SUPPLY CHAIN AND SUSTAINABILITY RESEARCH: SCSR
VOL.3, NO.3; JUL. – SEP. ; 2024, ISSN 2822-0412 (Online)

Envisioning the Future of Supply Chain Transportation Management-Exploring Crowd Density Image Recognition with Deep Learning

En-Fu Li *
Cheng-Chung Lee**
Guan-Hong Chen***

Received: June 14, 2024 / Revised: December 9, 2024 / Accepted: December 13, 2024

Abstract

Over the past few decades, artificial intelligence (AI) technology has made significant progress, especially in image recognition and natural language processing. Deep learning, as a multi-level artificial neural network architecture, can automatically identify and learn data features through large amounts of data to achieve high-precision prediction and identification. This research aims to explore the practical application of deep learning in the field of image recognition, and designs an automated image recognition system, taking the identification of crowd density in a train carriage as an example. The system can accurately detect the number of passengers in each carriage, thereby optimizing urban rail transit strategies, ensuring even distribution of passengers and providing the best commuting experience. This technology also has important implications for supply chain management. Accurate people flow identification technology can improve resource utilization efficiency, optimize logistics resource allocation, and reduce transportation time and costs through real-time monitoring and prediction of people flow density. It can also improve demand forecasting and analyze historical data through deep learning algorithms to help companies predict demand more accurately and avoid excessive transportation and shared traffic. This technology enhances supply chain transparency and promotes sustainable development by monitoring the entire transportation process in real time. These applications can improve the operational efficiency of the transportation system and significantly improve the quality and efficiency of supply chain management, demonstrating the huge potential of deep learning technology in modern supply chain management.

Keywords: Artificial Intelligence, Deep Learning, Image Recognition, Machine Vision, Supply Chain

Corresponding*, Student, Department of Computer Science and Information Engineering, National United University, Miaoli Taiwan., E-mail: L0958242580@icloud.com

**, Assistant Professor, Department of Marketing and Logistics Management, Yu Da University of Science and Technology, Miaoli Taiwan

***, Student, Department of Computer Science and Information Engineering, National United University, Miaoli Taiwan

Introduction

Preamble

The fusion of data and technology has marked a new revolutionary turning point. Among these, artificial intelligence (AI) has showcased its revolutionary potential across various domains. From healthcare and finance to every aspect of daily life, AI technology is profoundly impacting the way we work and live.

In recent years, especially deep learning technology, it has exhibited unparalleled capabilities in handling complex image and language data. These advancements are not just theoretical breakthroughs but have changed the way we approach problems in real-world applications. For instance, utilizing deep learning to optimize metropolitan rail transport strategies can fundamentally enhance the commuting experience for passengers.

However, the genuine integration of these technologies into practical applications and the realization of their potential value remains a significant challenge for many researchers and engineers. In light of this, this research delves deep into the practical applications of deep learning in the field of image recognition, taking the identification of human density within train carriages as a case study. The aim is to offer fresh perspectives and solutions for the optimization of metropolitan rail transport strategies.

Research Background

With the accelerated development of global informatization and digitization, data has become the core asset of modern society. Among these, image data, due to its vast informational content and intricate structure, poses massive challenges for automatic recognition and analysis. In the past, traditional image processing techniques could extract basic features and recognize images, but their performance was often limited in highly complex and dynamically changing environments.

In recent years, deep learning has spearheaded a new wave of artificial intelligence, especially in the domain of image recognition, demonstrating capabilities that surpass traditional methods. Yet, translating these advanced techniques into practical applications still requires overcoming various technical and engineering obstacles.

Rail transportation, being a crucial pillar of urban development, faces burgeoning passenger flows with accelerating urbanization. Ensuring a balanced distribution of passengers within carriages to provide a comfortable commuting experience has become an urgent demand in the transport industry. Exploring the practical applications of deep learning in this domain holds profound academic significance and offers novel solutions for the real-world operations of urban rail transport.

Research Motivation

With global rapid development, rail transportation has increasingly become the central transportation infrastructure of cities, where its efficiency directly affects the quality of life for urban residents and the economic vitality of the city. However, during peak transportation hours, drastic variations in human density within carriages lead to diminished passenger experiences and unbalanced utilization of transport resources. This not only inconveniences passengers but also poses challenges to the operational efficiency and safety of metropolitan rail systems.

