที่ไม่เหมาะสมสำหรับภาพดิจิทัลที่สามารถตรวจสอบและทำการเบลอส่วนของพฤติกรรมที่ไม่เหมาะสมภายใต้กฎข้อบังคับซึ่งมีค่า TP คือค่าที่เป็นจริงและระบบทำนายว่ามีภาพพฤติกรรมที่ไม่เหมาะสมแสดงดัง Figure 9 – 14



(a)

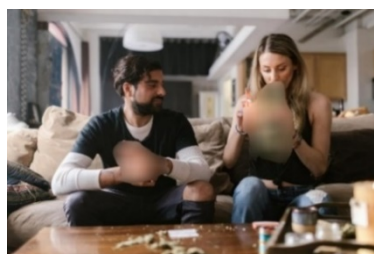


(b)

Figure 9 (a) Original image (b) Correctly blurred inappropriate image

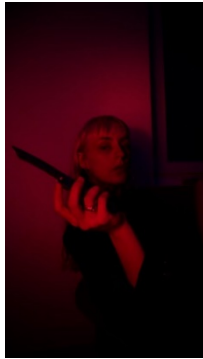


(a)

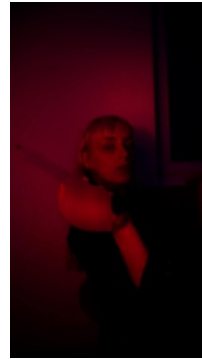


(b)

Figure 10 (a) Original image (b) Correctly blurred inappropriate image



(a)



(b)

Figure 11 (a) Original image (b) Correctly blurred inappropriate image



(a)



(b)

Figure 12 (a) Original image (b) Correctly blurred inappropriate image



(a)



(b)

Figure 13 (a) Original image (b) Correctly blurred inappropriate image



(a)



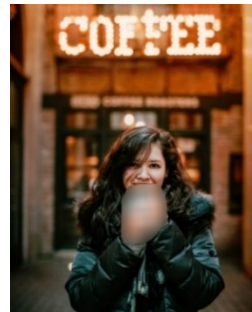
(b)

Figure 14 (a) Original image (b) Correctly blurred inappropriate image

ตัวอย่างภาพของผลการทดลองการเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับภาพดิจิทัลที่ตรวจสอบและทำการเบลอส่วนของพฤติกรรมที่ไม่เหมาะสมได้ไม่ถูกต้อง ซึ่งมีค่า FP คือค่าที่ไม่จริงแต่ระบบทำนายว่ามีภาพพฤติกรรมที่ไม่เหมาะสมแสดงดัง Figure 15 -17



(a)



(b)

Figure 15 (a) Original image (b) Incorrectly blurs the inappropriate image



(a)



(b)

