Received: 29 เม.ย. 2566 **Revised:** 24 ส.ค. 2566 **Accepted:** 28 ส.ค. 2566

การแบ่งส่วนภาพปลานิลสองขั้นตอนโดยใช้การเรียนรู้อย่างลึกด้วยเครือข่ายคอนโวลูชันพีระมิดเชิงพื้นที่

Two-Stage Tilapia Image Segmentation using Deep Learning with Spatial Pyramid Convolutional Network

เพชรรัช ปะระไทย[1], นฤพร เต็งไตรรัตน์[1*] และ ชัชวาลย์ ชัยชนะ[2]
สาขาวิชาวิศวกรรมซอฟต์แวร์ มหาวิทยาลัยพายัพ[1] และ
ภาควิชาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่[2]

Phetcharat Parathai[1], Naruephorn Tengtrairat[1*] and Chatchawan Chaichana[2]
School of Software Engineering, Payap University[1] and
Department of Mechanical Engineering, Faculty of Engineering, Chiang Mai University[2]
[*]Corresponding author: naruephorn_t@payap.ac.th

Abstract

Tilapia is one of the most important fish in the aquaculture industry of Thailand. This research proposes a method for creating a Tilapia Image Segmentation model using deep learning with spatial pyramid convolutional network based YOLOv8 to identify the pixels of Tilapia swimming in turbid water. The process involves creating a set of images of Tilapia in water using an underwater camera. The training process of Tilapia Image Segmentation is done using Supervised Deep Transferred Learning. The structure of the YOLOv8 model is elucidated and discussed. The performance of Tilapia Image Segmentation based on YOLOv8 is evaluated and compared to a Mask R-CNN model. The results show that under a small-sized image dataset, the YOLOv8 model has better accuracy in determining Tilapia pixels from various perspective images than the model-based Mask R-CNN method, with an accuracy improvement of up to 6% on average. The YOLOv8 model has an Area Under the Curve (AUC) value of 0.988, indicating its ability to identify Tilapia pixels in images with turbid water and unclear fish shapes. This method does not require any image quality enhancement processes, which benefits the reduction of complexity in development, while still accurately

identifying Tilapia pixels. Moreover, using the small-sized Tilapia-image dataset in this method reduces data preparation costs and low computational complexity.

Keywords: *Image segmentation, deep learning, YOLO, machine learning, computer vision*

บทคัดย่อ

ปลานิลเป็นปลาที่สำคัญที่สุดในอุตสาหกรรมการเลี้ยงสัตว์น้ำของประเทศไทย งานวิจัยนี้เสนอ วิธีการการสร้างแบบจำลอง Tilapia Image Segmentation ด้วยวิธีการเรียนรู้อย่างลึกด้วยเครือข่ายคอนโวลูชันพีระมิดเชิงพื้นที่โดยใช้ YOLOv8 เพื่อระบุพิกเซลของปลานิลที่ว่ายอยู่ในน้ำขุ่น โดยสร้างชุดรูปภาพปลานิลในน้ำจากกล้องบันทึกภาพใต้น้ำ กระบวนการฝึกฝน Tilapia Image Segmentation แบบ Supervised Deep Transferred Learning การวิเคราะห์โครงสร้างแบบจำลอง YOLOv8 และเปรียบเทียบประสิทธิภาพการแบ่งส่วนภาพปลานิลระหว่างแบบจำลองจากวิธี YOLOv8 และวิธี Mask R-CNN จากผลการทดสอบแสดงให้เห็นว่า แบบจำลองแบ่งกลุ่มพิกเซลปลานิลของวิธี YOLOv8 มีความแม่นยำในการแบ่งส่วนภาพปลานิลในหลากหลายมุมมอง ได้ดีกว่าแบบจำลองที่สร้างด้วยวิธี Mask R-CNN ถึงร้อยละ 6 ของค่าเฉลี่ยความแม่นยำ วิธี YOLOv8 ใช้จำนวนชุดคำสั่งในการฝึกฝนเรียนรู้ของแบบจำลองน้อยกว่าวิธี Mask R-CNN ถึงร้อยละ 54.6 และมีค่าพื้นที่ใต้เส้นโค้ง (Area Under the Curve: AUC) ที่ระดับ 0.988 นอกจากนี้วิธี YOLOv8 สามารถแบ่งพิกเซลของปลานิลจากรูปภาพที่มีความขุ่นของน้ำและความไม่ชัดเจนของตัวปลาได้ โดยไม่มีกระบวนการรปรับปรุงคุณภาพของรูปภาพ

คำสำคัญ: *การแบ่งส่วนภาพ, การเรียนรู้เชิงลึก YOLO, การเรียนรู้ของเครื่อง, วิสัยทัศน์คอมพิวเตอร์*

1. บทนำ

ผลผลิตและมูลค่าการเพาะเลี้ยงสัตว์น้ำในประเทศไทยเพิ่มขึ้นอย่างต่อเนื่อง ปี พ.ศ. 2565 - 2568 คาดการณ์ว่าจะมีแนวโน้มเพิ่มขึ้นอัตราร้อยละ 1.92 และ 3.63 ต่อปี ตามลำดับ (กลุ่มสถิติการประมง กองนโยบายและแผนพัฒนาการประมง, 2565) กรมประมงรายงานว่า การเพาะเลี้ยงสัตว์น้ำจืด ปลานิลเป็นหนึ่งในปลาที่สำคัญที่สุดในอุตสาหกรรมการเลี้ยงสัตว์น้ำของประเทศไทย ปลานิลเป็นปลาน้ำจืดที่เลี้ยงเพื่อการค้าอย่างแพร่หลายเนื่องจากมีอัตราการเติบโตเร็ว ตามรายงานจากกรมประมง (กลุ่มสถิติการประมง กองนโยบายและแผนพัฒนาการประมง, 2565) คาดการณ์ผลผลิตเฉลี่ย พ.ศ. 2565 - 2568 ผลผลิตปลานิลคิดเป็น 25% ของผลผลิตและมูลค่าสัตว์น้ำจากการเพาะเลี้ยงของ

ประเทศไทย ปัจจุบันการวัดขนาด นับจำนวน และชั่งน้ำหนักปลา ใช้เวลาและแรงงานคนอย่างมาก นอกจากนี้ยังทำให้เกิดความเครียดหรือเกิดการบาดเจ็บกับตัวปลาได้อีกด้วย การนำเทคโนโลยีซอฟต์แวร์มาช่วยในการติดตามการเจริญเติบโต โดยระบุขนาดตัวปลาขั้นตอนแรกเพื่อให้ได้ข้อมูลขนาดตัวปลา จากนั้นจะนำข้อมูลขนาดของตัวปลาไปประมวลผลวิเคราะห์น้ำหนักของปลาต่อไป ซึ่งสามารถใช้น้ำหนักตัวเพื่อกำหนดความต้องการสารอาหาร ปริมาณยา และความพร้อมในการออกสู่ตลาดได้

วิธีการตรวจหาและระบุขนาดของตัวปลาเป็นหนึ่งในปัญหาที่ท้าทายที่สุดในการเลี้ยงสัตว์น้ำ การระบุขนาดปลาด้วยอัลกอริทึมการแบ่งกลุ่มพิกเซลปลาได้ถูกพัฒนาและนำเสนออย่างต่อเนื่อง อาทิ วิธีการปรับปรุงการแบ่งกลุ่มปลาในภาพใต้น้ำโดยใช้เทคนิค Convolutional Neural Networks (CNNs) (Tengtrairat, 2022) และเทคนิค Data Augmentation (Hong Khai, 2022) เป็นการแบ่งกลุ่มปลาด้วยวิธี U-Net ที่เป็นการใช้เทคนิค CNNs ในการดึงคุณลักษณะสำคัญของปลาจากรูปภาพ และทำการประมวลผลภาพโดยใช้เทคนิค Data Augmentation โดยใช้การหมุนภาพแบบสุ่ม การปรับขนาดภาพ และการพลิกภาพเพื่อสร้างตัวอย่างการฝึกอบรมใหม่ ทำให้เพิ่มขนาดของชุดข้อมูลซึ่งเป็นสิ่งสำคัญในการฝึกโมเดล Deep Learning เพื่อเพิ่มความแม่นยำของโมเดล นอกจากนี้ Huang และทีม (Huang, 2022) ได้นำเสนอวิธีการแยกส่วนปลาจากวิดีโอใต้น้ำอย่างแม่นยำโดยใช้เทคนิค Deep CNNs ที่ใช้โมเดลต่าง ๆ ได้แก่ U-Net, VGG-16, และ MobileNet วิธีการที่นำเสนอมีความสามารถในการตรวจจับและแบ่งส่วนปลาหลายตัวในเฟรมวิดีโอเดียว