Traditional methods of crowd monitoring and analysis often rely on manual counts or rudimentary sensor technologies, such as CCTV, infrared sensors, or manual tallying. These methods struggle to achieve real-time, high-precision analysis when faced with large-scale, high-frequency, and dynamically changing crowd data. Hence, there's a need for a method that can automatically and accurately identify and analyze human density within train carriages.

In recent years, deep learning has achieved astonishing feats in the realms of image recognition and analysis. Its potent feature extraction and pattern recognition capabilities make it an ideal choice for optimizing metropolitan rail transport strategies. This research, grounded in deep learning, explores its potential in recognizing human density within train carriages, aspiring to offer a more efficient and intelligent solution for existing transport strategies. The ultimate goal is to enhance the commuting experience for passengers and boost the operational efficiency of rail transportation.

The benefits of supply chain management and the application of deep learning

In contemporary business operations, an effective supply chain management system is an indispensable cornerstone to support daily business. It is directly related to whether a company can stay ahead in market competition. The lack of effective supply chain management puts businesses at risk of operational disruption, underscoring the importance of supply chain management. The challenge for companies is whether to continue maintaining outdated systems or respond to rapid technological and social changes and build a supply chain that can grow, modernize, and digitally optimize with the changing times.

Optimizing supply chain management can bring the following main benefits:

Improve productivity: Through enterprise asset equipment management systems and predictive maintenance, we can improve the operating efficiency of machines and systems, solve production bottlenecks, and improve work processes, thereby significantly improving productivity. Leveraging deep learning technology to automate processes and data analysis can speed up product shipping and delivery times.

Reduce supply chain costs: Deep learning technology can perform accurate predictive analysis, effectively eliminate "uncertainty" and reduce the risk of inventory waste or shortages. The Internet of Things (IoT) combined with deep learning provides flexible asset management and optimized workflows.

Improve product quality: Combining customer feedback with deep learning analysis, the R&D team can more accurately grasp market demand and effectively improve product design to meet consumer expectations and trends.

Improve customer service: Modern supply chain management should be customer-centric. Deep learning technology makes the supply chain more flexible and adaptable, providing personalized customer experience and improving customer satisfaction.

Improve transparency and sustainability: Deep learning technology can provide transparency at every stage from product design to logistics, helping companies work with suppliers to improve their environmental footprint and achieve sustainable development goals.

In this study, deep learning technology is particularly applied to image recognition of crowd density. This technology can not only improve the passenger flow distribution of the rail transit system, but also provide real-time data support for logistics distribution in supply chain management, further optimizing the entire supply chain. efficiency and response speed.

Literature Review

Artificial Intelligence (Development and Current Status of AI)

Since its introduction in the 1950s, Artificial Intelligence (AI) has experienced several ups and downs. In its early stages, researchers explored rule-based systems, such as expert systems, to mimic human thought processes. Due to hardware limitations and the complexity of knowledge representation, initial AI research progressed slowly. As computational power improved, data-driven AI research became dominant. In the early 21st century, with the rise of big data technologies and a significant boost in hardware computational capacity, the field of AI experienced rapid growth. In today's AI domain, its applications have permeated all aspects of life, from voice assistants and facial recognition to self-driving cars. Particularly in image recognition and natural language processing, AI has not only achieved performances surpassing human levels but also attained commercial value in various real-world scenarios (Russell & Norvig, 2020 & LeCun et al., 2015).

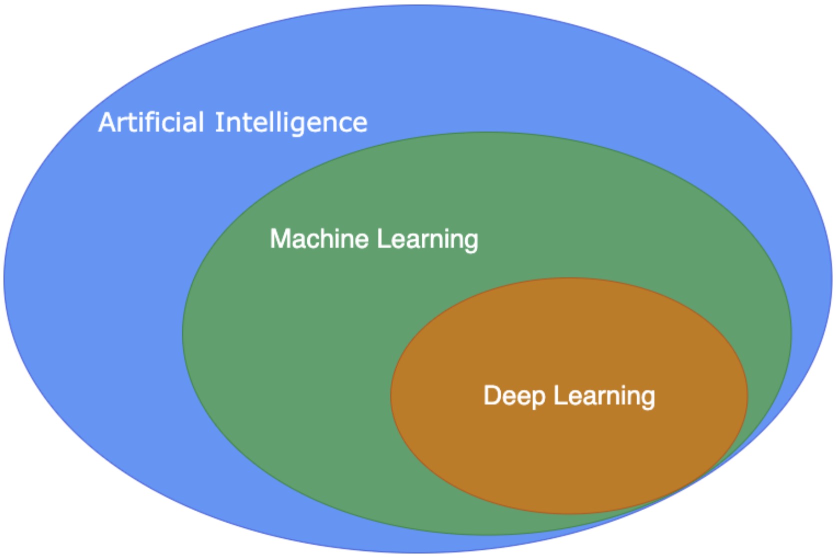

Figure1 Machine Learning Relationship Diagram

Machine learning

Machine Learning is a subfield of AI, centered on enabling computers to perform specific tasks using data and algorithms without explicit instructions. Over the past few decades, machine learning has become the driving force behind many technological and commercial applications. It involves building models that learn from data to make predictions or decisions (Goodfellow et al., 2016).

