บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็นการศึกษาและพัฒนาระบบการทำนายและป้องกันการชนกันของหุ่นยนต์ส่งของอัตโนมัติกับรถที่ไม่มีระบบช่วยเหลือหรือเซนเซอร์ใด ๆ ที่บริเวรทางแยกโดยใช้กล้องวงจรปิดเพื่อจับภาพและเปลี่ยนมุมมองไปยังมุมมองแบบ Bird Eye View เพื่อสร้างกริดอ้างอิงตำแหน่งจากนั้น ใช้ YOLO V8 ตรวจจับตำแหน่งรถยนต์บนกริดนั้นและส่งข้อมูลไปยังระบบ DeepSort เพื่อติดตามการเคลื่อนไหว ระบบจะแปลงตำแหน่ง X,Y จาก bounding box ให้อยู่ในระบบพิกัดเดียวกับหุ่นยนต์ส่งของอัตโนมัติผ่านการเปลี่ยนแปลงเชิงโฮโมจีเนียส และใช้เครือข่ายประสาทเทียม (ANN) ในการทำนายตำแหน่งถัดไปของรถที่ตรวจจับโดยกล้องวงจรปิดและหุ่นยนต์ส่งของอัตโนมัติเมื่อระบบทำนายตำแหน่งถัดไปและเปรียบเทียบระหว่างรถยนต์กับหุ่นยนต์ หากพบว่ามีความเป็นไปได้ที่จะเกิดการชนกัน ระบบการชะลอความเร็วเพื่อทำการหลีกเลี่ยงการชนนั้น ผู้วิจัยได้ทำการศึกษาค้นคว้าวรรณกรรมและงานวิจัยต่าง ๆ ที่เกี่ยวข้องเพื่อใช้ในการวิจัยและพัฒนาโดยในหัวข้อต่อไปนี้

2.1 การแปลงมุมมองภาพ

การเปลี่ยนมุมมองภาพ หรือ Perspective Transformation, เป็นกระบวนการที่ซับซ้อนและมีความสำคัญในการเปลี่ยนแปลงและสร้างมุมมองใหม่ของภาพที่ถูกจับจากมุมมองปกติเป็นมุมมองแบบภาพมุมสูง (Bird's eye view) (Li, H., Sima, C., Dai, J., Wang, W., Lu, L., Wang, H., ... & Qiao, Y. 2023). กระบวนการนี้ไม่เพียงแต่เปลี่ยนแปลงตำแหน่งของพิกเซลภายในภาพ แต่ยังรวมถึงการปรับมุมมองและขนาดของสิ่งที่ปรากฏในภาพเพื่อสร้างภาพที่มีมิติและลึกซึ้ง การประยุกต์ใช้ของกระบวนการนี้หลากหลาย ตั้งแต่การวางแผนเมือง การวิเคราะห์พื้นที่ การช่วยเหลือการขับขี่ในยานยนต์ ไปจนถึงการใช้งานในด้านกีฬาและบันเทิงในการทำให้ภาพมีมุมมองแบบภาพมุมสูงจำเป็นต้องมีการคำนวณทางคณิตศาสตร์อย่างละเอียด โดยใช้เทคนิคต่าง ๆ เช่น homography และการแปลงเมทริกซ์ ซึ่งช่วยในการปรับแต่งมุมมองและขนาดของภาพให้เป็นไปตามต้องการ กระบวนการนี้ยังต้องการความเข้าใจในหลักการของแสงและเงา เพื่อสร้างภาพที่สมจริงและเปี่ยมไปด้วยมิติผลลัพธ์ที่ได้จากการเปลี่ยนมุมมองแบบภาพมุมสูงนั้นไม่เพียงแต่เป็นการแสดงภาพที่กว้างขวางและลึกซึ้งเท่านั้น แต่ยังเปิดโอกาสให้ได้เห็นโลกในมุมที่แตกต่างจากมุมมองปกติ นำเสนอมุมมองที่ไม่ธรรมดาและมีความโดดเด่น ซึ่งสามารถนำไปใช้

ในหลากหลายด้าน ทั้งในงานวิจัย การพัฒนาเทคโนโลยี และการสร้างสรรค์ผลงานทางศิลปะและการออกแบบ ในทางปฏิบัติ การเปลี่ยนมุมมองภาพเป็นเครื่องมือที่มีค่าในการมองเห็นและเข้าใจโลกในมิติที่กว้างขึ้น และช่วยเพิ่มความสามารถในการวิเคราะห์และตีความข้อมูลที่มีอยู่ในภาพ โดยเฉพาะอย่างยิ่งในการวิเคราะห์ภาพที่ต้องการความแม่นยำสูงและมุมมองที่ครอบคลุม

2.1.1 Homography

Homography (Hartley, R., & Zisserman, A. 2003) คือ กระบวนการ ทางคณิตศาสตร์ที่ใช้ในการแปลงภาพสองมิติ ซึ่งช่วยให้สามารถเปลี่ยนภาพจากมุมมองหนึ่งไปยังมุมมองอื่นโดยไม่เปลี่ยนแปลงความสัมพันธ์เชิงเรขาคณิตระหว่างจุดในภาพ ในทางปฏิบัติ Homography ใช้เมทริกซ์ 3x3 ในการแปลงพิกัดของจุดในภาพหนึ่งไปยังพิกัดที่สอดคล้องในภาพอื่นสมการพื้นฐานของ Homography สามารถแสดงได้ดังนี้

กำหนดให้ (x, y) แทนพิกัดของจุดในภาพต้นฉบับและ (x', y') แทนพิกัดของจุดเดียวกันในภาพที่ปรับแปลงแล้ว สมการ Homography สามารถเขียนได้เป็น

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

โดยที่ H คือ เมทริกซ์ Homography ขนาด 3x3 ซึ่งประกอบไปด้วยสมาชิกและสามารถเขียนได้เป็น

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

เมทริกซ์ H นี้ใช้ในการแปลงพิกัดจุดในภาพอย่างเป็นระบบ ทำให้สามารถสร้างมุมมองใหม่ที่เหมือนกับการมองจากด้านบน การคำนวณเมทริกซ์ h_{ij} นั้น จำเป็นต้องมีการทำการวัดและประเมินค่าจากภาพหรือข้อมูลที่มีอยู่เพื่อที่จะสร้างการแปลงที่เหมาะสมที่สุด

2.2 การแปลงระบบพิกัด

การแปลง homogeneous transformation matrix (Shirley, P., Ashikhmin, M., & Marschner, S. 2009) ในทางหุ่นยนต์และกราฟิกคอมพิวเตอร์เป็นเทคนิคที่ใช้ในการแปลงตำแหน่งและการวางตัวของวัตถุจากระบบพิกัดหนึ่งไปยังอีกระบบพิกัดหนึ่ง โดยทั่วไปจะใช้ในการแปลง

ตำแหน่งจากระบบพิกัดท้องถิ่น (local coordinate system) เป็นระบบพิกัดทั่วโลก (global coordinate system) หรือในทางกลับกัน

การเคลื่อนที่ (Translation) ในการแปลงโฮโมจีเนียสไม่ได้จำกัดอยู่แค่การเพิ่มค่าเวกเตอร์โดยตรงเข้ากับพิกัดเท่านั้น แต่ยังช่วยให้สามารถผสานการเคลื่อนที่เข้ากับการหมุนและการสเกลได้ในเมทริกซ์เดียวกันทำให้การคำนวณการเปลี่ยนแปลงตำแหน่งเป็นไปอย่างราบรื่นและต่อเนื่อง นี่คือสมการการเคลื่อนที่

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & d_x \\ 0 & 1 & d_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

การเคลื่อนที่เป็นกระบวนการย้ายวัตถุจากตำแหน่งหนึ่งไปยังอีกตำแหน่งหนึ่งในปริภูมิสามมิติโดยไม่เปลี่ยนแปลงในรูปร่างหรือขนาดการเคลื่อนที่ใช้เวกเตอร์ของการเคลื่อนที่ d_x, d_y เพื่อเพิ่มค่าเหล่านี้ไปยังพิกัด x, y ปัจจุบันของวัตถุ

การหมุน (Rotation) เป็นหัวใจของการเปลี่ยนแปลงทิศทาง โดยเฉพาะในการจำลองหรือการเรนเดอร์ภาพสามมิติ การหมุนสามารถจำลองการเคลื่อนไหวของวัตถุได้เหมือนจริง เช่น การหมุนของเครื่องจักรหรือการเปลี่ยนทิศทางของรถยนต์นี่คือสมการการการหมุน

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

การหมุนเป็นกระบวนการเปลี่ยนทิศทางของวัตถุโดยรักษาตำแหน่งศูนย์กลางของมันเอาไว้ การหมุนใช้มุมหมุน θ และสามารถหมุนรอบแกนใดแกนหนึ่งหรือหลายแกนในการหมุนบนแกน x และ y จะใช้ค่า $\cos\theta$ และ $\sin\theta$ ในการคำนวณผลของการหมุน

การสเกล (Scaling) เป็นการเปลี่ยนแปลงที่ใช้ในการปรับขนาดของวัตถุ โดยการคูณพิกัดด้วยตัวปรับสเกลทำให้วัตถุขยายหรือย่อส่วนได้โดยไม่เปลี่ยนรูปร่างของมัน การสเกลมีประโยชน์ในการจำลองการเติบโตหรือการหดตัวของวัตถุในการจำลองกราฟิกหรือในการวิเคราะห์ข้อมูลจากภาพที่ได้จากอุปกรณ์ต่าง ๆ เช่น กล้องหรือเครื่องสแกน นี่คือสมการการสเกล

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

การสเกล คือ การเปลี่ยนขนาดของวัตถุ การเปลี่ยนขนาดในแนวแกน x และ y ทำได้โดยการคูณพิกัดด้วย s_x และ s_y ซึ่งเป็นตัวปรับสเกลที่กำหนด