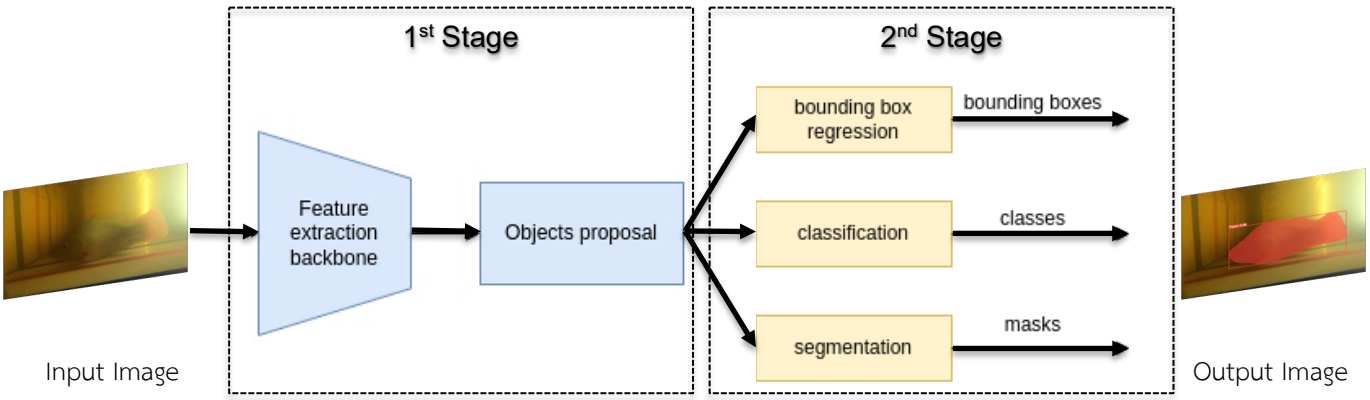
การแบ่งส่วนภาพปลาด้วยโมเดล U-Net, VGG-16, และ MobileNet ต้องการข้อมูลจำนวนมากเพื่อฝึกโมเดลได้อย่างมีประสิทธิภาพ โครงสร้างของโมเดลมีความลึก และจำเป็นต้องใช้การคำนวณมาก ทำให้ใช้เวลานานและค่าใช้จ่ายสูงในการพัฒนาแบบจำลอง ตัวอย่างเช่น รูปภาพขนาด 256x256 พิกเซลและขนาดแบทช์จำนวน 1 รูปภาพ โมเดล U-Net จะใช้จำนวน Floating-Point Operations (FLOPs) ประมาณ 300-400 พันล้านฟลอป (FLOPs) ในขณะที่โมเดล VGG-16 และโมเดล MobileNet ประมาณ 30-40 พันล้าน FLOPs และ 600-700 ล้าน FLOPs ตามลำดับ โดยทั่วไปโมเดล MobileNet มีระดับความแม่นยำน้อยกว่าโมเดล U-Net และ VGG-16 จึงอาจไม่เหมาะสำหรับการแบ่งแยกภาพที่ซับซ้อน สำหรับการบันทึกภาพใต้น้ำได้รับอิทธิพลจากปัจจัยเชิงซ้อนที่ซับซ้อน อาทิ การเปลี่ยนแปลงความส่องสว่าง ความขุ่น พื้นหลังที่หลากหลาย สภาพแวดล้อมที่ไม่แน่นอนซึ่งเกิดจากการเปลี่ยนแปลงของแสงและเงา รวมถึงการเคลื่อนไหวของสัตว์น้ำ ส่งผลให้ภาพที่บันทึกได้มีคุณภาพต่ำ ดังนั้นการบันทึกภาพปลาใต้น้ำส่วนใหญ่ต้องใช้กล้องพิเศษที่มีความซับซ้อนสำหรับบันทึกรูปปลาในน้ำ ได้แก่ Li และทีม (Li, 2023) ใช้อุปกรณ์โซนาร์ที่เฉพาะเจาะจงสำหรับการใช้ในการบันทึกภาพ Side-scan sonar ใต้น้ำ เพื่อนำรูปภาพที่ได้มาใช้สำหรับการแบ่งกลุ่มพิกเซลปลา

งานวิจัยนี้เสนอวิธีการแบ่งส่วนภาพปลานิล (Tilapia Image Segmentation) ในน้ำขุ่น โดยใช้กล้องบันทึกภาพใต้น้ำ 1 ตัว (หาซื้อได้ทั่วไป) แบบไม่สัมผัสและติดตั้งอุปกรณ์ที่ตัวปลา เพื่อช่วยลดความเครียดและการบาดเจ็บของปลาจากการสัมผัสตัวปลา และใช้ชุดข้อมูลรูปภาพปลานิลขนาดเล็กสำหรับการสร้างแบบจำลอง Tilapia Image Segmentation ที่สามารถแบ่งกลุ่มพิกเซลปลานิลได้อย่างมีประสิทธิภาพ

โครงสร้างของบทความนี้ประกอบด้วยส่วนหลัก 5 ส่วน ได้แก่ ส่วนที่ (1) บทนำ อธิบายถึงที่มา ความสำคัญ งานวิจัยที่มีความเกี่ยวข้องกับ ส่วนที่ (2) นำเสนออัลกอริทึมการการแบ่งส่วนภาพปลานิล ส่วนที่ (3) แสดงผลการทดสอบประสิทธิภาพของวิธีการการแบ่งส่วนภาพปลาทิลาเปียสองขั้นตอนโดยใช้การเรียนรู้อย่างลึกด้วยเครือข่ายคอนโวลูชันพีระมิดเชิงพื้นที่และเปรียบเทียบกับวิธีการแบ่งกลุ่มพิกเซลที่ทันสมัย ในส่วนที่ (4) อธิบายประสิทธิภาพของอัลกอริทึมการแบ่งส่วนภาพปลานิลและส่วนสุดท้าย ส่วนที่ (5) สรุปผลงานวิจัย

2. การแบ่งส่วนภาพปลานิล (Tilapia Image Segmentation)

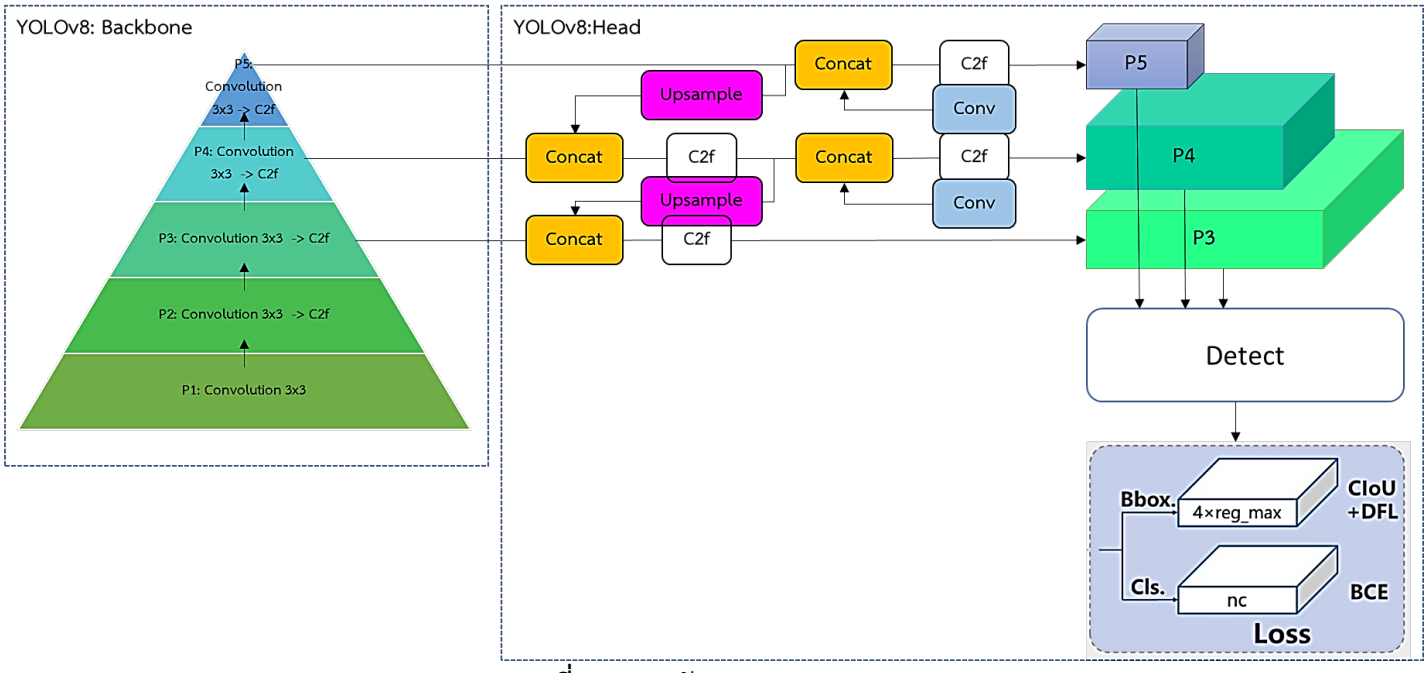
Tilapia Image Segmentation ได้ถูกพัฒนาขึ้นบนหลักการแบ่งกลุ่มแบบ 2 ขั้นตอน (Two-stage instance segmentation) ประกอบด้วยสองขั้นตอนหลัก คือการตรวจหาวัตถุและการแบ่งกลุ่มแบบอินสแตนซ์ ในขั้นตอนแรก จะใช้อัลกอริทึมการตรวจหาวัตถุเพื่อระบุวัตถุที่มีอยู่ในภาพและวาดกรอบสี่เหลี่ยมรอบวัตถุเรียกว่า กล่องขอบเขต (Bounding boxes) ในขั้นตอนที่สอง จะใช้อัลกอริทึมการแบ่งกลุ่มแบบอินสแตนซ์เพื่อปรับปรุงผลการตรวจหาวัตถุและแบ่งกลุ่มแต่ละวัตถุเป็นอินสแตนซ์ต่าง ๆ ดังแสดงแผนภาพการทำงานของวิธี Two-stage instance segmentation ในรูปที่ 2.1