Supervised Learning is the most common learning strategy in machine learning. In this approach, we provide the model with input data and corresponding output results, enabling it to learn predictions for unknown outputs from these paired data. In other words, supervised learning requires a labeled dataset where each data sample has a predefined label or result. The goal is to discover the relationship between inputs and outputs to predict new, unseen data. Common supervised learning tasks include classification (e.g., categorizing emails as "spam" or "non-spam") and regression (e.g., predicting house prices based on their features).

Unsupervised Learning differs from supervised learning as it doesn't rely on predefined output labels. Instead, it focuses on the underlying structure or patterns in the data. It seeks to automatically extract and learn useful information or features from input data without explicit output labels. This makes unsupervised learning particularly suitable for exploratory data analysis as it can reveal hidden structures in data. Typical unsupervised learning tasks are clustering (e.g., segmenting customers into different market segments) and dimensionality reduction (e.g., projecting high-dimensional data into a two-dimensional space for visualization (Murphy, 2012).

The primary distinction between supervised and unsupervised learning lies in whether explicit output labels guide the learning process. Supervised learning focuses on predicting precise output results, whereas unsupervised learning emphasizes discovering hidden data structures or patterns.

Deep Learning

Originating from artificial neural networks, Deep Learning has surged in popularity in recent years due to enhanced computational capabilities and the availability of vast datasets. This learning strategy mimics the functioning of neurons in the human brain, using multi-layered neural structures to model and abstract data.

In practical applications, deep learning has been successfully applied to voice recognition, image processing, and natural language processing. Its success lies in its ability to capture intricate patterns in vast datasets. As the depth of the model increases, it can capture more abstract and complex features (Hinton et al., 2006).

However, deep learning also presents challenges. These models require vast amounts of data for training, limiting their potential in data-scarce scenarios. Moreover, deep learning models often demand substantial computational resources, which can restrict their application in some situations (Schmidhuber, 2015).

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are designed specifically for handling image and video data within deep learning. Unlike traditional fully connected neural networks, CNNs use convolutional operations to process data, allowing them to capture spatial structures in images (Krizhevsky et al., 2012).

Basic components of CNNs include convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract features from input images, pooling layers reduce the spatial dimensions of features, and fully connected layers are used for classification and regression tasks.

CNNs have achieved remarkable results in various fields, especially in image recognition, face recognition, and object detection. Their efficiency largely stems from their ability to capture relevancy within local regions and identify features at different levels.

At the same time, CNNs face challenges. Deep CNN models can contain millions or even billions of parameters, making the training process both time-consuming and resource-intensive. Moreover, like all deep learning models, the interpretability of CNN models remains an ongoing issue.

YOLO

Object detection has always been a core research topic in the field of computer vision. Earlier methods, such as the R-CNN series, required multiple analyses of images or their segments. YOLO (You Only Look

Once) offers a faster and more efficient detection method. It assesses the entire image in just one pass, treating the detection task as a form of regression (Redmon et al., 2016)

What's unique about YOLO is that it doesn't rely on traditional sliding windows or region proposals. Instead, it directly evaluates the whole image. This means that during training and testing, it can fully understand all the contents of an image, enabling it to better recognize global classes and features. Compared to systems that only capture local information, like Fast R-CNN, YOLO is more precise in differentiating between backgrounds and objects, with relatively fewer false positives.

Moreover, YOLO has demonstrated strong generalization capabilities. When trained on real-world images, its performance often surpasses that of models like DPM and R-CNN when evaluating images from specialized domains like artwork. This suggests that YOLO can maintain good recognition performance even when faced with images it hasn't been trained on (Liu et al., 2016).

Despite these advantages, YOLO still lags behind some advanced detection systems in terms of accuracy. While it can locate objects quickly, it sometimes struggles with the precise positioning of smaller objects.

In conclusion, YOLO is an efficient object detection model that excels in real-time processing and generalization. However, in certain situations, it might compromise on accuracy.