2.3 การเรียนรู้ของเครื่องจักร

Machine Learning (การเรียนรู้ของเครื่องจักร) คือ สาขาวิทยาการที่อยู่ภายใต้ปัญญาประดิษฐ์ (Artificial Intelligence) โดยมีเป้าหมายในการพัฒนาเทคนิคที่ทำให้คอมพิวเตอร์สามารถเรียนรู้และทำการตัดสินใจหรือทำนายผลลัพธ์ด้วยตัวเองจากข้อมูลที่ได้รับ โดยไม่ต้องการการแทรกแซงหรือการเขียนโปรแกรมคำสั่งอย่างเฉพาะเจาะจงจากมนุษย์ในการเรียนรู้ของเครื่องจักร ระบบจะวิเคราะห์ข้อมูลจำนวนมาก เพื่อหาความสัมพันธ์ รูปแบบ หรือแนวโน้มที่ซ่อนอยู่ และใช้ข้อมูลเหล่านี้เพื่อสร้างโมเดลที่สามารถทำนายหรือตัดสินใจในสถานการณ์ใหม่ ๆ การเรียนรู้นี้อาจจะดำเนินการผ่านหลายวิธี ซึ่งหนึ่งในวิธีที่พบบ่อยคือการเรียนรู้แบบมีผู้ดูแล (Supervised Learning) ซึ่งระบบจะถูกฝึกฝนด้วยข้อมูลที่มีคำตอบหรือป้ายกำกับอยู่แล้ว เพื่อให้ระบบเรียนรู้และทำนายคำตอบสำหรับข้อมูลใหม่ที่ไม่มีป้ายกำกับและยังมีการเรียนรู้แบบไม่มีผู้ดูแล (Unsupervised Learning) ซึ่งเป็นการเรียนรู้จากข้อมูลที่ไม่มีป้ายกำกับ โดยที่ระบบจะพยายามหาความสัมพันธ์หรือกลุ่มของข้อมูลที่มีลักษณะคล้ายกัน และการเรียนรู้แบบเสริมพลัง (Reinforcement Learning) ซึ่งเป็นการเรียนรู้จากการลองผิดลองถูก โดยระบบจะได้รับรางวัลหรือการลงโทษจากการกระทำต่าง ๆ ที่ทำในสภาพแวดล้อมเสมือนจริงMachine Learning จึงเป็นเครื่องมือที่มีความสำคัญมากในยุคปัจจุบัน เนื่องจากช่วยให้คอมพิวเตอร์สามารถแก้ไขปัญหาที่ซับซ้อนและทำนายผลลัพธ์ได้ด้วยตัวเองโดยไม่ต้องโปรแกรมคำสั่งอย่างละเอียด

2.3.1 ANN (Artificial Neural Networks)

Artificial Neural Networks (ANN) เป็นหนึ่งในเทคโนโลยีหลักที่ขับเคลื่อนโลกของปัญญาประดิษฐ์และการเรียนรู้ของเครื่องจักร มีการออกแบบมาเพื่อเลียนแบบการทำงานของเครือข่ายประสาทในสมองมนุษย์ ซึ่งทำให้คอมพิวเตอร์สามารถประมวลผลข้อมูลและทำการตัดสินใจหรือทำนายผลลัพธ์ได้อย่างฉลาดโครงสร้างพื้นฐานของ ANN ประกอบด้วยชั้นของนิวรอนเทียมที่เชื่อมต่อกัน ซึ่งแต่ละนิวรอนสามารถรับและประมวลผลข้อมูลที่ป้อนเข้ามา และจากนั้นส่งผลลัพธ์ไปยังนิวรอนในชั้นถัดไป โครงสร้างนี้ช่วยให้ ANN สามารถรับรู้และแปลความหมายจากข้อมูลที่ซับซ้อนได้ในแต่ละนิวรอนเทียม การประมวลผลนี้จะเกิดขึ้นโดยใช้ฟังก์ชันการกระตุ้น ซึ่งควบคุมว่านิวรอนนั้นควรส่งสัญญาณต่อหรือไม่ ฟังก์ชันเหล่านี้สามารถเป็นแบบเชิงเส้นหรือไม่เชิงเส้น เช่น Sigmoid,

ReLU หรือ Tanh ซึ่งช่วยให้เครือข่ายสามารถรับมือกับปัญหาที่ซับซ้อนได้ารเรียนรู้ใน ANN เกิดขึ้นผ่านการปรับน้ำหนักและเอนเอียงของแต่ละนิวรอน กระบวนการนี้เรียกว่า "การส่งย้อนกลับ" (Backpropagation) ซึ่งเป็นการปรับปรุงโมเดลให้สามารถทำนายหรือตัดสินใจได้อย่างแม่นยำมากขึ้น โดยการปรับน้ำหนักเหล่านี้จะเกิดขึ้นเพื่อลดความแตกต่างระหว่างผลลัพธ์ที่ทำนายโดยเครือข่ายกับค่าจริงที่รู้อยู่แล้ว ANN ถูกใช้งานในหลายแอปพลิเคชัน ตั้งแต่การจำแนกประเภทข้อมูล การทำนายข้อมูลทางการเงิน การประมวลผลภาษาธรรมชาติ ไปจนถึงการมองเห็นของคอมพิวเตอร์ เช่น การจดจำวัตถุในภาพหรือวิดีโอ ความสามารถของ ANN ในการเรียนรู้และปรับตัวตามข้อมูลที่ซับซ้อนทำให้มันเป็นเครื่องมือที่ทรงพลังและหลากหลายในการจัดการกับปัญหาที่ต้องการความเข้าใจและการวิเคราะห์ข้อมูลในระดับลึกการวัดความแม่นยำของโมเดลในการเรียนรู้ของเครื่องจักรและการวิเคราะห์ข้อมูลเป็นส่วนสำคัญที่ช่วยให้เราประเมินและเข้าใจคุณภาพของโมเดลที่สร้างขึ้น มีหลายวิธีในการวัดความแม่นยำและประสิทธิภาพของโมเดล ซึ่งรวมถึง R-squared, Mean Squared Error (MSE), และ Mean Absolute Error (MAE) แต่ละวิธีนี้มีลักษณะเฉพาะและสำคัญในการทำความเข้าใจและปรับปรุงโมเดล R^2 (R-squared or Coefficient of Determination)

Mean Squared Error (MSE) เป็นวิธีการวัดความแตกต่างระหว่างค่าที่ประมาณโดยโมเดลกับค่าจริง โดยการคำนวณค่าเฉลี่ยของผลต่างที่ยกกำลังสองค่า MSE ที่ต่ำหมายถึงโมเดลที่มีความแม่นยำสูง เนื่องจากผลต่างระหว่างค่าที่โมเดลทำนายกับค่าจริงมีน้อย MSE คำนวณจากสมการ

$$\mathrm{M}SE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

y_i คือ ค่าจริง

$\hat{y}_i$ คือ ค่าทำนายโดยโมเดล

n คือ จำนวนของข้อมูล

MSE นับเป็นค่าเฉลี่ยของผลต่างที่ยกกำลังสองระหว่างค่าที่คาดการณ์และค่าจริง

Mean Absolute Error (MAE) คือ ค่าเฉลี่ยของค่าสัมบูรณ์ของผลต่างระหว่างค่าที่คาดการณ์โดยโมเดลกับค่าจริงคล้ายกับ MSE, ค่า MAE ที่ต่ำแสดงถึงความแม่นยำที่สูงของโมเดล MAE มีความไวต่อค่าผิดพลาดที่ใหญ่กว่า (outliers) MAE คำนวณจากสมการ

$$\mathrm{M}AE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2$$

y_i คือ ค่าจริง

$\hat{y}_i$ คือ ค่าทำนายโดยโมเดล

n คือ จำนวนของข้อมูล

MAE นับเป็นค่าเฉลี่ยของค่าสัมบูรณ์ของผลต่างระหว่างค่าทำนายและค่าจริง

R-squared (R^2) เป็นวิธีการวัดที่ใช้ในการถดถอยสถิติ (statistical regression) เพื่อกำหนดคุณภาพของการประมาณการและแสดงถึงสัดส่วนของความแปรปรวนในตัวแปรตามที่ถูกอธิบายโดยตัวแปรอิสระในโมเดลค่า R^2 มีช่วงตั้งแต่ 0 ถึง 1 โดยค่าที่ใกล้เคียง 1 หมายความว่าโมเดลสามารถอธิบายความแปรปรวนของข้อมูลได้ดี ในขณะที่ค่าที่ใกล้เคียง 0 แสดงว่าโมเดลไม่สามารถอธิบายความแปรปรวนของข้อมูลได้ R-squared คำนวณจากสมการ

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

y_i คือ ค่าจริง

$\hat{y}_i$ คือ ค่าทำนายโดยโมเดล

$\bar{y}$ คือ ค่าเฉลี่ยของ y_i

n คือ จำนวนของข้อมูล

ค่า R-squared นี้แสดงถึงสัดส่วนของความแปรปรวนในตัวแปรตามที่ถูกอธิบายโดยโมเดล ค่าที่ใกล้ 1 แสดงถึงโมเดลที่มีคุณภาพสูงในการอธิบายข้อมูล

การเลือกใช้วิธีการวัดความแม่นยำขึ้นอยู่กับบริบทและประเภทของข้อมูลที่เรามี แต่ละวิธีนี้ให้มุมมองที่แตกต่างกันในการประเมินและเข้าใจคุณภาพของโมเดล โดย R^2 มักใช้เพื่อประเมินความสามารถของโมเดลในการอธิบายข้อมูล ในขณะที่ MSE และ MAE เน้นการวัดความแม่นยำของการทำนายผลลัพธ์

2.3.2 CNN (Convolutional Neural Networks)

Convolutional Neural Networks (CNN) เป็นหนึ่งในนวัตกรรมปัญญาประดิษฐ์และการเรียนรู้ของเครื่องจักร โดยถูกออกแบบมาเพื่อจัดการกับข้อมูลที่มีโครงสร้างเช่นภาพและวิดีโอ มันมีความสามารถในการรับรู้และจดจำรูปแบบภายในข้อมูลภาพที่ซับซ้อน ทำให้เหมาะสำหรับการประยุกต์ใช้งานในหลากหลายด้าน เช่น การจำแนกประเภทภาพ การตรวจจับวัตถุ และการวิเคราะห์วิดีโอโครงสร้างพื้นฐานของ CNN ใจกลาง คือ Convolutional Layer ที่ทำหน้าที่สกัดคุณลักษณะ