รูปที่ 2.1 โครงสร้างการตรวจจับวัตถุแบบ Two-stage instance segmentation

ขั้นตอนแรกคือการตรวจจับวัตถุ (object detection) ที่อัลกอริทึมจะระบุวัตถุที่อยู่ในภาพและระบุพื้นที่ของกล่องบริบทในภาพ วิธีการตรวจจับวัตถุที่ใช้กันอย่างแพร่หลาย เช่น Faster R-CNN หรือ YOLO โดยทั่วไปในขั้นตอนนี้จะใช้โมเดลการเรียนรู้ลึก (deep learning) เพื่อตรวจจับและจำแนกวัตถุตามหมวดหมู่ที่กำหนดไว้ล่วงหน้า ในขั้นตอนที่สอง อัลกอริทึมจะดำเนินการแบ่งส่วนระดับพิกเซลภายในกล่องขอบเขตที่ได้รับจากขั้นตอนแรก ซึ่งหมายความว่าวัตถุที่ตรวจจับได้ในขั้นตอนแรกจะถูกปรับปรุงเพิ่มเติมเพื่อกำหนดเส้นขอบของแต่ละวัตถุให้แม่นยำที่ระดับพิกเซล เป้าหมายคือการแบ่งส่วนแต่ละอินสแตนซ์ในภาพอย่างถูกต้องและกำหนดป้ายชื่อที่ไม่ซ้ำกันให้กับแต่ละพิกเซลที่เป็นของวัตถุเฉพาะ

โมเดล Tilapia Image Segmentation ถูกสร้างตามหลักการ Two-stage instance segmentation โดยใช้วิธี YOLOv8 ซึ่งเป็นสถาปัตยกรรมล่าสุด (state-of-the-art:SOTA) ของวิธี You Only Look Once (YOLO) ถูกเผยแพร่ครั้งแรกในปี 2559 (Wageeh, 2021) เป็นอัลกอริทึมการเรียนรู้ด้วยวิธี Deep Learning ที่ออกแบบมาเพื่อตรวจจับวัตถุภายในภาพและระบุตำแหน่งของวัตถุด้วย Bounding boxes จากนั้นได้ถูกพัฒนาอย่างต่อเนื่อง YOLOv8 ได้ถูกนำเสนอในปี 2566 โดย Ultralytics (2023) สถาปัตยกรรมของระบบประกอบด้วย 2 ส่วนหลัก ดังแสดงในรูปที่ 2.2 ได้แก่ 1) Backbone เป็นส่วนที่ใช้สำหรับการแยกแยะลักษณะเด่นในอัลกอริทึม โดยทั่วไปแล้ว Backbone จะประกอบด้วยชั้น CNNs หลายชั้น ตามด้วย pooling layers และชั้นอื่นๆ ที่รับผิดชอบการประมวลผลภาพนำเข้าและแยกแยะลักษณะเด่นที่เกี่ยวข้องที่สามารถใช้สำหรับการตรวจจับวัตถุได้ ผลลัพธ์ของ Backbone คือเซ็ตของ feature maps ซึ่งจะถูกส่งเข้าไปยัง detection head ของอัลกอริทึมซึ่งรับผิดชอบการทำนาย Bounding boxes ล้อมรอบวัตถุและคลาสวัตถุ (Cheng, 2020) 2) Head เป็น เลเยอร์ Convolutional หลายชั้นที่ทำนายความน่าจะเป็นของคลาส มีหน้าที่ทำนายตำแหน่งและขนาดของวัตถุในภาพ



รูปที่ 2.2 สถาปัตยกรรม YOLOv8

หมายเหตุ จาก https://github.com/ultralytics/ultralytics/issues/189

จากรูปที่ 2.2 ฟังก์ชัน Concat ใช้สำหรับการเชื่อมต่อแผนที่ลักษณะจากชั้นคอนโวลูชันสุดท้ายและชั้นก่อนหน้าที่มีความละเอียดสูงกว่า เนื่องจากความละเอียดของแผนที่ลักษณะในชั้นนี้อาจจะต่ำเกินพอที่จะตรวจจับวัตถุขนาดเล็กอย่างแม่นยำ การใช้ฟังก์ชันนี้จะช่วยให้โครงข่ายสามารถใช้ข้อมูลจากทั้งสองชั้นเพื่อทำนายและเพิ่มความแม่นยำในการตรวจจับวัตถุขนาดเล็กได้ เป้าหมายของฟังก์ชัน Upsample ในโครงสร้าง YOLO คือเพื่อเพิ่มความละเอียดของแผนที่ลักษณะเชิงพื้นที่ (feature map) โดยทั่วไปแล้วจะทำการขยายขนาดก่อนที่จะนำ convolutional layers มาใช้กับ feature map เนื่องจากสามารถช่วยให้สามารถตรวจจับวัตถุได้ง่ายขึ้น โดยเฉพาะอย่างยิ่งในการตรวจจับวัตถุที่มีขนาดเล็ก

3. ผลการทดลอง

ปลานิลอายุ 28 สัปดาห์ จำนวน 30 ตัว ที่ถูกเลี้ยงด้วยวิธีไบโอฟล็อค ในบ่อต้นแบบมีรูปทรงกลมด้านผ่านศูนย์กลาง 1.5 เมตรและความลึก 1.8 เมตร วิธีการเพาะเลี้ยงเรียกว่าวิธีการเพาะเลี้ยงไบโอฟล็อค ไบโอฟล็อคเป็นวิธีการเลี้ยงปลาที่ใช้หลักการนำตะกอนจุลินทรีย์มาช่วยย่อยสลายของเสียในบ่อปลา (แอมโมเนีย) และใช้จุลินทรีย์ในการบำบัดน้ำ เปลี่ยนของเสียให้กลายเป็นของดีที่มีประโยชน์ต่อสัตว์น้ำ จึงเป็นสาเหตุทำให้น้ำในบ่อมีตะกอนและขุ่นมากกว่าวิธีการเลี้ยงปลานิลแบบทั่วไป การ

บันทึกภาพปลานิลในน้ำ จะทำการสุ่มปลานิลจำนวน 5 ตัว จากบ่อเลี้ยงมาชั่งน้ำหนักบนเครื่องชั่งดิจิตัลพร้อมวัดขนาดความกว้างและความยาวของตัวปลา ก่อนนำใส่ในตู้ปลาขนาด 45x80 เซ็นติเมตร ได้ทำการติดตั้งไฟ LED เพื่อส่องสว่างไว้ปลายตู้ด้านหนึ่ง และติดตั้งกล้องวีดีทัศน์สำหรับบันทึกภาพไว้บริเวณตรงกลางที่ปลายตู้อีกด้านหนึ่ง

กระบวนการพัฒนาระบบ Tilapia Image Segmentation ประกอบด้วย 4 ขั้นตอน ได้แก่

1) การสร้างชุดข้อมูลรูปภาพปลานิล (Tilapia image dataset): ชุดรูปภาพปลานิลใต้น้ำถูกบันทึกวิดีทัศน์ด้วยกล้อง GoPro Hero เวอร์ชัน 8 พร้อมเคสกันน้ำ ติดตั้งด้านในตู้สำหรับบันทึกภาพ วิดีทัศน์ถูกบันทึกที่ความละเอียด 1,920 x 1,080 พิกเซล อัตราเฟรม 60 เฟรมต่อวินาที รูปภาพปลามีความละเอียด 24 บิตต่อพิกเซล และมีระดับความเข้ม 256 ระดับ การบันทึกวิดีทัศน์แต่ละครั้งมีความยาว 10 นาที จากนั้นผ่านกระบวนการแปลงวิดีทัศน์เป็นรูปภาพ (video-to-image) เป็นการลดลงของสัญญาณเวลาต่อเนื่อง (continuous-time signal) ด้วยอัตรา 1 ภาพต่อ 1 วินาที หลังจากได้รูปภาพปลาแล้วจะทำกระบวนการ data cleaning โดยการคัดเลือกรูปที่เหมาะสมและลบรูปที่ตัวปลาอยู่ใกล้กล้องเกินไปหรือปลาอยู่ไกลจนเกินไปจนทำให้ไม่สามารถระบุตัวปลาได้ จากขั้นตอนนี้ได้ชุดรูปภาพปลานิลทั้งสิ้น 136 รูปภาพ