While YOLO may not be as accurate as variants of R-CNN, it significantly reduces the chances of background misidentification. By sacrificing a bit of accuracy in favor of reducing false positives and increasing speed, it becomes a very practical choice in many real-world considerations.

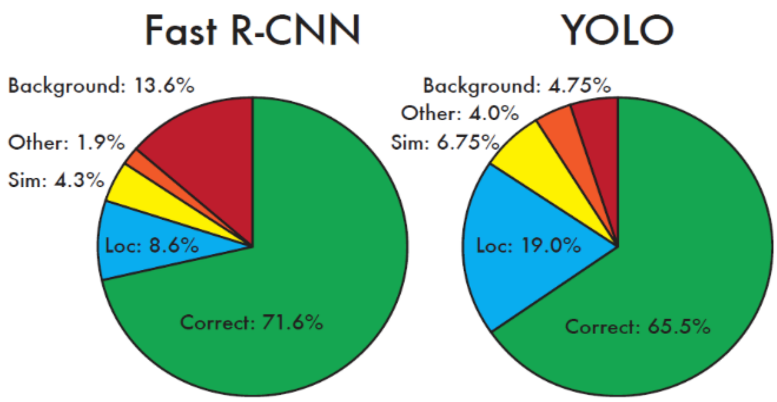

Figure 2 Comparison between CNN and YOLO

Train Passenger Flow Analysis System

In recent years, with the rapid development of smart technologies, various fields have begun to harness the power of these technologies to enhance operational efficiency. Among them, the "Train Passenger Flow Analysis System" has emerged as an excellent application. This technology, through advanced data analysis and real-time recording, can accurately count and effectively manage passenger flow. In train stations and on trains, the application of the passenger flow analysis system has become a crucial aspect of optimizing the transportation experience. Below are four major advantages of applying the passenger flow analysis system to trains (Yang et al., 2021; Lu et al., 2018 & Behrooz et al., 2022).

Advantage 1: Enhanced Safety

The Train Passenger Flow Analysis System can monitor the flow of people inside and outside the station in real-time, especially during peak hours, helping to prevent overcrowding situations. The system can quickly detect areas with excessive passenger flow and alert station management through a warning system, thus avoiding potential safety risks.

Advantage 2: Reduced Wait Times for Passengers

With the Train Passenger Flow Analysis System, transportation management can accurately predict the passenger flow at stations and allocate an appropriate number of trains according to different times, reducing the waiting time for passengers on platforms. This not only enhances passenger satisfaction but also makes more efficient use of train resources.

Advantage 3: Efficient Management of Onboard Facilities

The Train Passenger Flow Analysis System can also monitor passenger flow within the train. Based on the flow in different carriages, the system can adjust facilities such as air conditioning and seat arrangements, offering a more comfortable travel experience. Moreover, the system can detect overcrowded areas within the train and guide passengers to less crowded carriages, ensuring a balanced distribution of people within the train.

Advantage 4: Flexible Train Operations

With the Train Passenger Flow Analysis System in place, transportation management can adjust train operations more flexibly. Based on special events, holidays, and other scenarios, the system can predict peak flows, allowing for reasonable allocation of transportation resources and ensuring the smooth operation of stations and trains.

The application of the Train Passenger Flow Analysis System not only elevates the management efficiency of train transportation but also provides passengers with a safer and more convenient travel experience. The introduction of this technology will effectively revolutionize the operational model of train transportation, making the entire system more intelligent and user-friendly.

The relationship between supply chain and deep learning people flow management

With the accelerated development of global informatization and digitization, data has become the core asset of modern society. Among these data, due to their rich information content and complex structure, automatic recognition and analysis of image data are extremely challenging for traditional technologies. In the past, traditional image processing technology could extract basic features and perform image recognition, but its performance was often limited in highly complex and dynamically changing environments.

In recent years, deep learning has set off a new trend in the field of artificial intelligence, especially in the field of image recognition, showing its ability to surpass traditional methods. However, translating these advanced technologies into practical applications still requires overcoming various technical and engineering obstacles.

As an important part of modern enterprise operations, supply chain management is facing growing demands and challenges. The effective operation of the supply chain is crucial to an enterprise's production efficiency, cost control and market competitiveness. With the acceleration of urbanization, transportation, as a key link in supply chain management, is also facing an increasing passenger flow. How to use advanced technological means to optimize passenger distribution and improve transportation efficiency has become an urgent problem to be solved.