จากข้อมูลภาพ การทำ Convolution คือ การนำเคอร์เนลหรือฟิลเตอร์ไปประยุกต์ใช้กับข้อมูลภาพ เพื่อจับจุดเด่นเช่นขอบ, รูปร่าง หรือลักษณะเฉพาะอื่น ๆ ในภาพ การนี้ช่วยให้เครือข่ายสามารถจดจำลักษณะที่สำคัญของวัตถุหรือฉากในภาพได้ตามมาด้วย Pooling Layer ซึ่งทำหน้าที่ลดขนาดของข้อมูลที่ได้จากชั้น Convolutional ลดความซับซ้อนและการคำนวณที่จำเป็น พูลลิ่งมักจะใช้วิธีการ เช่น Max Pooling ที่เลือกค่าสูงสุดในบล็อกของข้อมูลในท้ายสุดของ CNN คือ Fully Connected Layer ที่นำข้อมูลที่ผ่านการสกัดแล้วมาเชื่อมต่อกับเครือข่ายประสาทเทียมที่มีการเชื่อมต่ออย่างเต็มที่ ในส่วนนี้เป็นที่ทำการจำแนกประเภทหรือทำนายผลลัพธ์จากข้อมูลที่ได้มาการทำงานและประสิทธิภาพมีความสามารถพิเศษในการเรียนรู้คุณลักษณะและรูปแบบของข้อมูลภาพโดยอัตโนมัติ การทำงานของมันเลียนแบบวิธีที่มนุษย์มองและรับรู้วัตถุ ซึ่งทำให้มันสามารถจดจำและจำแนกประเภทวัตถุในภาพได้อย่างแม่นยำ CNN ยังรักษาความเชื่อมโยงทางพื้นที่ของข้อมูลภาพได้ดี สามารถจำแนกและจดจำวัตถุและรูปแบบในภาพที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ และยังใช้พารามิเตอร์ที่มีประสิทธิภาพ โดยมีการใช้ฟิลเตอร์ที่ใช้ร่วมกันซึ่งช่วยลดจำนวนพารามิเตอร์ที่จำเป็นลงอย่างมากส่วนการประยุกต์ใช้งาน CNN ถูกใช้งานในหลายด้านที่ต้องการการมองเห็นของคอมพิวเตอร์ ตั้งแต่การจำแนกประเภทภาพ, การตรวจจับวัตถุ, การวิเคราะห์วิดีโอ, การจดจำใบหน้าไปจนถึงการประมวลผลภาษาธรรมชาติ ความสามารถของมันในการจดจำรูปแบบที่ซับซ้อนทำให้ CNN เป็นเครื่องมือที่มีค่าในการแก้ไขปัญหาที่มีความซับซ้อนในโลกแห่งข้อมูลภาพการพัฒนาและการประยุกต์ใช้งานของ CNN จึงเป็นหนึ่งในด้านที่สำคัญที่สุดในการวิจัยและพัฒนาในด้านปัญญาประดิษฐ์ ช่วยให้เราสามารถสร้างระบบที่สามารถจดจำและตีความข้อมูลภาพได้อย่างฉลาดและยืดหยุ่นเพื่อวัดความแม่นยำของโมเดลในการเรียนรู้ของเครื่องจักรและการวิเคราะห์ข้อมูล มีหลายเครื่องมือและเทคนิคที่สามารถใช้ได้ แต่ละเครื่องมือหรือเทคนิคมีลักษณะเฉพาะที่ช่วยให้เราประเมินคุณภาพและประสิทธิภาพของโมเดลในแง่ต่าง ๆ ต่อไปนี้ คือ เครื่องมือหลักที่ใช้ในการวัดความแม่นยำ

Confusion Matrixในงานการจำแนกประเภทแบบสองคลาส (binary classification) Confusion Matrix ประกอบด้วยสี่ส่วนหลัก

True Positives (TP): จำนวนครั้งที่โมเดลทำนายผลบวกและถูกต้อง

True Negatives (TN): จำนวนครั้งที่โมเดลทำนายผลลบและถูกต้อง

False Positives (FP): จำนวนครั้งที่โมเดลทำนายผลบวกแต่ไม่ถูกต้อง (ข้อผิดพลาดประเภท I)

False Negatives (FN): จำนวนครั้งที่โมเดลทำนายผลลบแต่ไม่ถูกต้อง (ข้อผิดพลาดประเภท II)

การคำนวณ Accuracy (ความแม่นยำโดยรวม), Precision (ความแม่นยำในการทำนายผลบวก), Recall (ความสามารถในการจำแนกผลบวกที่แท้จริง), และ F1 Score (ค่าเฉลี่ยแบบ harmonic mean ระหว่าง Precision และ Recall) จาก Confusion Matrix เราสามารถคำนวณค่าต่าง ๆ ต่อไปนี้

Accuracy: คำนวณจากสมการ

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

ค่านี้แสดงถึงความแม่นยำโดยรวมของโมเดลค่า Accuracy ที่สูงแสดงว่าโมเดลมีประสิทธิภาพสูงในการทำนายทั้งผลบวกและผลลบในกรณีที่ข้อมูลมีความไม่สมดุล (Imbalanced Data) เช่น หนึ่งคลาสมีข้อมูลมากกว่าอีกคลาสมาก ค่า Accuracy อาจไม่สามารถสะท้อนประสิทธิภาพของโมเดลได้อย่างแท้จริง

Precision: คำนวณจากสมการ

$$Precision = \frac{TP}{TP + FP}$$

ค่านี้แสดงถึงความแม่นยำในการทำนายผลบวกค่า Precision ที่สูงบ่งบอกว่าโมเดลมีความแม่นยำสูงในการทำนายผลบวก ลดโอกาสของ False Positivesสถานการณ์ที่ False Positives นั้นมีความสำคัญ เช่น การทดสอบสำหรับโรคที่ร้ายแรง

Recall คำนวณจากสมการ

$$Recall = \frac{TP}{TP + FN}$$

ค่านี้แสดงถึงความสามารถของโมเดลในการจำแนกผลบวกที่แท้จริงค่า Recall ที่สูงแสดงว่าโมเดลสามารถจำแนกผลบวกที่แท้จริงได้ดี ลดโอกาสของ False Negativesสถานการณ์ที่ False Negatives นั้นมีความสำคัญ เช่น การตรวจจับภาวะที่เสี่ยงต่อชีวิต

F1 Score: คำนวณจากสมการ

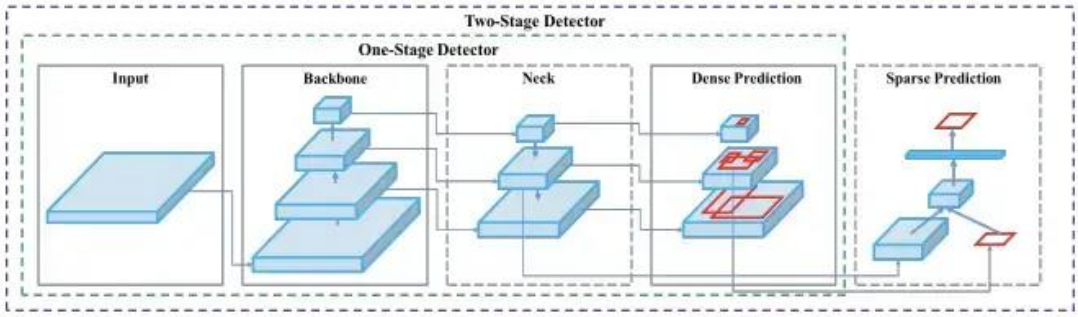
$$F1\ Score = 2 \times \frac{Percision \times Recall}{Percision + Recall}$$

F1 Score เป็นค่าเฉลี่ยแบบ harmonic mean ระหว่าง Precision และ Recall ซึ่งเป็นการวัดความสมดุลระหว่างความแม่นยำและความไวของโมเดล F1 Score ที่สูงบ่งบอกว่าโมเดลมี

ความสมดุลที่ดีระหว่าง Precision และ Recall สถานการณ์ที่ต้องการความสมดุลระหว่างการลด False Positives และ False Negatives

CNN ประกอบด้วยชั้นคอนโวลูชันที่ทำหน้าที่สกัดคุณลักษณะจากข้อมูลภาพ การทำ Convolution คือ การนำเคอร์เนลหรือฟิลเตอร์ไปประยุกต์ใช้กับข้อมูลภาพเพื่อจับจุดเด่นเช่นขอบ, รูปร่าง หรือลักษณะเฉพาะอื่น ๆ ในภาพ การนี้ช่วยให้เครือข่ายสามารถจดจำลักษณะที่สำคัญของวัตถุหรือฉากในภาพได้ ตามมาด้วย Pooling Layer ซึ่งทำหน้าที่ลดขนาดของข้อมูลที่ได้จากชั้น Convolutional ลดความซับซ้อนและการคำนวณที่จำเป็น พูลลิ่งมักจะใช้วิธีการเช่น Max Pooling ที่เลือกค่าสูงสุดในบล็อกของข้อมูล ในท้ายสุดของ CNN คือ Fully Connected Layer ที่นำข้อมูลที่ผ่านการสกัดแล้วมาเชื่อมต่อกับเครือข่ายประสาทเทียมที่มีการเชื่อมต่ออย่างเต็มที่ ในส่วนนี้เป็นที่ทำการจำแนกประเภทหรือทำนายผลลัพธ์จากข้อมูลที่ได้มา YOLO นำเสนอวิธีการตรวจจับวัตถุที่เปลี่ยนแปลงแนวทางจากเทคนิคที่ใช้ในอดีต โดยรวมการสกัดคุณลักษณะและการจำแนกประเภทเข้าด้วยกันในโมเดลเดียว ทำให้ YOLO มีความเร็วสูงและเหมาะสำหรับการใช้งานในระบบเรียลไทม์ การทำงานของ YOLO ทำการตรวจจับวัตถุในภาพเพียงครั้งเดียว โดยแบ่งภาพออกเป็นกริดขนาดเล็ก ๆ และทำการทำนายกรอบขอบเขต (bounding boxes) และความน่าจะเป็นของคลาสต่าง ๆ สำหรับแต่ละเซลล์ในกริด การนี้ช่วยลดเวลาและความซับซ้อนในการประมวลผล YOLO (Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016)นำเสนอการประมวลผลแบบ end-to-end ที่สามารถทำนายตำแหน่งและประเภทของวัตถุในภาพได้ในครั้งเดียว ทำให้มีประสิทธิภาพสูงในการตรวจจับวัตถุหลายอย่างในภาพได้อย่างรวดเร็วและแม่นยำ