2) การอธิบายประกอบรูปภาพ (Image Annotation) : ขั้นตอนนี้เป็นการให้คำอธิบายประกอบข้อมูลรูปภาพ โดยการติดป้ายกำกับปลาที่อยู่ในรูปภาพด้วยตนเอง (manual) ขั้นตอนนี้เป็นกระบวนการที่ใช้เวลานาน แต่จำเป็นอย่างยิ่งที่จะต้องฝึกฝนโมเดลการเรียนรู้ของเครื่องให้แม่นยำ คำอธิบายประกอบรูปภาพมีโครงสร้างที่แตกต่างกันขึ้นอยู่กับวิธีการสร้างแบบจำลองที่เลือกใช้ สำหรับวิธี YOLOv8 เป็นไฟล์ txt มีโครงสร้างดังนี้ <object-class><x><y><width><height> โดย txt ไฟล์นี้จะถูกสร้างสำหรับแต่ละรูปภาพ ส่วนวิธี Mask R-CNN คำอธิบายประกอบรูปภาพเป็นไฟล์ json มีโครงดังต่อไปนี้ {"filename":{"filename":"...","size":int,"regions":[{"shape_attributes":{"name":"polygon","all_points_x":[...],"all_points_y":[...]},"region_attributes":{"name":"Tilapia","type":"tilapia"}}] โดยคำอธิบายประกอบรูปภาพทุกรูปจะถูกรวมอยู่ในไฟล์เดียวกัน

3) การฝึกอบรม (Training): การฝึกโมเดล Tilapia Image Segmentation ให้แบ่งกลุ่มพิกเซลปลานิลในรูปภาพอย่างแม่นยำ ด้วยวิธี YOLOv8 การทดสอบการแบ่งส่วนปลานิล (Tilapia Image Segmentation) ด้วยวิธี YOLOv8 แบบการถ่ายโอนการเรียนรู้ (Transfer learning) จากโมเดล yolov8m-seg ซึ่งโมเดลนี้ได้รับการฝึกฝนจากชุดรูปภาพ Common Objects in Context (COCO) เป็นชุดข้อมูลที่ใช้ในการตรวจจับวัตถุขนาดใหญ่ โดยเปรียบเทียบประสิทธิภาพของ Tilapia Image Segmentation ด้วยวิธี YOLOv8 กับ Tilapia Image Segmentation ด้วยวิธี Mask R-CNN แบบ

Transfer learning จากโมเดล COCO-Instance Segmentation ที่ได้รับการฝึกฝนในชุดข้อมูลการแบ่งกลุ่ม COCO เช่นกัน

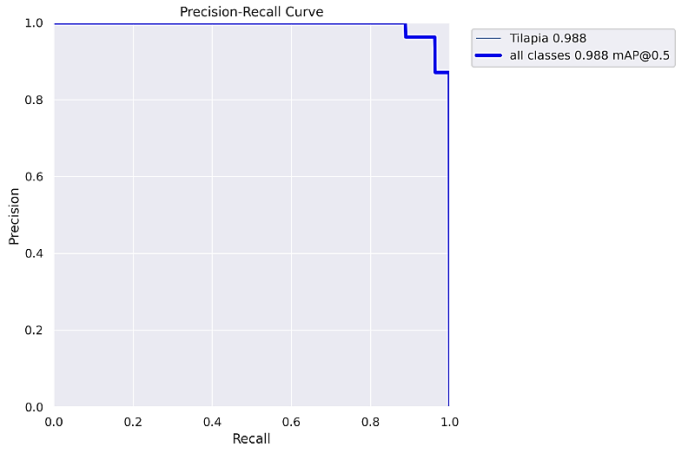
4) กระบวนการทดสอบ Tilapia Image Segmentation เริ่มจากการฝึกฝนโมเดลด้วยวิธี YOLOv8 และ Mask R-CNN โดยใช้ชุดรูปภาพปลานิลอายุ 28 สัปดาห์ในน้ำขุ่นจำนวน 136 รูป แบ่งออกเป็น 3 ส่วนได้แก่ ส่วนแรกสำหรับการเรียนรู้ของโมเดล (training) จำนวน 70% ของรูปภาพทั้งหมด ส่วนที่สองรูปภาพสำหรับตรวจสอบ (validation) การเรียนรู้ของโมเดลจำนวน 20% ของรูปภาพทั้งหมด และรูปภาพสำหรับทดสอบ (test) จำนวน 10% ของรูปภาพ การตั้งค่าตัวแปรในการฝึกฝนของวิธี YOLOv8 จำนวนการเรียนรู้ (epoch) 1,000 รอบ โมเดลยุติการเรียนรู้ที่ 120 รอบ เนื่องจากผลการเรียนรู้ไม่เปลี่ยนแปลง ส่วนวิธี Mask R-CNN ได้กำหนดค่าตัวแปรการเรียนรู้ของโมเดล ดังต่อไปนี้ จำนวนการเรียนรู้ (epoch) 1,000 รอบ อัตราการเรียนรู้ (learning rate) 0.0001 จำนวนแพตช์รูปภาพ (batch size) จำนวน 64 รูป

5) การตรวจสอบความถูกต้องและการทดสอบ (Validation and Testing): เมื่อแบบจำลองได้รับการฝึกฝนแล้ว จะมีการตรวจสอบความถูกต้องและทดสอบโดยใช้ชุดรูปภาพที่แยกต่างหากที่ไม่ได้ใช้ระหว่างการฝึกอบรม สิ่งนี้ช่วยในการประเมินประสิทธิภาพของแบบจำลอง การวัดประสิทธิภาพการแบ่งส่วนปลานิลด้วยวิธี Mean Average Precision (mAP) นำเสนอโดยเป็นค่าแสดงประสิทธิภาพในการตรวจหาวัตถุและการจำแนกภาพ โดยจะวัดความแม่นยำของอัลกอริทึมในการตรวจหาวัตถุภายในภาพ (Lin, 2018) และความสามารถในการระบุตำแหน่งของวัตถุ การคำนวณค่า mAP จะเริ่มต้นด้วยการคำนวณ Average Precision (AP) สำหรับแต่ละคลาสวัตถุในชุดข้อมูล งานวิจัยนี้ได้กำหนดคลาสเพียงหนึ่งคลาสได้แก่ คลาสของปลานิล ซึ่ง AP เป็นตัววัดความแม่นยำแบบ tradeoff ระหว่าง precision และ recall คะแนน mAP จะอยู่ในช่วง 0 - 1 โดยคะแนน 1 จะหมายถึงการทำงานที่สมบูรณ์แบบที่สุด โดยทั่วไปแล้ว ค่า mAP ที่สูงกว่านั้น แสดงให้เห็นถึงประสิทธิภาพของโมเดลตรวจจับวัตถุที่ดีขึ้นในแง่ความแม่นยำและความครอบคลุมในช่วงค่า Intersection over Union (IoU) ที่กว้างขึ้น โดยสรุปผลแล้ว มีค่า mAP ที่สูงขึ้นนั้นแสดงให้เห็นถึงประสิทธิภาพของโมเดลตรวจจับวัตถุที่ดีขึ้นในแง่ความแม่นยำและความครอบคลุมของการตรวจจับในช่วงค่า IoU ที่กว้างขึ้นเช่นกัน ประสิทธิภาพผลการเรียนรู้ของแบบจำลองที่เรียนรู้ด้วย วิธี Mask R-CNN และ วิธี YOLOv8 ในตารางที่ 3.1

ตารางที่ 3.1 เปรียบเทียบประสิทธิภาพการเรียนรู้ของวิธี Mask R-CNN และ YOLOv8 ด้วยค่า mAP