Figure 16 (a) Original image (b) Incorrectly blurs the inappropriate image

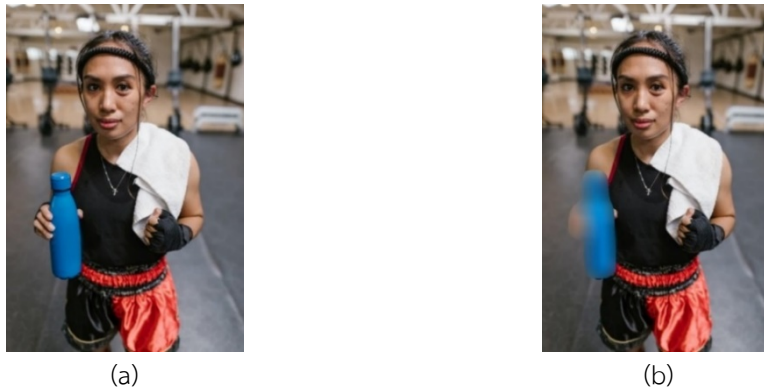


Figure 17 (a) Original image (b) Incorrectly blurs the inappropriate image

2.2 ผลการทดลองการเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับวิดีโอ

จากการทดสอบการเบลอภาพพฤติกรรมที่ไม่เหมาะสมสำหรับวิดีโอทั้งหมด 100 วิดีโอ (13,286 เฟรมภาพ) แบ่งเป็นวิดีโอพฤติกรรมการดื่มแอลกอฮอล์ 20 วิดีโอ (3,309 เฟรมภาพ) วิดีโอการถ่มน้ำลาย 13 วิดีโอ (868 เฟรมภาพ) วิดีโอการสูบบุหรี่ไฟฟ้า 20 วิดีโอ (3,034 เฟรมภาพ) วิดีโอการสูบบุหรี่ 19 วิดีโอ (2,803 เฟรมภาพ) วิดีโอการถืออาวุธปืน 18 วิดีโอ (1,821 เฟรมภาพ) วิดีโอการสูบกัญชา 10 วิดีโอ (1,451 เฟรมภาพ) โดยผลที่ได้จาก Confusion Matrix แสดงดัง Figure 18

		Actual Values	
		Positive	Negative
Predicted Values	Positive	12,260	865
	Negative	1,111	1,915

Figure 18 The Confusion Matrix of blurring inappropriate behavior in videos

ค่าของ TP เท่ากับ 12,260 ค่าของ TN เท่ากับ 1,915 ค่าของ FP เท่ากับ 865 และค่าของ FN เท่ากับ 1,111 จากนั้นทำการคำนวณหาค่า Accuracy, Precision, Recall และ F1 Score โดยแสดงดัง Table 3

Table 3 Experimental results of inappropriate blurring behavior for videos

Type	Accuracy	Precision	Recall	F1 Score
Inappropriate	0.88	0.93	0.92	0.93

ตัวอย่างผลการทดลองการเบลอฟฤติกรรมที่ไม่เหมาะสมสำหรับวิดีโอแสดงดัง Figure 19 – 21



(a) Original image frame



(b) Blurred image frame



(c) Original image frame



(d) Blurred image frame



(e) Original image frame



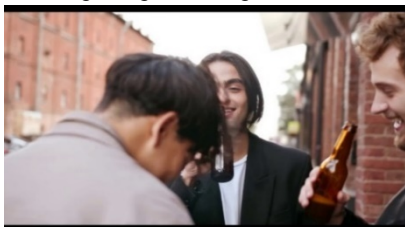
(f) Blurred image frame



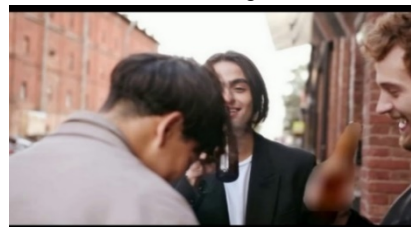
(g) Original image frame



(h) Blurred image frame



(i) Original image frame



(j) Blurred image frame

Figure 19 Example of the original image frame (Left) The result of blurring inappropriate behavior (Right)



(a) Original image frame



(b) Blurred image frame



(c) Original image frame



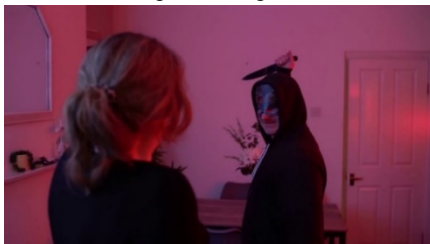
(d) Blurred image frame



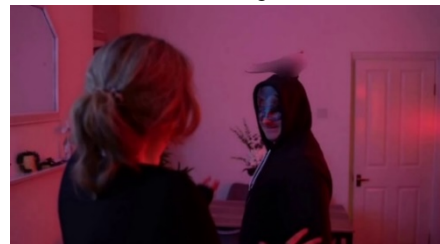
(e) Original image frame



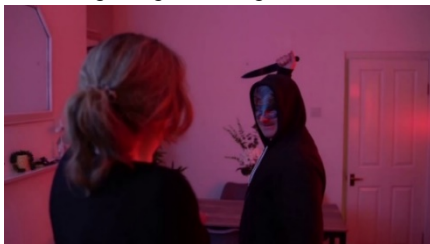
(f) Blurred image frame



(g) Original image frame



(h) Blurred image frame



(i) Original image frame



(j) Blurred image frame

Figure 20 Example of the original image frame (Left) The result of blurring inappropriate behavior (Right)