You Only Look Once (YOLO) เป็นอัลกอริธึมการตรวจจับวัตถุที่เปิดตัวในปี 2558 ในรายงานการวิจัยโดย Joseph Redmon, Santosh Divvala, Ross Girshick และ Ali Farhadi สถาปัตยกรรมของ YOLO เป็นการปฏิวัติครั้งสำคัญใน พื้นที่ การตรวจจับวัตถุ แบบเรียลไทม์ ซึ่งเหนือกว่ารุ่นก่อน นั่นคือ Convolutional Neural Network (R-CNN) YOLO เป็นอัลกอริธึม ช็อตเดียวที่จัดประเภทวัตถุโดยตรงในการส่งผ่านครั้งเดียว โดยมีโครงข่ายประสาทเทียมเพียงเครือข่ายเดียวทำนายกรอบขอบเขตและความน่าจะเป็นของคลาสโดยใช้รูปภาพเต็มเป็นอินพุต โมเดล YOLO มีการพัฒนาอย่างต่อเนื่อง ตั้งแต่นั้นเป็นต้นมา ทีมวิจัยหลายทีมได้เปิดตัว YOLO เวอร์ชันต่าง ๆ โดย YOLOv8 เป็นเวอร์ชันล่าสุด ส่วนต่อไปนี้จะสรุปภาพรวมของเวอร์ชันที่ผ่านมาทั้งหมดและการปรับปรุง

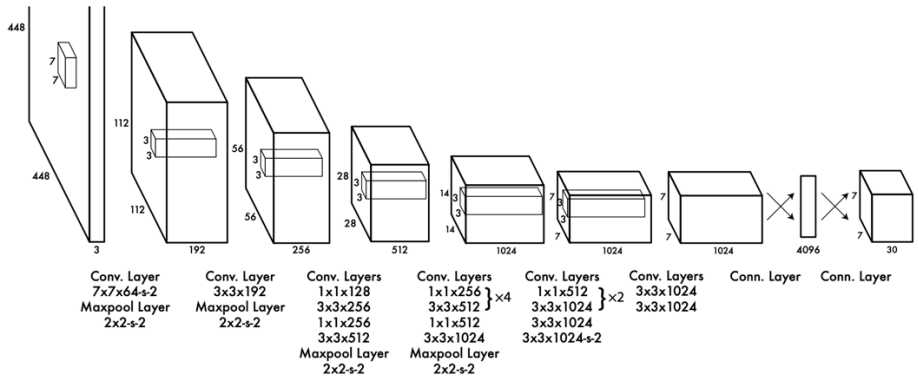


รูปที่ 2.1 แสดงกลไกที่สำคัญของแบบจำลองการตรวจจับวัตถุ (Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. 2020).

การทำงานพื้นฐานของโมเดลตรวจจับวัตถุประกอบไปด้วยสามส่วนหลัก ๆ คือ ส่วนหลัง (backbone), คอ (neck), และหัว (head) ส่วนหลังเป็นเครือข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN) ที่ได้รับการฝึกฝนมาแล้วเพื่อสกัดคุณลักษณะต่าง ๆ ที่มีระดับต่ำ, ปานกลาง, และสูงออกจากรูปภาพที่นำเข้ามา ส่วนคอจะทำหน้าที่ผสานคุณลักษณะเหล่านี้ด้วยบล็อกการรวมเส้นทางเช่นเครือข่ายพีระมิดคุณลักษณะ (Feature Pyramid Network - FPN) และส่งต่อไปยังส่วนหัวเพื่อจำแนกประเภทวัตถุและทำนายกรอบขอบเขตของวัตถุ ส่วนหัวอาจประกอบด้วยโมเดลการทำนายแบบหนึ่งขั้นตอนหรือการทำนายแบบหนาแน่น เช่น YOLO หรือ Single-shot Detector (SSD) หรืออาจใช้โมเดลการทำนายแบบสองขั้นตอนหรือการทำนายแบบกระจัดกระจาย เช่น R-CNN

ประวัติโดยย่อของ YOLO

YOLOv1 เป็นเวอร์ชันแรกที่นำเสนอวิธีการทำนายกรอบขอบเขตและความน่าจะเป็นของประเภทวัตถุในครั้งเดียว โดยการแบ่งภาพเป็นกริดหลายช่องและคำนวณความมั่นใจและกรอบขอบเขตสำหรับแต่ละช่องกริด ทำให้มีความเร็วและความแม่นยำสูงกว่า R-CNN ที่ใช้ก่อนหน้านี้



รูปที่ 2.2 สถาปัตยกรรม Yolo V1 (Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016).

YOLOv2 นำเสนอ 'anchor boxes' ซึ่งเป็นกรอบขอบเขตที่กำหนดไว้ล่วงหน้า ช่วยให้การทำนายตำแหน่งของวัตถุมีความแม่นยำมากขึ้น โดยสามารถทำนายวัตถุได้มากกว่า 9000 ประเภท และมีการปรับปรุงในเรื่องความเร็วและความแม่นยำ (Redmon, J., & Farhadi, A. 2017)

ตารางที่ 2.1 ทดสอบการตรวจจับ Pascal VOC 2007 YOLOv2 เร็วและแม่นยำกว่าการตรวจจับอื่น

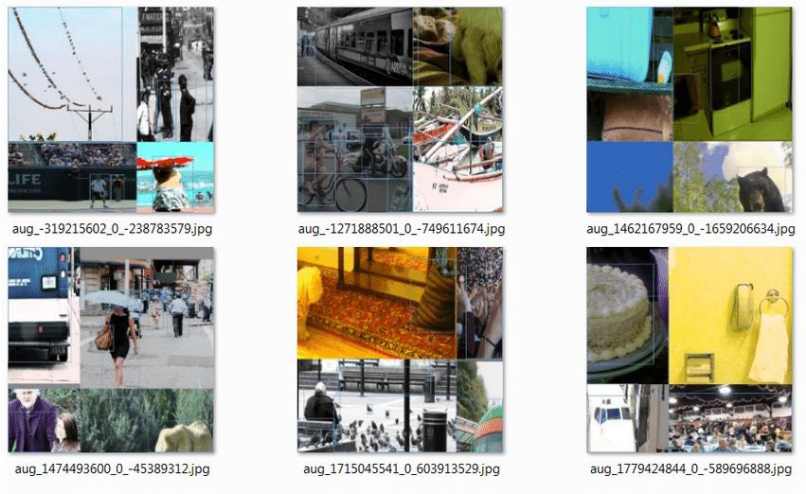
Detection Frameworks	**Train**	**mAP**	**FPS**
Fast R-CNN (Girshick, R. B. 2015)	2007+2012	70.0	0.5
Faster R-CNN VGG-16 (Sun, J. 2015)	2007+2012	73.2	7
Faster R-CNN ResNet (Sun, J. 2016)	2007+2012	76.4	5
YOLO (Farhadi, A. 2016)	2007+2012	63.4	45
SSD300 (Berg, A. C. 2016)	2007+2012	74.3	46
SSD500 (Berg, A. C. 2016)	2007+2012	76.8	19
YOLOv2 288 x 288	2007+2012	69.0	91
YOLOv2 352 x 352	2007+2012	73.7	81
YOLOv2 416 x 416	2007+2012	76.8	67
YOLOv2 480 x 480	2007+2012	77.8	59
YOLOv2 544 x 544	2007+2012	78.6	40

YOLOv3 เป็นการอัพเกรดที่มีความแม่นยำสูงขึ้น โดยใช้ Darknet-53 เป็นส่วนหลัง และใช้ logistic classifiers แทน softmax รวมถึงการใช้ Binary Cross-entropy (BCE) loss ในการทำนายประเภทของวัตถุ ทำให้สามารถจำแนกประเภทวัตถุได้แม่นยำขึ้น Redmon, J., & Farhadi, A. (2018)

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

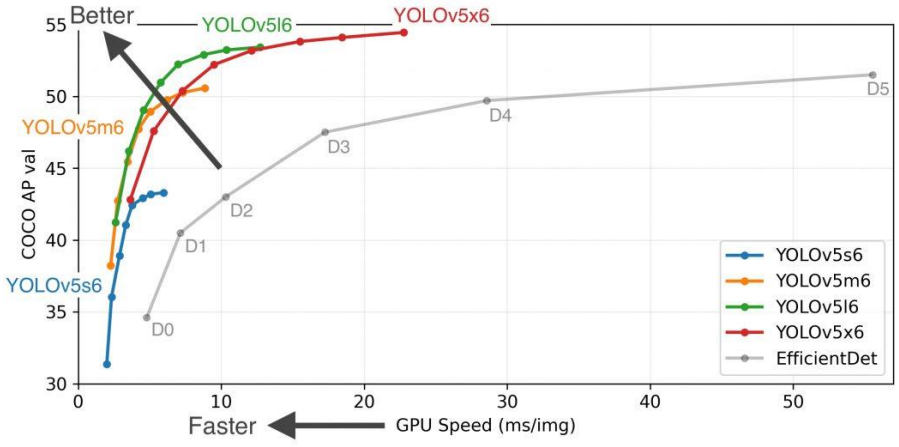
รูปที่ 2.3 Darknet-53 (Chen, R. C. 2019)

YOLOV4 นำเสนอความคิดใหม่อย่าง 'Bag of Freebies' (BoF) และ 'Bag of Specials' (BoS) ซึ่งเป็นเทคนิคที่ช่วยเพิ่มความแม่นยำโดยไม่เพิ่มต้นทุนในการคำนวณ ใช้เทคนิคการเพิ่มข้อมูลแบบใหม่อย่าง Mosaic ที่รวมภาพการฝึกฝนสี่ภาพเข้าด้วยกัน ทำให้ได้ข้อมูลเพิ่มเติมสำหรับการฝึกฝน Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020).