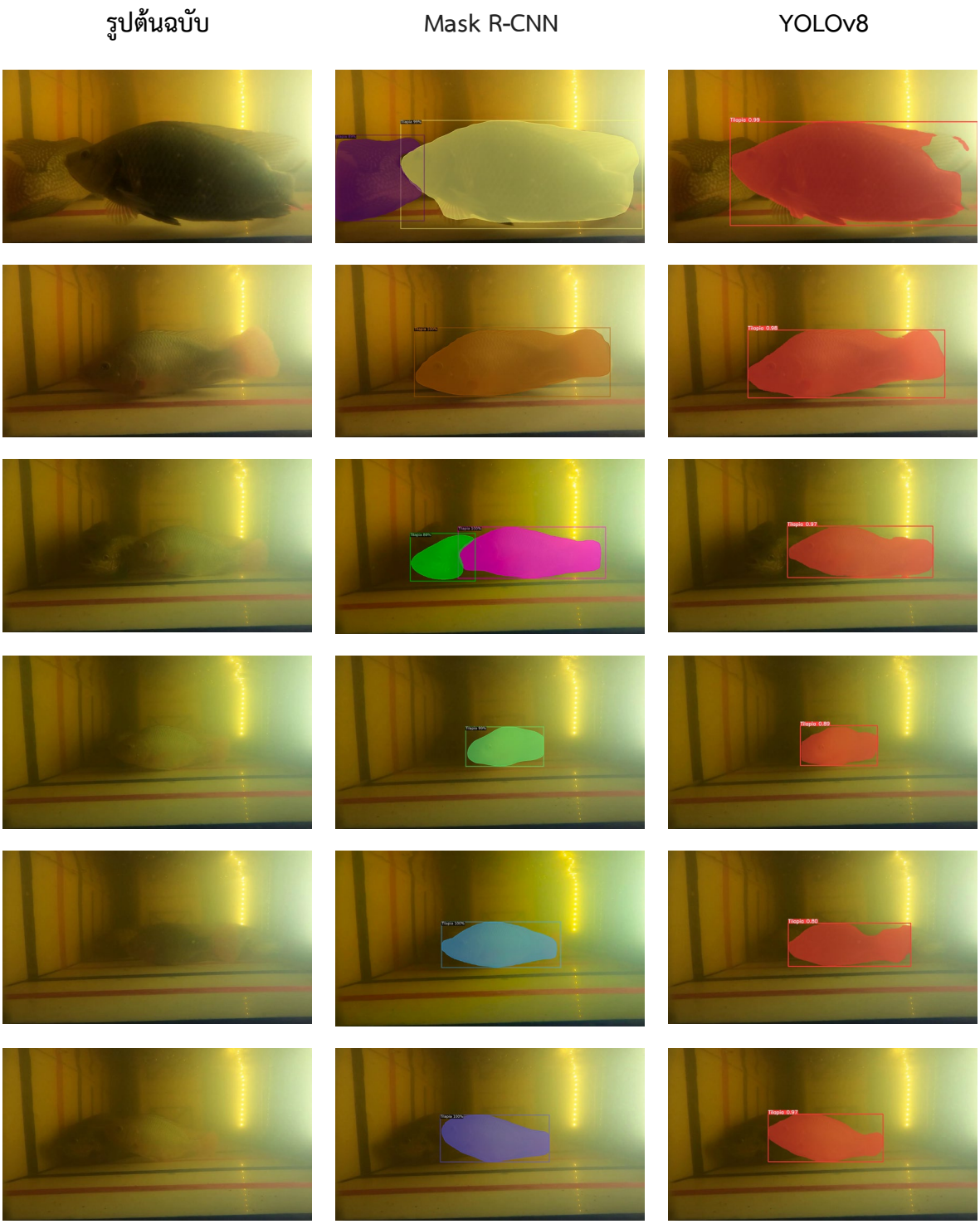
วิธี\ ค่า mAP	mAP50	mAP50-95	mAP
Mask R-CNN based Tilapia	0.94	0.73	0.84
YOLOv8 based Tilapia	0.995	0.77	0.88

แสดงค่า mAP50 (ค่าเฉลี่ยความแม่นยำเฉลี่ยที่ความถี่เกินที่ 50%) เป็นการวัดประสิทธิภาพของโมเดลตรวจจับวัตถุโดยวัดความแม่นยำเฉลี่ยที่ช่วงค่า IoU ที่ 50% โดยที่ IoU เป็นการวัดความซ้อนทับกันระหว่างกรอบสี่เหลี่ยมที่ทำนายได้กับกรอบสี่เหลี่ยมตามจริงของวัตถุ หากค่า IoU ถูกตั้งค่าไว้ที่ 50% หมายความว่ากรอบสี่เหลี่ยมที่ทำนายได้ถูกพิจารณาว่าเป็นการตรวจจับที่ถูกต้องถ้ามีการซ้อนทับกับกรอบสี่เหลี่ยมตามจริงอย่างน้อย 50% และ mAP50-95 เป็นการคำนวณค่าความแม่นยำเฉลี่ยสำหรับโมเดลตรวจจับวัตถุในช่วงค่า IoU หลายๆ ระดับตั้งแต่ 0.50 ถึง 0.95 โดยมีขนาดขั้นตอนที่เท่ากับ 0.05 ช่วงค่า IoU กว้างขึ้นและเป็นตัวชี้วัดที่ครอบคลุมมากขึ้นสำหรับการประเมินประสิทธิภาพของโมเดลตรวจจับวัตถุ มันช่วยให้เข้าใจได้ดียิ่งขึ้นว่าโมเดลทำงานได้ดีแค่ไหนในระดับการซ้อนทับกรอบสี่เหลี่ยมที่ทำนายได้และกรอบสี่เหลี่ยมตามจริง และสุดท้ายได้แก่ ค่า mAP เป็นค่าเฉลี่ยของค่า Average Precision (AP) ของทุกคลาสในโมเดล 2 mAP สามารถใช้เปรียบเทียบประสิทธิภาพของโมเดล จากตารางที่ 3.1 แบบจำลองการแบ่งกลุ่มพิกเซลด้วยวิธี YOLOv8 มีค่าเฉลี่ยความแม่นยำสูงกว่าวิธี Mask R-CNN ทั้งแบบ mAP50 mAP50-95 และ mAP โดยมีประสิทธิภาพแม่นยำกว่าคิดเป็นร้อยละ 5.36 ประสิทธิภาพแบบจำลองการแบ่งส่วนภาพปลานิลด้วยวิธี YOLOv8 แสดงในรูปแบบเส้นโค้ง Precision-Recall (PR) เป็นความสัมพันธ์ระหว่าง Precision และ Recall สำหรับค่าเขตของการจัดหมวดหมู่ข้อมูลแบบไบนารีที่แตกต่างกัน ในที่นี้คือการจัดกลุ่มพิกเซลปลานิลและพิกเซลที่ไม่ใช่ปลานิล PR curve เป็นตัวชี้วัดความแม่นยำ (Precision) และความครอบคลุม (Recall) ที่ใช้ประเมินประสิทธิภาพของโมเดลการจำแนก จากรูปที่ 3.1 แสดงพื้นที่ใต้กราฟ (Area Under the Curve: AUC) ที่ 0.988 แสดงว่าแบบจำลองการแบ่งกลุ่มพิกเซลด้วยวิธี YOLOv8 มีความแม่นยำสูงในการระบุพิกเซลของปลานิล



รูปที่ 3.1 เส้นโค้ง Precision-Recall (PR) เป็นความสัมพันธ์ระหว่าง Precision และ Recall ของแบบจำลองการแบ่งส่วนภาพปลานิลด้วยวิธี YOLOv8

