

# ระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่น (Crowd Counting System)

ณัฐวีร์ นริศชาติ, ปิยพันธ์ บุญเลิศอนันต์, สุรัชย์ สันติภาพ, วรรณนิภา อุบลบาน

ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ จังหวัดกรุงเทพมหานครฯ 10800

## Abstract

ระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่น (Crowd Counting System) มีความสำคัญในการช่วยประเมินจำนวนคนในพื้นที่ต่าง ๆ เช่น สนามกีฬา โรงอาหาร หรือสถานที่จัดกิจกรรมที่มีผู้คนแออัด การจัดทำโครงการนี้มีวัตถุประสงค์เพื่อพัฒนาระบบการนับจำนวนคนในสนามบาสโดยใช้เทคนิคการประมวลผลภาพและการเรียนรู้เชิงลึก (Deep Learning) ด้วยโมเดลที่เหมาะสม เช่น YOLOv5, YOLOv8, Mask R-CNN และ Faster R-CNN เพื่อนำมาประมวลผลข้อมูลภาพรวมถึงการนำข้อมูลมาแสดงผลข้อมูลในรูปแบบกราฟ ผลการทดสอบแสดงให้เห็นว่าแต่ละโมเดลที่ใช้มีความแม่นยำและมีประสิทธิภาพที่ต่างกันในแต่ละ นอกจากนี้ยังพบว่าการปรับแต่งพารามิเตอร์ของแต่ละโมเดลสามารถช่วยเพิ่มความแม่นยำในการประมวลผลได้อย่างมีนัยสำคัญ ข้อสรุปจากการทำโครงการนี้ชี้ให้เห็นว่าระบบการนับจำนวนคนที่พัฒนาขึ้นสามารถนำไปใช้งานได้หลากหลายบริบทและมีศักยภาพในการนำไปประยุกต์ใช้ในงานด้านการจัดการฝูงชน ความปลอดภัยสาธารณะ และการวิเคราะห์ข้อมูลเพื่อการตัดสินใจในการจัดสรรพื้นที่ในระดับองค์กร โดยสามารถดูโค้ดการพัฒนาระบบได้จากลิงก์ <https://github.com/mossZmossZ/basketball-court-crowd-counting.git>

## Introduction

การนับจำนวนคน (Crowd Counting) เป็นเทคโนโลยีที่มีบทบาทสำคัญในหลายด้านของสังคมยุคปัจจุบัน ไม่ว่าจะเป็นด้านความปลอดภัยสาธารณะ การจัดการฝูงชน ในสถานการณ์ที่มีผู้คนจำนวนมากอยู่รวมกัน เช่น คอนเสิร์ต สนามกีฬา หรือโรงอาหาร การที่สามารถนับหรือระบุจำนวนคนที่แม่นยำได้มีความสำคัญในการวางแผน

และจัดสรรทรัพยากรและพื้นที่ให้เพียงพอเพื่อความสะดวก ความปลอดภัยของบุคคล ในอดีตการนับจำนวนคน อาจใช้เซนเซอร์พื้นฐานหรือการนับด้วยตนเอง ซึ่งมีข้อจำกัดในด้านความแม่นยำและไม่สามารถรองรับการประมวลผลในพื้นที่กว้างรวมถึงพื้นที่ที่มีความหนาแน่นสูงได้

ปัจจุบันการพัฒนาเทคโนโลยีปัญญาประดิษฐ์ (AI) และการเรียนรู้เชิงลึก (Deep Learning) ได้เปิดทางให้การนับจำนวนคนทำได้อย่างรวดเร็วและแม่นยำมากขึ้น โดยสามารถประมวลผลข้อมูลภาพจากกล้องหรือภาพถ่ายในสถานะที่ซับซ้อนได้อย่างมีประสิทธิภาพ โมเดลการประมวลผลภาพขั้นสูง เช่น YOLOv5, YOLOv8, Mask R-CNN และ Faster R-CNN ได้รับความสนใจในการพัฒนาระบบการนับจำนวนคนเนื่องจากมีความแม่นยำในการตรวจจับและประมวลผลสูง ระบบเหล่านี้สามารถนับจำนวนคนในภาพแบบเรียลไทม์ ทำให้สามารถตอบสนองต่อสถานการณ์ได้ทันที ซึ่งเป็นประโยชน์อย่างยิ่งในกรณีฉุกเฉินหรือการบริหารจัดการในพื้นที่ที่มีการแออัด