รูปที่ 2.4 Mosaic represents

YOLOv5 ซึ่งถูกพัฒนาโดย Ultralytics เป็นเวอร์ชันที่โดดเด่นในซีรีส์ของ YOLO (You Only Look Once) สำหรับการตรวจจับวัตถุ YOLOv5 ได้รับความนิยมอย่างกว้างขวางและการใช้งานอย่างแพร่หลายในชุมชนของการเรียนรู้ด้วยเครื่องจักร หนึ่งในคุณสมบัติหลักของ YOLOv5 คือ การใช้งานที่ง่ายและการปรับแต่งที่ยืดหยุ่นการทำงานของ YOLOv5 มีการออกแบบที่เน้นความเรียบง่าย โดยการใช้ PyTorch, ไลบรารีการเรียนรู้เชิงลึกที่นิยมใช้ ซึ่งทำให้โมเดลนี้เข้าถึงได้ง่ายสำหรับนักพัฒนาและนักวิจัยที่หลากหลาย การเลือกใช้ PyTorch ทำให้โมเดลมีความยืดหยุ่นและง่ายต่อการใช้งาน YOLOv5 ยังใช้ไฟล์ YAML (Yet Another Markup Language) สำหรับการกำหนดค่าแทนการใช้ไฟล์ CFG (configuration file) ที่ใช้ในเวอร์ชันก่อน ๆ YAML ซึ่งเป็นมาตรฐานการบันทึกข้อมูลที่เข้าใจได้ง่ายสำหรับมนุษย์ ทำให้กระบวนการกำหนดค่าของโมเดลนั้นง่ายขึ้นและเข้าถึงได้ง่ายกว่าคุณสมบัติอีกประการหนึ่งของ YOLOv5 คือ ความง่ายในการฝึกฝนและการใช้งาน โครงสร้างของโมเดลและกระบวนการฝึกฝนถูกออกแบบมาเพื่อให้เข้าใจง่ายและมีประสิทธิภาพ ช่วยให้ผู้ใช้สามารถฝึกฝนและใช้งานโมเดลการตรวจจับวัตถุได้โดยใช้ความพยายามและทรัพยากรคอมพิวเตอร์ที่น้อยลงเมื่อเทียบกับเวอร์ชันก่อน ๆ ด้วยคุณสมบัติเหล่านี้ ทำให้ YOLOv5 เป็นเครื่องมยู่ในชุดเครื่องมือที่สำคัญสำหรับนักพัฒนาและนักวิจัยในด้านการมองเห็นด้วยคอมพิวเตอร์ แม้จะไม่มีการเผยแพร่งานวิจัยอย่างเป็นทางการ แต่ YOLOv5 ได้แสดงให้เห็นถึงประสิทธิภาพที่เข้มข้นในงานตรวจจับวัตถุและได้รับการยอมรับอย่างกว้างขวางในการประยุกต์ใช้งานที่หลากหลาย ตั้งแต่งานอุตสาหกรรมไปจนถึงโครงการวิจัย



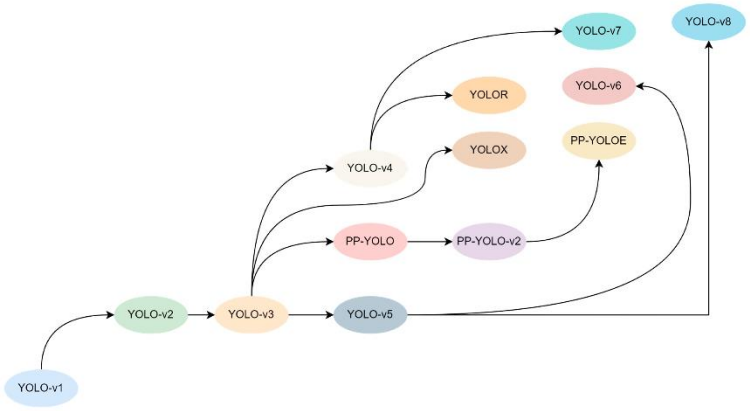
รูปที่ 2.5 กราฟแสดงความเร็วในการตรวจจับของ YOLOv5 Ultralytics. (2020).

YOLOv6 ถูกนำเสนอโดย Meituan จากประเทศจีน, ได้เปิดตัวเป็นหนึ่งในโมเดลการตรวจจับวัตถุที่มีความเร็วและความแม่นยำสูงสำหรับการใช้งานในอุตสาหกรรมการพัฒนานี้มุ่งเน้นไปที่การปรับปรุงหลายด้าน เพื่อเพิ่มประสิทธิภาพของโมเดลและทำให้มันเหมาะกับการใช้งานในสภาพแวดล้อมที่มีความต้องการความเร็วสูงคุณสมบัติหลายประการที่ทำให้มีความเร็วและความแม่นยำที่สูงขึ้น เมื่อเทียบกับเวอร์ชันก่อน ๆ คือ การตรวจจับแบบไม่ใช้ Anchor (Anchor-free): โมเดล YOLOv6 เปลี่ยนไปใช้วิธีการตรวจจับที่ไม่ใช้ anchor ซึ่งช่วยเพิ่มความสามารถในการทั่วไปของโมเดลและลดเวลาในการประมวลผลหลังการตรวจจับ ทำให้ YOLOv6 เร็วขึ้นถึง 51% เมื่อเทียบกับโมเดลที่ใช้วิธี anchor-based (Wei, X. 2022).

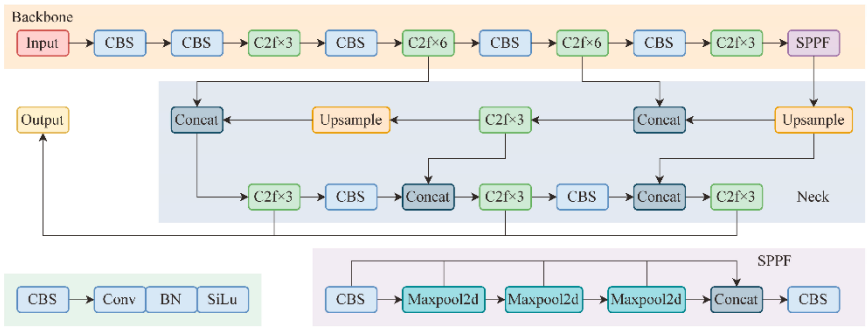
YOLOv7 เป็นโมเดลที่มีความเร็วและความแม่นยำสูงสุดในขณะนั้น ใช้ Extended Efficient Layer Aggregation Network (E-ELAN) เป็นส่วนหลังและใช้ compound scaling ซึ่งช่วยให้การฝึกฝนและทำนายมีประสิทธิภาพสูงขึ้น (Liao, H. Y. M. 2023).

YOLOv8 มีการปรับปรุงให้มีความแม่นยำและความเร็วในการทำนายที่สูงขึ้น โดย YOLOv8 (Ultralytics. 2023) มีการใช้งานที่ง่ายด้วย Python package และ CLI และสามารถทำการตรวจจับวัตถุในภาพหรือวิดีโอสดได้ efficaciously. ในเวอร์ชั่นนี้สามารถตรวจจับวัตถุที่มีขนาด

เล็กได้ดี (Liao, H. Y. M. 2023) การตรวจจับวัตถุทางอากาศไร้คนขับ (UAV)นำมามาใช้เพื่อเพิ่มประสิทธิภาพในการตรวจจับที่ดีขึ้น



รูปที่ 2.6 แผนภาพแสดงการพัฒนาของ (YOLO Zhang, R. 2023).



รูปที่ 2.7 ส่วนประกอบหลักของสถาปัตยกรรม (YOLOv8 Chen, H. 2023).

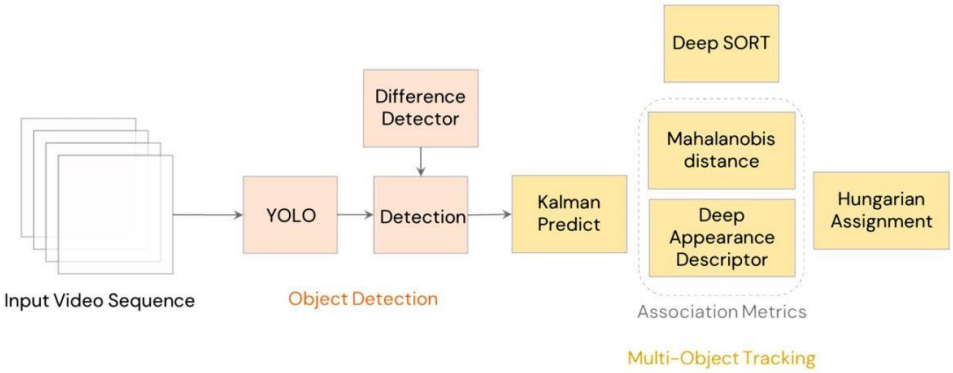
2.4 การติดตามวัตถุ

การติดตามวัตถุ (Object Tracking) ในวิดีโอเป็นหนึ่งในงานท้าทายในด้านการมองเห็นของคอมพิวเตอร์ งานนี้เกี่ยวข้องกับการระบุตำแหน่งและการติดตามการเคลื่อนไหวของวัตถุหนึ่งหรือหลายวัตถุในซีเควนซ์วิดีโอต่อเนื่อง วิธีการติดตามวัตถุมีหลากหลายรูปแบบ ตั้งแต่การติดตามวัตถุด้วยเทคนิคทางเรขาคณิตไปจนถึงการใช้แบบจำลองการเรียนรู้เชิงลึกหนึ่งในเทคนิคที่โดดเด่นในการติดตามวัตถุคือ DeepSORT ซึ่งเป็นการพัฒนาต่อจากเทคนิค SORT โดยเพิ่มความสามารถในการจดจำลักษณะเฉพาะของวัตถุโดยใช้ deep learning และนำมาประยุกต์ใช้เพื่อเพิ่มความแม่นยำในการติดตามวัตถุ โดยเฉพาะในสถานการณ์ที่วัตถุมีการเคลื่อนไหวเร็วหรือมีการเปลี่ยนแปลงลักษณะ

การใช้ DeepSORT ในการติดตามวัตถุช่วยเพิ่มความสามารถในการติดตามวัตถุได้อย่างต่อเนื่องและลดข้อผิดพลาดที่อาจเกิดจากการปกปิดหรือการสับเปลี่ยนระหว่างวัตถุ