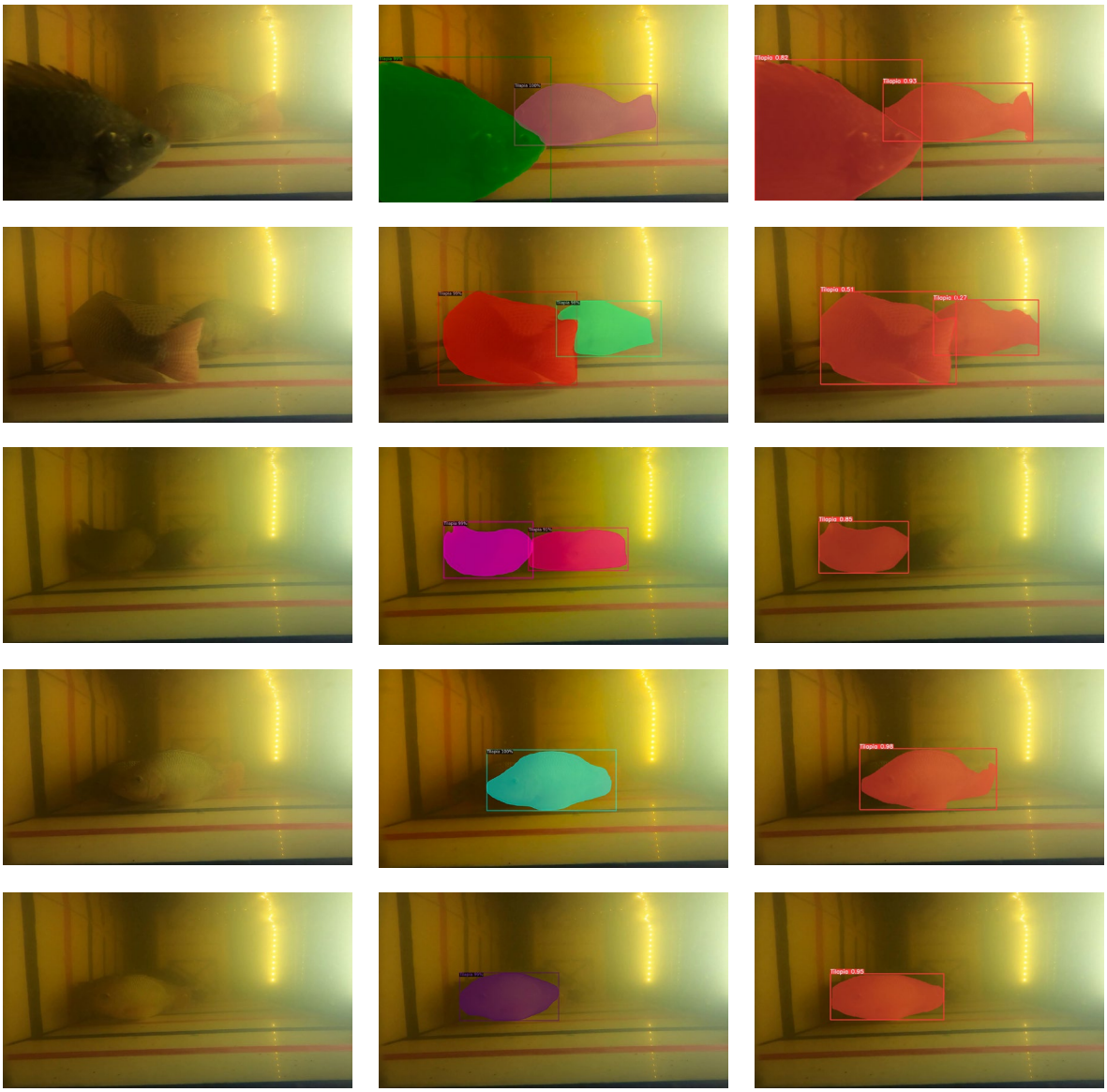
การทดสอบ Tilapia Image Segmentation ด้วยวิธี Mask R-CNN และ YOLOv8 ทำการทดสอบกับรูปภาพปลานิลสำหรับทดสอบจำนวน 13 รูป ที่ได้จากการสุ่มเลือกรูปภาพจากชุดข้อมูลรูปภาพ ชุดรูปภาพทดสอบแบ่งออกเป็น 2 กลุ่มได้แก่ กลุ่มแรกภายในรูปภาพแสดงปลานิล 1 ตัวในลักษณะด้านข้างลำตัวอยู่ในแนวขวางขนานกับความกว้างของตู้ปลาในระยะห่างจากกล้องต่างๆ ผลการทดสอบแบบจำลองจากวิธี Mask R-CNN และ YOLOv8 กับรูปภาพปลานิลกลุ่มที่หนึ่งและกลุ่มที่สอง แสดงรูปตัวอย่างในรูปที่ 3.2 และ 3.3 ตามลำดับ



รูปที่ 3.2 เปรียบเทียบการแบ่งกลุ่มปลานิลของรูปปลานิลด้านข้างลำตัวขนานกล้อง 1 ตัวในระยะต่างๆ ด้วยวิธี Mask R-CNN และ YOLOv8

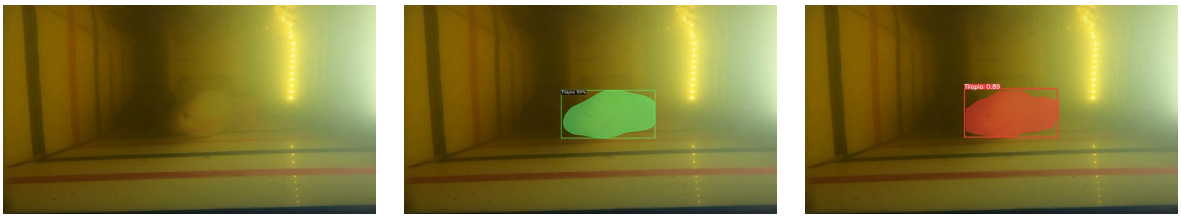
จากรูปที่ 3.2 ทั้งวิธี Mask R-CNN และ YOLOv8 สามารถแบ่งกลุ่มพิกเซลปลานิลด้านข้าง ทั้งในกรณีที่ปลาอยู่ใกล้และอยู่ไกลได้ สำหรับกรณีที่ปลาอยู่ใกล้กล้องบันทึกภาพตัวปลาจะมีขนาดใหญ่ ดังแสดงในรูปแถวที่ 1 ถึงแถวที่ 3 กรณีนี้วิธี Mask R-CNN สามารถระบุพิกเซลตัวปลานิลได้สมบูรณ์ และสามารถระบุพิกเซลปลานิลตัวที่เห็นเพียงบางส่วนของลำตัวได้กว่าวิธี YOLOv8 สำหรับกรณีที่ปลาอยู่ไกลจากกล้อง ตัวปลาจะมีขนาดเล็กและไม่ชัดเจน นอกจากนี้ความขุ่นของน้ำยังส่งผลต่อความชัดเจนของตัวปลา โดยเฉพาะบริเวณหางปลาที่มีสีอ่อนกว่าลำตัว ทำให้สังเกตเห็นได้ยากในระยะไกล กรณีนี้วิธี YOLOv8 สามารถแบ่งกลุ่มปลานิลได้สมบูรณ์กว่าวิธี Mask R-CNN ซึ่งไม่สามารถแบ่งกลุ่มบริเวณหางของปลาได้

รูปต้นฉบับ Mask R-CNN YOLOv8



รูปที่ 3.3 เปรียบเทียบการแบ่งกลุ่มปลานิลของรูปภาพปลานิลในทิศทางและระยะต่างๆ ด้วยวิธี Mask R-CNN และ YOLOv8

รูปต้นฉบับ Mask R-CNN YOLOv8



รูปที่ 3.3 เปรียบเทียบการแบ่งกลุ่มปลานิลของรูปภาพปลานิลในทิศทางและระยะต่างๆ ด้วยวิธี Mask R-CNN และ YOLOv8 (ต่อ)

จากรูปที่ 3.3 จะเห็นได้ว่าประสิทธิภาพในการแบ่งปลานิลในทิศทางและระยะที่แตกต่างกันของทั้ง 2 วิธีให้ผลสอดคล้องกับผลที่ได้ในรูปภาพกลุ่มแรก

4. การวิเคราะห์ผลการทดลอง

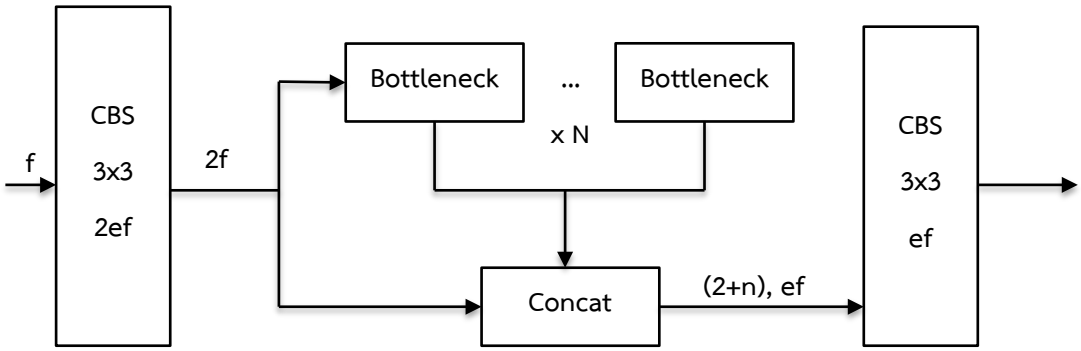
การระบุพิกเซลปลานิลจากรูปภาพปลาในน้ำขุ่น การแบ่งส่วนภาพปลานิลด้วยวิธี Mask R-CNN และวิธี YOLOv8 ทั้งสองวิธีมีโครงสร้างแบบ two-stage ซึ่งจะทำการตรวจจับ bounding boxes ของวัตถุ ก่อนทำการแบ่งกลุ่มพิกเซลตามอัลกอริทึมของแต่ละวิธี การสร้างแบบจำลอง Tilapia Image Segmentation นั้นวิธี YOLOv8 มีโครงสร้างเบากว่าวิธี Mask R-CNN โดยมีจำนวน parameters และ GFLOPs น้อยกว่าวิธี Mask R-CNN ถึงร้อยละ 48.97 และ 54.57 ตามลำดับ ดังแสดงจำนวนตัวแปรของโครงสร้างโมเดลของทั้ง 2 วิธี ในตารางที่ 4.1 นอกจากนี้วิธี YOLOv8 ไม่มีขั้นตอนการสร้าง masks ระดับพิกเซลสำหรับแต่ละวัตถุ ทำให้ใช้เวลาในการ Training น้อยกว่าวิธี Mask R-CNN