แรงจูงใจในการพัฒนาโครงการนี้ คือการสร้างระบบการนับจำนวนคนที่สามารถประมวลผลได้รวดเร็วและแม่นยำ เพื่อให้สามารถใช้ข้อมูลนี้ในการตัดสินใจด้านการจัดสรรพื้นที่รวมถึงการจัดการฝูงชน เพิ่มความสะดวกและความปลอดภัยให้กับบุคคล

## Related Work

---

ตัวอย่างงานวิจัยที่สำคัญคือการใช้ Convolutional Neural Networks (CNNs) ในการนับจำนวนคนจากภาพนิ่งและวิดีโอ ซึ่งงานวิจัยของ Yingying Zhang และทีมวิจัย ได้นำเสนอวิธีการใช้ CNN ในการนับจำนวนคนในภาพที่มีความหนาแน่นสูง โดยใช้ Density Map เพื่อระบุตำแหน่งของคนในภาพ งานนี้ได้สร้างมาตรฐานใหม่สำหรับการนับจำนวนคนโดยไม่ต้องพึ่งพาการตรวจจับวัตถุแต่ละตัวในภาพโดยตรง

อีกหนึ่งงานวิจัยที่มีบทบาทสำคัญคือการพัฒนาโมเดล Single Shot Multibox Detector (SSD) และ You Only Look Once (YOLO) ซึ่งถูกออกแบบมาเพื่อการตรวจจับวัตถุในภาพแบบเรียลไทม์ งานวิจัยของ Joseph Redmon และทีมวิจัย ที่พัฒนา YOLO ได้ทำให้การตรวจจับวัตถุสามารถทำได้ในอัตราความเร็วสูงเหมาะสำหรับการใช้งานในสภาพแวดล้อมที่มีการเปลี่ยนแปลงตลอดเวลา นอกจากนี้ Kaiming He และทีมวิจัย ได้พัฒนา Mask R-CNN ซึ่งเป็นโมเดลที่สามารถตรวจจับและแบ่งขอบเขตของวัตถุในภาพได้อย่างแม่นยำ และมีประโยชน์สำหรับการนับจำนวนคนในสภาพแวดล้อมที่ซับซ้อน

โครงการนี้ได้นำหลักการของ YOLOv5, YOLOv8, Mask R-CNN และ Faster R-CNN มาปรับใช้เพื่อพัฒนาระบบการนับจำนวนคนโดยเน้นที่ความแม่นยำและการประมวลผลภาพ เปรียบเทียบกับงานวิจัยก่อนหน้านี้

โครงการนี้มีความมุ่งมั่นในการประยุกต์ใช้โมเดลหลากหลายและพื้นที่ที่มีความหนาแน่นสูง การเลือกใช้ YOLOv5 และ YOLOv8 ช่วยเพิ่มความเร็วในการประมวลผล ในขณะที่ Mask R-CNN และ Faster R-CNN ถูกใช้ในการเพิ่มความแม่นยำในการระบุตำแหน่งของคนในภาพ การเปรียบเทียบนี้แสดงให้เห็นว่า การใช้โมเดลที่หลากหลายสามารถช่วยพัฒนาระบบการนับจำนวนคนที่มีประสิทธิภาพสูงขึ้น และยังสามารถใช้ในระบบที่หลากหลายได้

## Proposed Method

---

ในโครงการนี้ ระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่น (Crowd Counting System) ถูกพัฒนาขึ้นโดยใช้เทคนิคการประมวลผลภาพและการเรียนรู้เชิงลึก (Deep Learning) เพื่อให้สามารถตรวจจับและนับจำนวนคนได้อย่างแม่นยำในสภาวะแวดล้อมที่มีความหนาแน่นและความแออัดสูง โดยมีวิธีการดังนี้

1. โมเดลที่ใช้ในโครงการ
  - YOLOv5(yolov5xu.pt), YOLOv8(yolov8x.pt), Mask R-CNN และ Faster R-CNN
2. การเตรียมข้อมูลและการสร้าง Dataset
  - ชุดข้อมูลที่ใช้มีดังนี้ COCO (Common Objects in Context) ส่วนข้อมูลภาพจะใช้เป็นภาพนิ่งขนาดไม่น้อยกว่า 800\*800 pixel
3. การทดสอบและประเมินผล
  - ในการทดสอบจะใช้ภาพนิ่งขนาด 800\*800 pixel โดยแต่ละโมเดลจะใช้ภาพเดียวกันในการทดสอบ