2.4.1 Deepsort

Deep SORT (Simple Online and Realtime Tracking) เป็นเทคนิคขั้นสูงในการติดตามวัตถุหลายวัตถุภายในวิดีโอสตรีมมิ่ง (Nunes, U. J. 2022) วิธีนี้พัฒนาต่อยอดมาจาก SORT (Simple Online and Realtime Tracking) โดยใช้แผนการของตัวกรองคาลมานในการติดตามวัตถุ Deep SORT นำเสนอการวิเคราะห์การเชื่อมโยงของวัตถุโดยใช้คุณลักษณะลึกซึ่งได้รับการเรียนรู้จากเครือข่ายประสาทเทียมขั้นสูง ทำให้สามารถจัดการได้กับสถานการณ์ที่วัตถุอาจหายไปชั่วคราวหรือถูกบดบัง Deep SORT ยังรวมถึงการกำหนดรหัสประจำตัว (ID) สำหรับติดตามแต่ละวัตถุในหลายๆ เฟรมภาพ ซึ่งมีความสำคัญต่อการประยุกต์ใช้งานในด้านต่าง ๆ เช่น การเฝ้าระวังด้วยกล้องวงจรปิด ระบบยานพาหนะอัตโนมัติ และการโต้ตอบระหว่างมนุษย์กับคอมพิวเตอร์ อัลกอริธึมนี้ทำงานโดยการประมวลผลในสองขั้นตอน การสร้างการตรวจจับวัตถุก่อน แล้วต่อด้วยการเชื่อมโยงการตรวจจับเหล่านี้กับเส้นทางติดตามที่มีอยู่ โดยรวมแล้ว Deep SORT นำเสนอความก้าวหน้าที่สำคัญในด้านความแม่นยำและความเสถียรของการติดตาม เมื่อเทียบกับเทคนิคการติดตามแบบเดิม ทำให้เป็นเครื่องมือที่มีคุณค่าในด้านการมองเห็นด้วยคอมพิวเตอร์และแอปพลิเคชันด้านปัญญาประดิษฐ์



รูปที่ 2.8 สถาปัตยกรรม Deep SORT (Ikomia 2023).

Deep SORT ประกอบด้วยองค์ประกอบสำคัญ 4 ประการ ดังนี้

1) การตรวจจับและการแยกคุณสมบัติ

อัลกอริธึม Deep SORT เริ่มต้นด้วยกระบวนการตรวจจับวัตถุ โดยมักใช้เครือข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) เช่น YOLO (You Only Look Once) เพื่อจำแนกวัตถุภายในแต่ละเฟรม การตรวจจับแต่ละครั้งจะเชื่อมโยงกับตัวแทนลักษณะที่มีมิติสูงซึ่งสกัดโดย CNN โดยตัวแทนเหล่านี้จะทำการตรวจจับลักษณะภายนอกของวัตถุที่ถูกตรวจจับเพื่อใช้ในการจับคู่

2) ตัวกรองคาลมานในการทำนายสถานะ

ตัวกรองคาลมาน (Kalman Filter) ในระบบ Deep SORT มีบทบาทสำคัญในการทำนายสถานะของวัตถุ โดยทำหน้าที่เป็นอัลกอริธึมหลักในการประมาณค่าสถานะของกระบวนการที่มีความไม่แน่นอน เช่น ตำแหน่ง, ความเร็ว, และความเร่งของวัตถุ ความสามารถในการทำนายตำแหน่งและเส้นทางเคลื่อนที่ของวัตถุในเฟรมต่อไปด้วยความแม่นยำสูงเป็นคุณสมบัติที่โดดเด่นของตัวกรองนี้ตัวกรองคาลมานใช้ข้อมูลสถานะล่าสุดและการวัดที่ได้จากเซ็นเซอร์หรือตัวตรวจจับเพื่อทำการประมาณสถานะของวัตถุในเฟรมปัจจุบัน โดยคำนึงถึงตำแหน่ง, ความเร็ว, และความเร่ง การประมาณการนี้ไม่เพียงช่วยในการทำนายตำแหน่งของวัตถุ แต่ยังช่วยในการคำนวณความไม่แน่นอนหรือข้อผิดพลาดในการประมาณการด้วย ซึ่งเป็นสิ่งจำเป็นในการคำนวณความเป็นไปได้หรือความน่าเชื่อถือของสถานะที่ประมาณได้หนึ่งในข้อดีของตัวกรองคาลมานคือความสามารถในการปรับปรุงสถานะของวัตถุอย่างต่อเนื่องผ่านแต่ละเฟรม โดยอาศัยข้อมูลปัจจุบันและประวัติการเคลื่อนไหวที่ผ่านมา เมื่อมีข้อมูลใหม่มาจากการตรวจจับในเฟรมต่อไป ตัวกรองคาลมานจะอัพเดตประมาณการสถานะโดยพิจารณาจากข้อมูลใหม่นี้ ทำให้การประมาณการมีความแม่นยำมากขึ้นตัวกรองคาลมานยังได้รับการออกแบบมาเพื่อจัดการกับความไม่แน่นอนและข้อผิดพลาดที่เกิดจากการวัดหรือสภาวะแวดล้อมที่ไม่สมบูรณ์ ช่วยให้การทำนายสถานะของวัตถุมีความน่าเชื่อถือมากขึ้น การใช้ตัวกรองคาลมานใน Deep SORT จึงทำให้มีความสามารถในการติดตามวัตถุที่มีความแม่นยำสูงและสามารถจัดการกับการเปลี่ยนแปลงของสถานะวัตถุที่รวดเร็วหรือไม่แน่นอนได้ดีขึ้น ซึ่งเป็นสิ่งสำคัญในการประยุกต์ใช้ในการติดตามวัตถุในวิดีโอสตรีมมิ่งแบบเรียลไทม์การเชื่อมโยงข้อมูลด้วยเมตริกลักษณะที่ปรากฏเชิงลึกการเชื่อมโยงข้อมูลด้วยเมตริกลักษณะที่ปรากฏเชิงลึกใน Deep SORT เป็นกระบวนการที่เข้าถึงการติดตามวัตถุอย่างซับซ้อนและแม่นยำมากขึ้น โดยไม่ได้พึ่งพาเพียง IoU (Intersection over Union) แต่ใช้การผสมผสานของข้อมูลการเคลื่อนไหวและลักษณะที่ปรากฏของวัตถุ ซึ่งประกอบด้วยสองส่วนหลักระยะทางของ Mahalanobis สำหรับการเคลื่อนไหว: ระยะทาง Mahalanobis เป็นวิธีการวัดที่ใช้ในการประเมินความสัมพันธ์ระหว่างการเคลื่อนที่ของวัตถุ มันช่วยให้สามารถคำนวณความแตกต่างระหว่างตำแหน่งปัจจุบันและตำแหน่งที่ทำนายไว้ของวัตถุ โดยพิจารณาจากการเคลื่อนไหวและการเปลี่ยนแปลงของวัตถุในแต่ละเฟรม วิธีนี้ช่วยลดความผิดพลาดจากการติดตามวัตถุที่เคลื่อนที่อย่างรวดเร็วหรือมีการเปลี่ยนแปลงทิศทางระยะทางโคไซน์สำหรับความคล้ายคลึงของรูปลักษณ์: ระยะทางโคไซน์ใช้ในการวัดความคล้ายคลึงของลักษณะที่ปรากฏของวัตถุ โดยเปรียบเทียบลักษณะที่สกัดจากเครือข่ายประสาทเทียมกับลักษณะที่มีอยู่ในแทร็กที่มีอยู่ วิธีนี้ช่วยให้สามารถจดจำและติดตามต่อไปได้ แม้ว่าวัตถุจะมีการเปลี่ยนแปลงในรูปลักษณ์เล็กน้อย หรือมีการบดบังเป็นระยะเวลาสั้น ๆ การใช้เมตริกซ์ที่รวมทั้งสองนี้ช่วยให้ Deep SORT สามารถจับคู่วัตถุกับแทร็กที่มีอยู่ได้อย่างแม่นยำ โดยใช้อัลกอริธึมของฮังการี Kuhn, H. W. (1955). ในการคำนวณการ

จับคู่ที่เหมาะสมที่สุด วิธีการนี้ช่วยให้สามารถติดตามวัตถุได้อย่างต่อเนื่องและแม่นยำ แม้ในสภาพที่มีการบดบังหรือการเปลี่ยนแปลงในลักษณะที่ปรากฏของวัตถุ

3) การจัดการติดตาม

Deep SORT ปรับปรุงวิธีการจัดการติดตามวัตถุในวิดีโอสตรีมมิ่งโดยใช้กลไกที่ซับซ้อนในการดูแลรักษาแทร็กหรือเส้นทางการเคลื่อนที่ของวัตถุตลอดวงจรชีวิตของมัน กลไกนี้รวมถึงหลายขั้นตอนที่สำคัญ เช่น การยืนยันแทร็ก, การบำรุงรักษาแทร็ก และการลบแทร็กที่ไม่ใช้งาน การยืนยันแทร็กเกิดขึ้นเมื่อแทร็กได้รับการตรวจพบอย่างต่อเนื่องในหลายเฟรมติดต่อกัน ทำให้สามารถระบุและยืนยันว่าแทร็กนั้นแสดงถึงวัตถุที่กำลังถูกติดตามจริง แทร็กที่ได้รับการยืนยันนี้จะถูกประมวลผลต่อเพื่อรักษาความสม่ำเสมอและความน่าเชื่อถือในการติดตามในขั้นตอนของการบำรุงรักษาแทร็ก, Deep SORT จะอัพเดตและปรับปรุงแทร็กตามข้อมูลใหม่ที่ได้รับจากการตรวจจับในเฟรมล่าสุด การปรับปรุงนี้รวมถึงการปรับตำแหน่งแทร็ก, ความเร็ว และการเปลี่ยนแปลงอื่น ๆ ที่เกี่ยวข้องกับการเคลื่อนที่ของวัตถุ ทำให้สามารถติดตามวัตถุได้อย่างแม่นยำและต่อเนื่องการลบแทร็กที่ไม่ใช้งานเป็นขั้นตอนสุดท้ายในการจัดการติดตาม ซึ่งใช้พารามิเตอร์อายุเพื่อลบแทร็กที่ไม่ได้รับการตรวจพบในเฟรมล่าสุดหรือเฟรมส่วนใหญ่ หากวัตถุที่เคยติดตามหายไปจากภาพหรือไม่ได้รับการตรวจจับเป็นเวลานาน แทร็กนั้นจะถูกลบออกจากระบบ เพื่อป้องกันไม่ให้ระบบถูกโหลดด้วยข้อมูลที่ล้าสมัยหรือไม่แม่นยำการจัดการติดตามใน Deep SORT ด้วยวิธีนี้ช่วยให้ระบบสามารถรักษาความแม่นยำและประสิทธิภาพในการติดตามวัตถุได้อย่างมีประสิทธิผล โดยทำให้ระบบสามารถตอบสนองและปรับตัวตามการเปลี่ยนแปลงของวัตถุที่ติดตามได้อย่างรวดเร็วและแม่นยำ