(a) Original image frame



(b) Blurred image frame



(c) Original image frame



(d) Blurred image frame



(e) Original image frame



(f) Blurred image frame



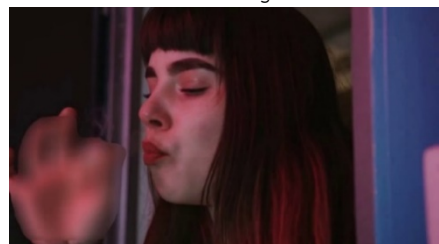
(g) Original image frame



(h) Blurred image frame



(i) Original image frame



(j) Blurred image frame

Figure 21 Example of the original image frame (Left) The result of blurring inappropriate behavior (Right)

อภิปรายและสรุปผลการวิจัย

เนื่องจากค่าของการวัดประสิทธิภาพของระบบในงานวิจัยนี้มีค่าที่สูงส่วนหนึ่งเป็นผลมาจากการกำหนดกลุ่มข้อมูลที่น้อยในงานวิจัยนี้ได้กำหนดไว้ 6 กลุ่มข้อมูลคือ การสูบบุหรี่ การสูบบุหรี่ไฟฟ้า การสูบกัญชา การถืออาวุธปืน การถือมีด และการถือเครื่องดื่มแอลกอฮอล์ ซึ่งในความเป็นจริงแล้วภาพพฤติกรรมที่ไม่เหมาะสมไม่ได้มีเพียง 6 กลุ่มข้อมูลที่กำหนดภายในงานวิจัยนี้ ยังขาดภาพในลักษณะอื่นๆอีก เช่น ภาพการทารุณกรรมสัตว์ ภาพของการทำร้ายร่างกาย ภาพของการที่บุคคลใช้ความรุนแรงในการทำร้ายบุคคลอื่น และยังมีภาพที่บุคคลสามารถใช้อุปกรณ์ที่ไม่ใช่อาวุธเช่นมีด อาจจะเป็นค้อนหรือไม้มาเป็นอาวุธ เพราะฉะนั้นด้วยจำนวนกลุ่มข้อมูลที่น้อยจึงมีผลทำให้การตรวจสอบข้อมูลของระบบทำได้แม่นยำสูง ดังนั้นหากเพิ่มกลุ่มข้อมูลเหล่านี้เข้าไปเพื่อให้สามารถตรวจสอบภาพพฤติกรรมที่ไม่เหมาะสมได้มากขึ้น ผลของค่าที่ใช้วัดประสิทธิภาพก็น่าจะมีค่าลดลง จากความผิดพลาดในการตรวจสอบที่เห็นได้จาก Figure 15 – 17 แสดงให้เห็นว่ากลุ่มข้อมูลที่ใช้ในการฝึกสอนยังขาดความหลากหลายและมีจำนวนข้อมูลที่ใช้ฝึกสอนระบบน้อยเกินไป จึงทำให้การตรวจสอบว่าการถือถ้วยกาแฟดัง Figure13 กลายเป็นภาพพฤติกรรมที่ไม่เหมาะสมและด้วยเหตุผลเดียวกันกับที่เกิดกับ Figure16 และ 17

สรุปผลการวิจัย

ในงานวิจัยนี้ได้พัฒนาระบบการเรียนรู้แบบเชิงลึกเพื่อทำการเบลอภาพของพฤติกรรมที่ไม่เหมาะสมภายในภาพหรือวิดีโอโดยใช้โมเดลการเรียนรู้เชิงลึกแบบ Mask R-CNN กระบวนการฝึกสอนโมเดลการเรียนรู้ได้ใช้ภาพจำนวน 7,550 ภาพในการฝึกสอนระบบการเรียนรู้แบบเชิงลึกให้เกิดองค์ความรู้ใหม่เพื่อนำมาใช้ในการตรวจสอบ จากการทำงานของระบบพบว่าระบบสามารถทำการเบลอภาพพฤติกรรมที่ไม่เหมาะสมได้ถูกต้องโดยในงานวิจัยนี้ได้ใช้ตัววัดประสิทธิภาพของการเรียนรู้แบบเชิงลึกสำหรับภาพดิจิทัลโดยมีค่า Accuracy, Precision, Recall และ F1 Score ระบบสามารถเบลอในส่วนของพฤติกรรมที่ไม่เหมาะสมในภาพดิจิทัลและคลิปวิดีโอได้อย่างเหมาะสม