## Experiments

---

การตั้งค่าสภาพแวดล้อมและเครื่องมือที่ใช้ในโครงการ (Environment and Tools Setup) มีดังนี้

- Hardware: ดำเนินการบนเครื่อง Windows 11 ที่มี CPU Intel Core i5 และ RAM ขนาด 16GB โดย CPU ถูกใช้เพื่อประมวลผลแบบ inference ของโมเดล ทำให้สามารถประมวลผลภาพความละเอียดสูงได้ในเวลาที่รวดเร็ว
- Software and Libraries: ใช้ Python 3.11.7 และไลบรารีดังนี้ Built-in Libraries (os, sys, time, datetime, sqlite3 และ External Libraries (cv2 (OpenCV), numpy, torch, torchvision, torchvision.ops, matplotlib.pyplot, matplotlib.backends.backend\_qt5agg, PyQt5, PIL (Pillow), ultralytics)

## Results and Discussion



ผลลัพธ์จากทั้งสี่โมเดล YOLOv5, YOLOv8, Mask-R-CNN และ Faster-R-CNN แสดงให้เห็นอย่างชัดเจนว่าแต่ละโมเดลได้แสดงผลการตรวจจับจำนวนคนในสนามบาสเกตบอล ซึ่งมีความแตกต่างกันไป ดังนี้

- ภาพต้นฉบับ: ใช้การนับแบบแมนนวลได้จำนวน 14 คน ใช้เป็นเกณฑ์ในการเปรียบเทียบกับโมเดลอื่น
- YOLOv5: โมเดลนี้ตรวจจับได้ 11 คน ซึ่งทำงานได้ดีในการตรวจจับบุคคลภายในขอบเขตของสนาม แม้ว่าจะนับได้น้อยกว่าการนับแบบแมนนวลเล็กน้อย อาจเป็นเพราะมีการบดบังบางส่วนหรือการจัดการกับขนาดของวัตถุที่เล็กลง
- YOLOv8: ตรวจจับได้ทั้งหมด 5 คน ซึ่งน้อยกว่าทั้งการนับแบบแมนนวลและ YOLOv5 ความแตกต่างนี้ชี้ให้เห็นว่า YOLOv8 อาจมีความสามารถในการตรวจจับที่ต่ำกว่าในบางสถานะที่มีลักษณะภาพเฉพาะ เช่น แสงหรือการวางตำแหน่งของวัตถุ
- Mask-R-CNN: ตรวจจับได้เพียง 1 คน ซึ่งเป็นจำนวนน้อยกว่าคาดการณ์อย่างมาก อาจแสดงถึงข้อจำกัดของโมเดลในการแยกแยะบุคคลออกจากพื้นหลังของสนาม หรือจัดการกับการบดบัง
- Faster-R-CNN: โมเดลนี้ตรวจจับได้ 5 คน ซึ่งใกล้เคียงกับ YOLOv8 แต่ยังคงน้อยกว่าเกณฑ์การนับแบบแมนนวลซึ่งอาจเป็นเพราะมีข้อจำกัดในการตรวจจับวัตถุที่มีขนาดเล็กหรือถูกบดบังบางส่วน

จากผลการทดสอบแสดงให้เห็นว่า YOLOv5 มีประสิทธิภาพดีที่สุดในการตรวจจับทั้งหมดเมื่อเทียบกับโมเดลอื่น แม้ว่าจะนับได้น้อยกว่าจำนวนจริง ความแตกต่างของผลการตรวจจับแต่ละโมเดลสะท้อนถึงความสามารถที่ต่างกันภายใต้สภาวะแวดล้อมนี้ YOLOv8 แม้ว่าจะเป็นโมเดลที่มีความสามารถ แต่จำเป็นต้องปรับแต่งเพิ่มเติมเพื่อเพิ่มความแม่นยำในการตรวจจับในสถานการณ์เฉพาะนี้ ผลการตรวจจับของ Mask-R-CNN ที่ต่ำกว่าคาดการณ์ อาจแสดงให้เห็นว่าโมเดลนี้ไม่เหมาะสมสำหรับภาพประเภทนี้หากไม่มีการปรับแต่งเพิ่มเติมหรือการฝึกฝนด้วยข้อมูลที่เจาะจง