2.5 การระบุตำแหน่งของยานพาหนะอัตโนมัติ

การระบุตำแหน่งอย่างแม่นยำเป็นหน้าที่พื้นฐานและสำคัญที่สุดของยานพาหนะอัตโนมัติ ยานพาหนะเหล่านี้ต้องสามารถนำทางและตอบสนองต่อสภาพแวดล้อมได้อย่างเชื่อถือได้เพื่อความปลอดภัยและความราบรื่นในการขับขี่ การทำความเข้าใจกับสภาพแวดล้อมโดยละเอียดและการตรวจจับสิ่งกีดขวางได้อย่างรวดเร็วจึงเป็นสิ่งจำเป็น เทคโนโลยีที่มีอยู่ในปัจจุบันให้ความสามารถในการระบุตำแหน่งที่เหนือกว่าด้วยความช่วยเหลือจากแผนที่ความละเอียดสูง (HD Maps) และการวิเคราะห์ข้อมูลด้วยวิธี Normal Distribution Transform (NDT) ซึ่งมักจะถูกนำมาใช้เป็นชั้นการประมวลผลในการเปรียบเทียบข้อมูลเซ็นเซอร์กับแผนที่ ในการจบบทความนี้ สิ่งสำคัญคือการเน้นว่าการใช้ HD Map และ NDT นั้นช่วยให้ยานพาหนะอัตโนมัติสามารถระบุตำแหน่งตัวเองได้อย่างแม่นยำและเชื่อถือได้ โดยทำให้การนำทางเป็นไปได้ด้วยความปลอดภัยและเป็นระบบมากขึ้น ซึ่งจะนำพาเราเข้าสู่อนาคตของการขนส่งที่ไร้คนขับ

2.5.1 แผนที่ความละเอียดสูง (HD Maps)

การวิจัยและพัฒนาแผนที่ความละเอียดสูง (High-definition map, HD map) มีบทบาทสำคัญในระบบขับขี่อัตโนมัติ โดยเริ่มแรก HD map ได้รับการพัฒนาในโครงการ Bertha Drive Project ตั้งแต่ปี 2010 ซึ่งได้รับการทดสอบบน Mercedes Benz S-Class S 500 Stiller, C. (2014, June) และได้แสดงความสามารถในการนำทางอัตโนมัติได้ในระยะทางกว่า 103 กิโลเมตรบนท้องถนนจริงการพัฒนาและการปรับปรุงแผนที่ความละเอียดสูง (High-definition Maps, HD Maps) ที่สร้างขึ้นด้วยเทคโนโลยี Lidar (Light Detection and Ranging) Mayr, M. (2018, November). นับเป็นหนึ่งในส่วนประกอบหลักที่มีบทบาทสำคัญในการทำให้ระบบขับขี่อัตโนมัติของยานพาหนะมีประสิทธิภาพและความน่าเชื่อถือสูงLidar เป็นเครื่องมือที่ใช้วัดระยะทางโดยใช้แสงเลเซอร์ที่ส่องออกไปและวัดเวลาที่แสงนั้นเดินทางกลับมาหลังจากสะท้อนจากวัตถุต่าง ๆ ด้วยความสามารถนี้, Lidar จึงสามารถสร้างข้อมูลทางเรขาคณิตที่มีความละเอียดสูงของสภาพแวดล้อมรอบตัว ซึ่งเป็นประโยชน์อย่างมากในอุตสาหกรรมยานพาหนะอัตโนมัติการสร้างภาพแบบสามมิติหรือที่เรียกว่า 'point cloud' จากข้อมูล Lidar ช่วยให้ผู้ผลิตยานพาหนะสามารถมองเห็นโครงสร้างพื้นผิวอุปสรรคและรายละเอียดอื่น ๆ ของท้องถนนได้ด้วยความแม่นยำที่ไม่เคยมีมาก่อนแผนที่ HD ที่สร้างจาก Lidar มีความสำคัญต่อระบบขับขี่อัตโนมัติเพราะมันช่วยให้รถยนต์สามารถทำความเข้าใจและตอบสนองต่อสภาพแวดล้อมที่มีความซับซ้อนได้อย่างแม่นยำ ไม่ว่าจะเป็นการรับรู้เส้นทาง, การจดจำสัญญาณจราจร, หรือแม้กระทั่งการปรับเปลี่ยนเส้นทางตามสภาพการจราจรในเวลาจริง แผนที่ HD ยังมีการอัพเดตอย่างต่อเนื่องเพื่อให้สะท้อนถึงการเปลี่ยนแปลงในสภาพแวดล้อม เช่น การก่อสร้างใหม่หรือการเปลี่ยนแปลงในการจัดการจราจร ทำให้ระบบขับขี่อัตโนมัติมีข้อมูลล่าสุดและแม่นยำที่สุดที่สุดในการทำการตัดสินใจการปรับปรุงแผนที่อย่างต่อเนื่องนี้ช่วยลดความเสี่ยงของข้อมูลที่ล้าสมัยซึ่งอาจนำไปสู่การตัดสินใจที่ไม่ถูกต้องในระหว่างการขับขี่แผนที่ HD ที่สร้างด้วย Lidar มีคุณสมบัติที่เหนือกว่าเมื่อเทียบกับแผนที่แบบดั้งเดิมในหลายด้านสำหรับตัวอย่าง, ความสามารถในการระบุรายละเอียดและความแตกต่างของพื้นผิวถนนที่ไม่สามารถจับจดด้วยเทคโนโลยีอื่น ๆ ได้, การจับภาพสิ่งกีดขวางที่มีขนาดเล็กอย่างเช่น กิ่งไม้หรือสิ่งปลูกสร้างชั่วคราว, และความสามารถในการแสดงภาพความลึกของสภาพแวดล้อม ช่วยให้ระบบขับขี่อัตโนมัติสามารถทำความเข้าใจสภาพแวดล้อมได้ในมิติที่ลึกขึ้นการนำเสนอข้อมูลดังกล่าวในรูปแบบแผนที่ HD นั้นสำคัญยิ่งกว่าเมื่อมีการประมวลผลข้อมูลในเวลาจริง Fallah, Y. P. (2022). แผนที่ HD ช่วยให้ระบบขับขี่อัตโนมัติสามารถปรับแผนการนำทางและการตัดสินใจขับขี่อย่างรวดเร็ว เมื่อมีการเปลี่ยนแปลงของสภาพทางหรือสภาพการจราจรที่ไม่คาดคิด เช่น การปิดถนนเนื่องจากการก่อสร้างหรืออุบัติเหตุ ระบบขับขี่อัตโนมัติที่มีการอัพเดตข้อมูลแผนที่ HD อย่างสม่ำเสมอจะสามารถนำทางได้อย่างมั่นใจและปลอดภัย ห่างไกลจากความผิดพลาดที่อาจเกิดขึ้น

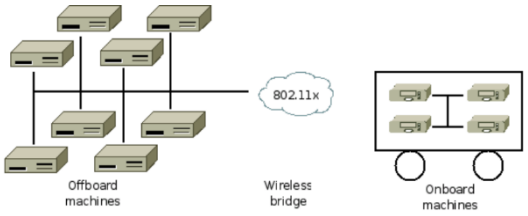
2.5.2 การระบุตำแหน่งด้วย Normal distribution transform

การระบุตำแหน่งของยานพาหนะด้วยการใช้แผนที่ความละเอียดสูง (HD map) และเทคนิค Normal Distribution Transform (NDT) ร่วมกับเซ็นเซอร์ Inertial Measurement Unit (IMU) เป็นกระบวนการที่สำคัญในการนำทางอัตโนมัติNDT เป็นอัลกอริทึมที่ใช้ในการจับคู่แผนที่ 3D point cloud ที่ได้จากเซ็นเซอร์ Lidar กับแผนที่ HD เพื่อระบุตำแหน่งและทิศทางของยานพาหนะในสภาพแวดล้อมจริงIMU ให้ข้อมูลเกี่ยวกับการเคลื่อนที่และการเอียงของยานพาหนะ, ช่วยให้ระบบสามารถประมวลผลการเปลี่ยนแปลงของทิศทางและตำแหน่งได้ด้วยความเร็วสูง NDT เป็นเทคนิคการจับคู่แผนที่ที่ใช้การแปลงการกระจายความน่าจะเป็นเพื่อระบุตำแหน่ง 3D point clouds ที่ได้จาก Lidar กับแผนที่ HD ที่มีอยู่อัลกอริทึมนี้ไม่ต้องพึ่งพาการจับคู่จุดทีละจุดแบบดั้งเดิม แต่ใช้ความน่าจะเป็นของการแจกแจงตามปกติ (normal distribution) ของความสูงและคุณสมบัติอื่น ๆ ของสภาพแวดล้อมที่สแกนมาวิธีนี้ช่วยให้การระบุตำแหน่งแม่นยำและเร็วขึ้นเนื่องจากสามารถประมวลผลข้อมูลที่ซับซ้อนได้อย่างรวดเร็ว IMU เป็นเซ็นเซอร์ที่วัดการเคลื่อนที่และการเอียงของยานพาหนะ ซึ่งรวมถึงเซ็นเซอร์เช่น gyroscopes และ accelerometers ที่วัดการเคลื่อนไหวในสามแกนคือ pitch, roll, และ yawข้อมูลจาก IMU เมื่อรวมกับข้อมูลจาก GPS และ Lidar ช่วยให้ระบบสามารถรักษาและอัปเดตตำแหน่งของยานพาหนะได้แม้ในสภาพแวดล้อมที่ GPS อาจมีความไม่แม่นยำ เช่น ในอุโมงค์หรือระหว่างอาคารสูงนักวิจัยอย่าง Naoki Akai. (2017) และทีมงานจาก Gwangju Institute of Science and Technology ได้นำเสนอการใช้ NDT ร่วมกับเซ็นเซอร์ VLP-32 Lidar ในการทดลองกับยานพาหนะ KIA Soul EV, แสดงให้เห็นถึงความสามารถในการนำทางอย่างแม่นยำในระยะทางกว่า 100 กิโลเมตร การวิจัยล่าสุดจาก Sijia Liu et al. (2021) ยังได้แสดงการประยุกต์ใช้ NDT แบบ realtime ร่วมกับข้อมูลจาก GPS และ IMU ในการระบุตำแหน่งของยานพาหนะ Electric Vehicle ที่ทดลองในมหาวิทยาลัย Wuhan University of Technology, โดยมีความแม่นยำในการระบุตำแหน่งด้วย NDT ในระดับ mean square error ที่ 0.15 เมตร การใช้ NDT และ IMU ในการระบุตำแหน่งยังช่วยลดความต้องการสำหรับการมีสัญญาณ GPS ที่แม่นยำตลอดเวลา, ซึ่งอาจได้รับผลกระทบจากสิ่งกีดขวางต่าง ๆ เช่น ตึกสูงหรือสภาพแวดล้อมในเมือง ข้อมูลจาก IMU สามารถใช้เพื่อปรับปรุงการระบุตำแหน่งของยานพาหนะในเวลาจริงและช่วยเพิ่มความเร็วในการตัดสินใจของระบบขับขี่อัตโนมัติ