ตารางที่ 4.1 จำนวนตัวแปรของโครงสร้างโมเดลของ วิธี Mask R-CNN และวิธี YOLOv8

วิธี\ จำนวน	layers	parameters (ล้าน)	GFLOPs
Mask R-CNN	238	53.3	243
YOLOv8	331	27.2	110.4

จากชุดข้อมูลรูปภาพปลานิลจำนวน 136 รูป นำมาสร้างแบบจำลองการแบ่งส่วนภาพปลานิล วิธี YOLOv8 สามารถเรียนรู้ได้อย่างมีประสิทธิภาพกว่า วิธี Mask R-CNN ด้วยปัจจัยสำคัญได้แก่ ฟังก์ชันในการทำ Feature Extraction ที่มีประสิทธิภาพด้วยวิธี C2f ทำให้แบบจำลองสามารถจดจำลักษณะเด่นของปลานิลได้ในขณะที่วิธี Mask R-CNN ใช้วิธี convolution มาตรฐานร่วมกับการคำนวณค่า

Residual Node ซึ่งต้องการข้อมูลจำนวนมากในการเรียนรู้เพื่อให้ได้ประสิทธิภาพที่ดี YOLOv8 ได้พัฒนาโมดูล Convolution ขึ้นใหม่ ได้แก่ Cross-Stage Partial Network 2 Features (C2F) แทนโมดูล Cross-Stage Partial Network (C3) ในสถาปัตยกรรม YOLOv5 โดยโมดูล C2F เอาต์พุตทั้งหมดจากส่วน Bottleneck เพื่อสกัดคุณลักษณะที่มีความหมายมากขึ้นและเพิ่มความแม่นยำในการตรวจจับวัตถุของโมเดล



รูปที่ 4.1 โมดูล C2F ของ YOLOv8

หมายเหตุ จาก https://github.com/ultralytics/ultralytics/issues/189

จากรูปที่ 4.1 CBS ประกอบด้วย 3 ชั้น ได้แก่ Conv, BatchNorm, SiLU ชั้น Convolutional Layer (Conv): ชั้นนี้ทำการดำเนินการ Convolution บนภาพนำเข้าหรือ Feature map ซึ่งเป็นเมทริกซ์ของตัวเลขและนำชุดของตัวกรองที่เรียนรู้ได้มาใช้เพื่อสร้างชุดของ Output feature maps โดยตัวกรองถูกใช้เพื่อตรวจจับลักษณะต่างๆในภาพ เช่นขอบ เหลี่ยมมุม และลายเส้น โดย Output feature maps นี้จะนำไปใช้ต่อในการตรวจจับวัตถุ ชั้น Batch Normalization Layer (BatchNorm): ชั้นนี้ทำการปรับค่า Output ของ Convolutional layer โดยการลบค่าเฉลี่ยและหารด้วยค่าเบี่ยงเบนมาตรฐาน ซึ่งจะช่วยเสถียรภาพกระบวนการเรียนรู้และปรับปรุงประสิทธิภาพของเครือข่าย (Loffe, 2015) ชั้น Sigmoid Linear Unit (SiLU): ชั้นนี้นำฟังก์ชันการกระตุ้น SiLU มาใช้กับ Output ของชั้น Batch Normalization Layer เพื่อให้ได้ Output ที่ไม่เชิงเส้น ซึ่งจะช่วยในการแก้ไขปัญหาของ "dead neurons" ซึ่งเกิดจากการใช้ฟังก์ชัน ReLU โดยชั้นนี้จะช่วยปรับปรุงประสิทธิภาพของเครือข่ายได้ดีขึ้น (Elfwing, 2018) การใช้งาน CBS ในโครงสร้าง YOLO ช่วยปรับปรุงความแม่นยำและความเร็วในการตรวจจับวัตถุ ชั้น Conv จะตรวจจับลักษณะต่างๆในภาพ ชั้น BatchNorm จะช่วยเสถียรภาพกระบวนการเรียนรู้ และชั้น SiLU จะนำเสนอความไม่เชิงเส้นเข้าไปในเครือข่าย รวมกันเป็นชุดสร้างเสริมกันที่สามารถใช้สร้างเครือข่ายที่ซับซ้อนมากขึ้นได้

Bottleneck จะลดความละเอียดเชิงพื้นที่ (Spatial Resolution) ของ Feature map ของรูปภาพ เป็นการลดจำนวนข้อมูลในการประมวลผลส่งผลให้ YOLOv8 ทำงานได้อย่างรวดเร็ว ส่วน Bottleneck นี้ประกอบด้วยลำดับชั้นของ CNNs สามารถแสดงด้วยสมการทางคณิตศาสตร์ดังแสดงในสมการที่ (1)

$$y = \sigma(W * x + b) \tag{1}$$

โดยตัวแปร x และตัวแปร y คือ Feature Map ที่รับเข้าเป็นอินพุตและผลลัพธ์ Feature Map ที่ได้จากฟังก์ชัน ตามลำดับ ตัวแปร W และ ตัวแปร b เป็นค่าถ่วงน้ำหนักของเทนเซอร์และค่าอคติของเทนเซอร์ ตัวแปร σ คือฟังก์ชัน Activation สุดท้ายเครื่องหมาย $*$ เป็นตัวดำเนินการ Convolution

นอกจากนี้ในส่วน Head เป็นการเปลี่ยนแปลงที่สำคัญที่สุดของโมเดล YOLOv8 โดยโมดูล Head ได้ถูกเปลี่ยนจากโครงสร้างการผูกติด (Anchor boxes) ของ YOLOv5 เดิมเป็นโครงสร้างการไม่ผูกติด (Anchor free) ในส่วน Head ของ YOLOv8 เป็นโมดูลที่ไม่มีแองค์เกอร์ (Anchor-Free) โมดูลจะทำการทำนายจุดศูนย์กลางของวัตถุโดยตรง แทนที่จะใช้การเปรียบเทียบระยะห่างจาก Anchor box การตรวจจับวัตถุโดยใช้วิธี Anchor-free นั้นจะไม่ใช้ Anchor boxes เพื่อกำหนดตำแหน่งของวัตถุ แต่จะใช้ Convolutional Neural Network (CNN) เพื่อค้นหา object center และ object boundary โดยตรง การตรวจจับด้วยวิธี Anchor free จะช่วยลดจำนวนการทำนายกล่องซึ่งจะทำให้การดำเนินการ Non-Maximum Suppression (NMS) เร็วขึ้น ซึ่ง NMS ทำหน้าที่ลดความซ้อนทับของกรอบสี่เหลี่ยมที่พบในขั้นตอนการตรวจจับวัตถุ เพื่อให้ได้กรอบสี่เหลี่ยมที่ไม่ซ้อนทับและไม่ซ้ำกัน ซึ่งสามารถช่วยลดการตรวจจับวัตถุซ้ำซ้อนได้ โดยเริ่มจากการเลือกกรอบสี่เหลี่ยมที่มีค่าความน่าจะเป็นสูงสุด (confidence score) จากนั้นตรวจสอบว่ากรอบสี่เหลี่ยมอื่นๆ ที่มีความซ้อนทับกับกรอบสี่เหลี่ยมที่เลือกนั้นเกินกำหนดหรือไม่ และลบกรอบสี่เหลี่ยมที่ซ้อนทับออกไปจนกว่าจะเหลือกรอบสี่เหลี่ยมที่ไม่ซ้ำกัน ซึ่งจะช่วยลด false positive และช่วยเพิ่มประสิทธิภาพในการตรวจจับวัตถุได้