ข้อจำกัดสำคัญในงานทดลองนี้คือประสิทธิภาพของโมเดลที่แตกต่างกันในการตรวจจับบุคคลได้อย่างแม่นยำ โดยเฉพาะในสถานการณ์ที่มีการบดบังหรือขนาดของวัตถุที่ต่างกัน การปรับปรุงในอนาคตอาจรวมถึงการปรับแต่งโมเดลด้วยข้อมูลเฉพาะทาง หรือการทดลองใช้งานสถาปัตยกรรมที่แตกต่างกันเพื่อเพิ่มความแม่นยำในการตรวจจับในภาพที่มีลักษณะเฉพาะเช่นสนามกีฬา

## Conclusions

โครงการระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่น (Crowd Counting System) ที่พัฒนาขึ้นนี้ได้รับการออกแบบเพื่อแสดงถึงความหนาแน่นของกลุ่มคนที่ใช้งานสนามบาส แรงจูงใจในการทำโครงการนี้มุ่งเน้นไปที่การตรวจจับและนับจำนวนคนที่เข้าใช้งานสนามบาสในมหาวิทยาลัยในวันและเวลาต่าง ๆ โดยการใช้โมเดล YOLOv5, YOLOv8, Mask R-CNN และ Faster R-CNN ผลการทดสอบแสดงให้เห็นว่าในแต่ละโมเดลมีความแม่นยำที่ค่อนข้างแตกต่างกัน รวมถึงยังมีความผิดพลาดในแต่ละโมเดลอยู่โดยดูได้จากผลการทดลอง

เมื่อเปรียบเทียบกับงานวิจัยก่อนหน้านี้ ระบบที่พัฒนาขึ้นนี้อาจมีความแม่นยำที่ต่ำกว่า โดยเฉพาะในการตรวจจับบุคคลที่ทับซ้อนกัน รวมถึงวัตถุรอบข้าง

ทิศทางในอนาคตสำหรับโครงการนี้อาจมุ่งเน้นการปรับปรุงโมเดลแต่ละตัว เพื่อให้มีความแม่นยำขึ้น รวมถึงการทำให้ระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่นแบบเรียลไทม์

## Contributions

ในการทำโครงการระบบตรวจจับจำนวนคนในสภาพแวดล้อมหนาแน่น (Crowd Counting System) สมาชิกในทีมแต่ละคนมีส่วนร่วมในการทำงานอย่างสมดุลและมีบทบาทสำคัญในแต่ละขั้นตอนของโครงการดังนี้

1. ณัฐวีร์ นริศชาติ และ ปิยพันธ์ บุญเลิศอนันต์: มีหน้าที่รับผิดชอบในส่วนของการศึกษาเกี่ยวกับโมเดลที่ใช้ในโครงการการนำโมเดลมาใช้และการ implement Code ต่างๆ
2. สุรัชย์ สันติภาพ และ วรรณนิภา อุบลบาน: มีหน้าที่ในการนำข้อมูลไป visualization ออกแบบหน้าต่างในการแสดงผลข้อมูลเป็นกราฟ รวมถึงการ Coding

ทุกคนมีส่วนร่วมในแต่ละขั้นตอนของโครงการอย่างเท่าเทียม โดยมีการประชุมร่วมกันเพื่อแบ่งปันข้อมูลและวิเคราะห์ผลการทดลองอย่างต่อเนื่อง ทำให้โครงการนี้เกิดจากการทำงานร่วมกันของทีมงานทั้งหมด

## References

---

- [1] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao and Yi Ma, “Single-Image Crowd Counting via Multi-Column Convolutional Neural Network”, CVPR, 2016.
- [2] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, CVPR, 2016.
- [3] Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross Girshick, “Mask R-CNN”, ICCV, 2017.
- [4] Deepak Babu Sam, Skand Vishwanath Peri, Mukuntha Narayanan Sundararaman, Amogh Kamath and R. Venkatesh Babu, “Locate, Size and Count: Accurately Resolving People in Dense Crowds via Detection”, cs.CV, 2020.
- [5] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors,” ICCV, 2005.

---