2.6 Ros (Robot Operating System)

Robot Operating System (ROS) เป็นเฟรมเวิร์กมาตรฐานสำหรับการพัฒนาซอฟต์แวร์สำหรับหุ่นยนต์ที่ช่วยให้นักพัฒนาสามารถสร้างและทดสอบอัลกอริธึมที่ซับซ้อนในสภาพแวดล้อมจำลองก่อนนำไปใช้งานจริงได้อย่างมีประสิทธิภาพ (Quigley, M., Conley, K., Gerkey, B., 2009)

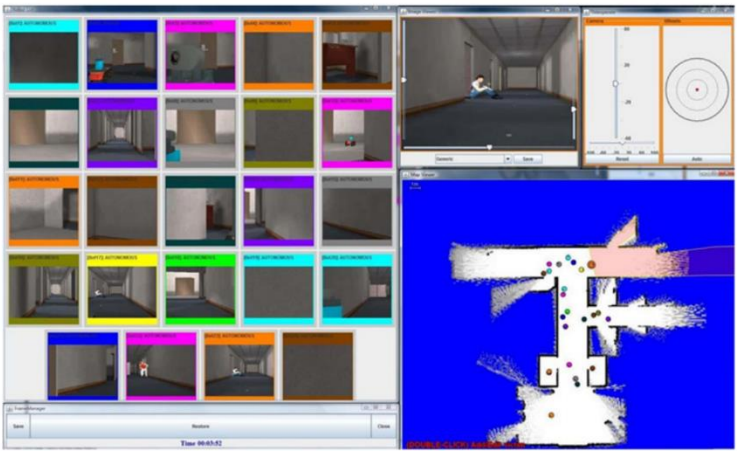
โครงสร้างพื้นฐานของ ROS ประกอบด้วย nodes ที่สื่อสารกันผ่าน messages, services และ actions ซึ่งอำนวยความสะดวกในการเขียนโค้ดที่มีโครงสร้างและแบ่งส่วนได้ชัดเจน



รูปที่ 2.9 A typical ROS network configuration (Quigley, M., Conley, K., Gerkey, B., 2009)

2.6.1 โครงสร้างและการทำงานของ ROS

ROS ประกอบด้วย nodes ที่สามารถสื่อสารกันผ่านการรับส่ง messages ทาง topics, services, และ action servers ซึ่งทำให้เกิดการทำงานร่วมกันได้ในระบบแบบกระจาย (Distributed computing) ช่วยให้หุ่นยนต์มีความยืดหยุ่นและการปรับตัวได้ดีในสภาพแวดล้อมที่แตกต่างกัน (Barber et al., 2011).



รูปที่ 2.10 multi-UV interface with map, camera views, and interaction through selection. (Barber et al., 2011).

2.6.2 ซอฟต์แวร์และไลบรารีใน ROS

ROS มีระบบแพ็คเกจที่อำนวยความสะดวกในการพัฒนาด้วยไลบรารีและเครื่องมือต่าง ๆ เช่น navigation stacks และ perception libraries ซึ่งเป็นแกนหลักในการพัฒนาหุ่นยนต์

อัตโนมัติและระบบการมองเห็นด้วยคอมพิวเตอร์ ROS ได้รับการนำไปใช้ในหลายโปรเจกต์หุ่นยนต์ระดับโลก ตั้งแต่หุ่นยนต์สำรวจดาวอังคารไปจนถึงหุ่นยนต์ทำความสะอาดในบ้าน (Toupet et al., 2020) ความสามารถในการปรับเปลี่ยนและรองรับการทำงานแบบเปิดทำให้ ROS เป็นเครื่องมือที่มีคุณค่าอย่างยิ่งในการพัฒนาและนำไปใช้ในงานวิจัยหุ่นยนต์

2.7 Roboflow

Roboflow เป็นแพลตฟอร์มที่ช่วยนักพัฒนาและนักวิจัยในงานคอมพิวเตอร์วิชั่น (Computer Vision) จัดการกับชุดข้อมูลได้อย่างครบวงจร ตั้งแต่การเตรียมข้อมูล, การติดป้ายกำกับ (Labeling), การทำ Data Augmentation และการสร้างชุดข้อมูลสำหรับฝึกฝนโมเดลปัญญาประดิษฐ์ Roboflow มีจุดเด่นที่ทำให้นักพัฒนาสามารถจัดการข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพ เช่น การรองรับไฟล์รูปแบบหลากหลายและมีเครื่องมือช่วยในกระบวนการจัดการข้อมูล ทำให้เหมาะสำหรับการใช้งานในโปรเจกต์วิจัยที่ต้องการความแม่นยำในการตรวจจับวัตถุ สามารถรวมเข้ากับเฟรมเวิร์กที่ใช้โมเดลปัญญาประดิษฐ์ยอดนิยม เช่น YOLOv8, TensorFlow, และ PyTorch ซึ่งทำให้กระบวนการฝึกฝนโมเดลสามารถทำได้อย่างรวดเร็วและแม่นยำมากขึ้น

การติดป้ายกำกับ (Labeling) Roboflow ช่วยในขั้นตอนการติดป้ายกำกับภาพ โดยการกำหนดกรอบ Bounding Box รอบวัตถุที่ต้องการตรวจจับ เช่น รถยนต์ บุคคล หรือสิ่งกีดขวางต่าง ๆ การกำหนดกรอบ Bounding Box จะช่วยให้โมเดลสามารถเรียนรู้ลักษณะของวัตถุได้อย่างแม่นยำ ขั้นตอนนี้เป็นหนึ่งในกระบวนการที่ใช้เวลามากหากทำด้วยมือ แต่ Roboflow มีอินเทอร์เฟซที่ใช้งานง่ายและสามารถจัดการได้อย่างรวดเร็วผ่านการลากและวางยังมีฟีเจอร์ที่ช่วยในการแบ่งประเภทของวัตถุ และมีการจัดการการติดป้ายกำกับในหลาย ๆ ภาพพร้อมกัน ทำให้สามารถลดเวลาในการเตรียมข้อมูลได้มาก

การทำ Data Augmentation ช่วยเพิ่มความหลากหลายของข้อมูลโดยไม่ต้องเก็บภาพเพิ่มเติม โดยใช้วิธีการหมุนภาพ, พลิกภาพ (Flip), ปรับความสว่าง (Brightness), ขยายหรือย่อภาพ และการปรับแต่งอื่น ๆ ซึ่งจะช่วยให้โมเดลสามารถเรียนรู้จากภาพที่มีความหลากหลายมากขึ้น ส่งผลให้โมเดลมีความทนทานในการทำงานกับข้อมูลใหม่ ๆ ที่ไม่เคยเห็นมาก่อนในสภาพแวดล้อมต่าง ๆ การทำ Data Augmentation มีความสำคัญเป็นพิเศษในงานคอมพิวเตอร์วิชั่นที่ต้องการให้โมเดลสามารถตรวจจับวัตถุในสภาพแวดล้อมที่หลากหลาย เช่น กลางวัน กลางคืน หรือในมุมมองที่ต่างกัน

การจัดการชุดข้อมูล Roboflow ช่วยในกระบวนการจัดการชุดข้อมูลโดยอัตโนมัติ เช่น การแบ่งชุดข้อมูลออกเป็น Training Set และ Test Set โดยทั่วไปแล้ว ข้อมูลจะถูกแบ่งเป็น 80% สำหรับการฝึกฝน (Training) และ 20% สำหรับการทดสอบ (Testing) ซึ่งช่วยให้สามารถตรวจสอบและประเมินประสิทธิภาพของโมเดลในการตรวจจับวัตถุที่ไม่เคยเห็นมาก่อนได้ดีขึ้น การจัดการข้อมูล

เช่นนี้ช่วยให้มั่นใจได้ว่าโมเดลจะไม่เกิดปัญหา Overfitting และมีความสามารถในการทำนายวัตถุใหม่ๆ อย่างแม่นยำ

รูปที่ 2.11 แพลตฟอร์มที่ช่วยในการจัดการข้อมูลสำหรับการเรียนรู้ของระบบคอมพิวเตอร์วิชั่น

Roboflow เป็นเครื่องมือที่มีความสำคัญในกระบวนการเตรียมข้อมูลสำหรับการตรวจจับวัตถุ ซึ่งช่วยให้สามารถจัดการข้อมูลได้อย่างสะดวกและรวดเร็ว Roboflow ยังมีฟีเจอร์สำคัญ เช่น การติดป้ายกำกับ, Data Augmentation, การจัดการชุดข้อมูล, การสร้างเวอร์ชันของชุดข้อมูล และการแปลงรูปแบบข้อมูลที่สนับสนุนการทำงานในงานคอมพิวเตอร์วิชั่นได้อย่างเต็มประสิทธิภาพ