ปัจจัยที่ 2 ที่ทำให้ YOLOv8 มีความแม่นยำสูงกว่า Mask R-CNN ได้แก่ การใช้ Focal Loss ซึ่งฟังก์ชันนี้คำนวณค่าความน่าจะเป็นของพิกเซลของแต่ละกลุ่มและกำหนดให้อยู่ในกลุ่มที่เหมาะสมที่สุดตามค่าถ่วงน้ำหนักจากการเรียนรู้ข้อมูลรูปภาพ ในขณะที่ Mask R-CNN ใช้การสร้างตัวกรองแบบไบนารี่ (Binary mask) สำหรับการแบ่งกลุ่มพิกเซล ส่งผลให้บริเวณหางปลาที่มีความขุ่นมัว Mask ไม่สามารถแยกออกได้ ฟังก์ชันการสูญเสียใหม่ (Loss Function) ของวิธี YOLOv8 ใช้ฟังก์ชัน Focal Loss สำหรับคำนวณค่าการสูญเสีย ฟังก์ชัน Focal Loss นี้สามารถเรียนรู้ชุดข้อมูลรูปภาพที่ไม่สมดุลได้ (Imbalance image dataset) โดยจะให้ค่าถ่วงน้ำหนักกับตัวอย่างยากๆ มากขึ้น และลดอิทธิพลของตัวอย่างง่ายๆ การคำนวณค่า Focal Loss แสดงในสมการที่ (2)

$$FL(p_t) = -(1 - p_t) \wedge \gamma * log(p_t) \tag{2}$$

โดยตัวแปร p_t คือความน่าจะเป็นที่ทำนายได้ว่าเป็นคลาสจริง ๆ ซึ่งอยู่ระหว่าง 0 ถึง 1 ส่วน γ เป็นพารามิเตอร์ที่สามารถปรับได้ เพื่อควบคุมการลดน้ำหนัก โดยเมื่อ γ ถูกกำหนดเป็น 0 แล้ว Focal Loss จะเทียบเท่ากับค่า Cross-entropy loss มาตรฐาน แต่เมื่อเพิ่ม γ ขึ้น ค่าถ่วงน้ำหนักของตัวอย่างที่ง่ายจะลดลง เพื่อให้ความสำคัญกับตัวอย่างที่ยาก ดังนั้น Focal Loss จะกำหนดค่าถ่วงน้ำหนักต่ำกว่าให้กับตัวอย่างที่ง่ายๆ หากมีปัจจัยเพิ่มขึ้นส่งผลให้ค่า p_t ลดลงแสดงว่าตัวอย่างนั้นจะเป็นตัวอย่างที่ยากในการจำแนก โดยการลดค่าถ่วงน้ำหนักตัวอย่างที่ง่าย Focal Loss สามารถจัดการความไม่สมดุลของคลาสได้อย่างมีประสิทธิภาพ ปรับปรุงความแม่นยำของโมเดลในตัวอย่างที่ยากและทำให้การตรวจจับวัตถุมีความถูกต้องยิ่งขึ้น

5. สรุปผล

บทความวิจัยนี้ได้นำเสนอ การสร้างแบบจำลอง Tilapia Image Segmentation เพื่อแบ่งกลุ่มพิกเซลปลานิล ด้วยวิธี YOLOv8 ประกอบด้วย 4 ขั้นตอน ได้แก่ 1) สร้างชุดรูปภาพปลานิลในน้ำโดยใช้กล้อง GoPro Hero 8 พร้อมเคสกันน้ำ ได้ชุดรูปภาพปลานิลในน้ำที่ผ่านการคัดเลือกจำนวน 136 รูป 2) การทำ Image Annotation ของชุดรูปภาพปลาทั้งหมดสำหรับวิธี YOLOv8 และวิธี Mask R-CNN 3) การฝึกฝนแบบจำลอง Tilapia Image Segmentation ด้วยวิธี Deep Transfer Learning เป็นการนำโมเดลที่ผ่านการฝึกฝนเบื้องต้นมาทำการฝึกฝนต่อ วิธีการนี้ช่วยให้ลดระยะเวลาในการเรียนรู้ของแบบจำลอง จากการทดลอง YOLOv8 ใช้จำนวนรอบในการฝึกฝนเพียง 137 รอบ และขั้นตอนสุดท้าย 4) การทดสอบประสิทธิภาพของแบบจำลอง Tilapia Image Segmentation ที่สร้างขึ้นด้วยวิธี YOLOv8 และ Mask R-CNN พบว่า YOLOv8 มีความแม่นยำในการแบ่งส่วนภาพปลานิล ทั้งระยะใกล้และไกลได้ดีกว่า Mask R-CNN ถึงร้อยละ 6 ของค่า mAP

วิธีการการสร้างแบบจำลอง Tilapia Image Segmentation ด้วยวิธี YOLOv8 เป็นการระบุพิกเซลของปลานิลที่ว่ายอยู่ในน้ำขุ่น มีจุดเด่นดังต่อไปนี้ 1) วิธีการที่นำเสนอเพื่อระบุพิกเซลของปลานิลจากรูปภาพที่มีความขุ่นของน้ำและความไม่ชัดเจนของตัวปลาเนื่องจากการเคลื่อนไหวของปลา วิธีการนี้ไม่มีกระบวนการรปรับปรุงคุณภาพของรูปภาพให้ชัดเจนจึงลดความซับซ้อนในการพัฒนา แต่ยังสามารถระบุพิกเซลของปลานิลได้อย่างแม่นยำ 2) วิธีการที่นำเสนอใช้ชุดรูปภาพปลาขนาดเล็กทำให้ลดต้นทุนการจัดเตรียมข้อมูลและ ลดระยะเวลาการฝึกฝนเรียนรู้ของแบบจำลอง และ 3) อุปกรณ์สำหรับบันทึกภาพปลาในน้ำเป็นอุปกรณ์ที่สามารถหาซื้อได้ทั่วไปการติดตั้งไม่ซับซ้อน

แนวทางการพัฒนาต่อยอด จาก Tilapia Image Segmentation การแบ่งกลุ่มพิกเซลเพื่อระบุขนาดตัวปลานิลได้แล้ว จากนั้นนำข้อมูลจำนวนพิกเซลที่ได้มาใช้สำหรับคำนวณน้ำหนักของปลานิลแต่ละตัว เพื่อใช้สำหรับติดตามการเจริญเติบโต และเป็นข้อมูลสำหรับการให้อาหารอย่างเหมาะสม

กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับการสนับสนุนทางการเงินจากสำนักงานวิทยาศาสตร์วิจัยและนวัตกรรมแห่งชาติ (สวทน.) ได้รับการสนับสนุนด้านเทคนิคจากศูนย์วิจัยเทคโนโลยีพลังงานเพื่อสิ่งแวดล้อม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ ประเทศไทย

เอกสารอ้างอิง

กลุ่มสถิติการประมง กองนโยบายและแผนพัฒนาการประมง. (2565). **ประมาณการผลผลิตและมูลค่าสัตว์น้ำจากการประมงของประเทศไทย พ.ศ. 2565 - 2568**. [ระบบออนไลน์]. แหล่งที่มา https://www4.fisheries.go.th/local/file_document/20230322114403_1_file.pdf

Cheng, R., Zhang, C., Xu, Q., Liu, G., Song, Y., Yuan, X., & Sun, J. (2020). **Underwater Fish Body Length Estimation Based on Binocular Image Processing**. Information, 11, 476. doi: 10.3390/info11100476

Elfwing, S., Uchibe, E., & Doya, K. (2018). **Sigmoid-weighted linear units for neural network function approximation in reinforcement learning**. Neural Networks, 107, 3-11.

Hong Khai, T.; Abdullah, S.N.H.S.; Hasan, M.K.; Tarmizi, A. **Underwater Fish Detection and Counting Using Mask Regional Convolutional Neural Network**. Water 2022, 14, 222.

Huang, A., Jiang, L., Zhang, J., & Wang, Q. (2022). **Attention-VGG16-UNet: a novel deep learning approach for automatic segmentation of the median nerve in ultrasound images**. Quantitative Imaging In Medicine And Surgery, 12(6), 3138-3150.

Li, J., Chen, L., Shen, J., Xiao, X., Liu, X., Sun, X., Wang, X., & Li, D. (2023). **Improved neural network with spatial pyramid pooling and online datasets preprocessing for underwater target detection based on side scan sonar imagery**. Remote Sensing, 15(2), 440.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. In 2017 IEEE International Conference on Computer Vision (ICCV) (pp. 2999-3007). Venice, Italy: IEEE. https://doi.org/10.1109/ICCV.2017.324.

Loffe, S., & Szegedy, C. (2015). **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**. Proceedings of the 32nd International Conference on Machine Learning (pp. 37:448-456). Available from https://proceedings.mlr.press/v37/ioffe15.html.

Tengtrairat, N.; Woo, W.L.; Parathai, P.; Rinchumphu, D.; Chaichana, C. **Non-Intrusive Fish Weight Estimation in Turbid Water Using Deep Learning and Regression Models**. Sensors 2022, 22, 5161.

Wageeh, Y., Mohamed, H.ED., Fadl, A., El-Sayed, H.M., & El-Mahdy, M. (2021). **YOLO fish detection with Euclidean tracking in fish farms. Journal of Ambient Intelligence and Humanized Computing**, 12, 5-12.

Ultralytics. (2023). **Brief summary of YOLOv8 model structure**. [ระบบออนไลน์]. แหล่งที่มา https://github.com/ultralytics/ultralytics/issues/189.