ในงานวิจัยนี้พบปัญหาที่เกิดขึ้นในระหว่างตรวจสอบพฤติกรรมที่ไม่เหมาะสมภายในภาพหรือภาพวิดีโอคือ มีบางส่วนที่ทำงานไม่ถูกต้องโดยเฉพาะกับภาพที่มีวัตถุที่คล้ายกับสิ่งที่ไม่เหมาะสม เช่น ขวดน้ำ ซึ่งคล้ายกับขวดแอลกอฮอล์ การจับปากกาที่มีรูปร่างคล้ายกับบุหรี่ เป็นต้น เนื่องจากภาพที่นำมาฝึกสอนนั้นมีไม่เพียงพอ

References

- [1] Bengio, Y. & Lecun, Y. (1998). Convolutional Networks for Images, Speech, and Time-Series. In Arbib, M. A. (Ed.). *The Handbook of Brain Theory and Neural Networks* (p. 255–258). Massachusetts: The MIT Press.
- [2] Girshick, R. (2015). Fast R-CNN. In *2015 IEEE International Conference on Computer Vision* (p. 1440-1448). 11 – 18 December, 2015, Araucano Park, Chile.
- [3] Girshick, R., et al. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (p. 580–587). 23 – 28 June, 2014, Columbus, Ohio, USA.
- [4] He, K. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer*

Research Article

Journal of Advanced Development in Engineering and Science

Vol. 14 • No. 39 • January – April 2024

- Vision (p. 2980–2988). 22 – 29 October, 2017, Venice, Italy.
- [5] Redmon, J. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* (p. 779 - 788). 26 June – 1 July, 2016, Las Vegas, NV, USA.
- [6] Wongta, P. (2017). *Vision-Based Bus Route Number Reader*, (Master thesis, Chulalongkorn University). (in Thai)
- [7] Sanuksan, J. & Surinta, O. (2019). Deep Convolutional Neural Networks for Plant Recognition in Natural Environment, *Journal of Science and Technology Mahasarakham University*, 38(2), 113–124. (in Thai)
- [8] Manokij, F. (2019). *Thailand's Precipitation Forecasting Using Deep Learning Approach*, (Master thesis, Chulalongkorn University). (in Thai)
- [9] Yongying, N. (2019). *A Deep Learning Approach on Road Detection from Unmanned Aerial Vehicle-Based Images in Rural Road Monitoring*, (Master thesis, Naresuan University). (in Thai)
- [10] Rattanachot, P. (2019). *Automated Plant Disease Detection using Drones and Deep Learning*, (Master Thesis, Dhurakij Pundit University). (in Thai)
- [11] Posawang, P., et al. (2021). The Road Surface Anomalies Detection using Deep Convolutional Neural Networks with Transfer Learning Technique. *Information Technology Journal*, 17(1), 31-42. (in Thai)
- [12] Metkarunchit, T. & Charoenpojvajara, K. (2020). Detection of COVID-19 using Deep Learning with CT Scan Images. *TNI Journal of Engineering and Technology*, 8(2), 8-17. (in Thai)
- [13] Tipakorn, K. & Tanasai, S. (2021). Rice Image Segmentation using Deep Learning. In *The 11th National and the 5th International PIM Conference* (p.510 – 523). 16 July, 2021, Nonthaburi, Thailand. (in Thai)
- [14] Worasit, T., et al. (2022). Rice Bacterial Blight And Blast Diseases Recognition using Deep Learning Techniques. *Khon Kaen Agriculture Journal*, 50(1), 216-228. (in Thai)
- [15] Parinee, A., et al. (2022). Laser Pointer Control for Deep Learning Human Detection. *Journal of Science and Technology Mahasaraknam University*, 41(3), 151-163. (in Thai)
- [16] Saisangchan, U. (2022). Analysis of Lime Leaf Disease using Deep Learning. *Journal of Applied Informatics and Technology*, 4(1), 71-86. (in Thai)