In this context, accurate crowd flow prediction and identification have become particularly important. Using deep learning technology, accurate prediction and real-time monitoring of the flow of people in the carriage can be achieved, thereby reducing the imbalance of population transportation and improving transportation efficiency.

Supply chain management challenges include:

Demand forecast accuracy:

Supply chains need to accurately predict population transportation needs to avoid overloads or shortages. Deep learning technology can provide more accurate demand forecasts by analyzing historical data. The same technology can also be used to predict the number of passengers in the carriage at different times and optimize passenger distribution.

Efficiency of logistics and distribution:

In the supply chain, logistics and distribution are key links. Deep learning can optimize logistics routes, reduce transportation time and costs, and improve overall distribution efficiency. People flow recognition technology can help optimize the distribution of people flow in the carriage, reduce crowding, and improve passenger travel efficiency and comfort.

Optimal allocation of resources:

Effective supply chain management requires flexible allocation of resources to respond to rapid changes in the market. Deep learning technology can provide real-time data analysis to help managers make more informed decisions. Through real-time people flow data analysis, transportation resources can be flexibly adjusted to avoid resource waste.

Supply chain transparency and sustainability:

Modern supply chains require transparency from production to logistics and consideration of environmental impact. Deep learning technology can provide comprehensive monitoring and analysis to promote the sustainable development of the supply chain. Through real-time monitoring and analysis of people flow, the transparency of the transportation process can be improved, further promoting environmental protection and energy saving.

Research Framework and Methodology

Research Framework

This study aims to utilize deep learning techniques for image recognition of pedestrian density. The framework employed is based on the YOLOv5 object detection model, trained and tested using the COCO dataset. The primary research structure consists of five steps: dataset preparation, model construction, model training, model testing, and result analysis, as illustrated in the figure.

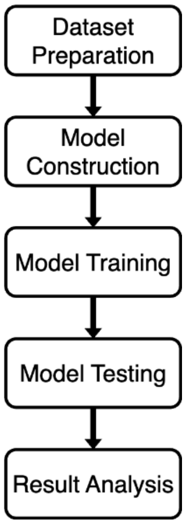

Figure 3 Research Architecture Diagram

Data Collection

For training, this study uses Microsoft's COCO dataset. This dataset provides comprehensive training data for machine learning and utilizing an existing dataset saves significant time.

COCO Dataset

The COCO dataset is one of the most commonly used datasets for machine learning, especially in areas like object detection, instance segmentation, and human key point detection.

Research Software and Tools

Development Environment

MacOS (Sonoma14.0 Beta version)

The advantages of using macOS as a development system include:

Stability: macOS is known for its exceptional stability, which is crucial for machine learning models that need to run for extended periods.

Unix Foundation: As macOS is based on Unix, it offers developers powerful terminal tools and environments. These tools simplify tasks like data processing, script execution, and many others.

Deep Learning Framework

Since real-time object detection is our goal, YOLO is the most widely used object detection method. We use YOLOv5, the latest version in the YOLO series, which strikes a good balance between speed and accuracy.

YOLOv5 utilizes the PyTorch deep learning framework, an open-source machine learning library developed by Facebook's AI Research lab (FAIR).

Model Optimization Method

The original COCO dataset is vast, with many images unrelated to humans, leading to excessive training time and resource consumption. Thus, we leveraged the labels provided in the dataset to extract images containing humans, optimizing the training efficiency of the model.

Model Selection

Among the many models in the YOLOv5 family, YOLOv5s is the most lightweight. As the model size increases, there are more convolution layers, implying higher computational costs, such as increased processing time. Hence, YOLOv5s is better suited for this study.

Training and Validation Strategy

At the start of training, we meticulously pre-processed the COCO dataset, specifically selecting images containing humans. We divided the data into 80% training, 10% validation, and 10% testing sets. Regarding training parameters, we chose a batch size of 16, ensuring computational efficiency and stable learning. We also set the total training epochs to 100, based on a preliminary assessment to ensure comprehensive model training and avoid overfitting. 3.7 Model Testing and Result Analysis

We evaluated the model using the previously reserved test set to ascertain its performance on unseen data. With the image detection program, each identified person is clearly marked, and the total number of detected individuals is displayed in the top-left corner of the image. Preliminary results indicate the model has commendable recognition abilities, but further in-depth evaluation is necessary for specific scenarios.

Training Implementation and Analysis

Dataset Preparation

The dataset used for training and testing is the COCO dataset provided by Microsoft. The ultimate purpose of the experiment is to identify the number of people, so we only used data labeled as "person" from the dataset.

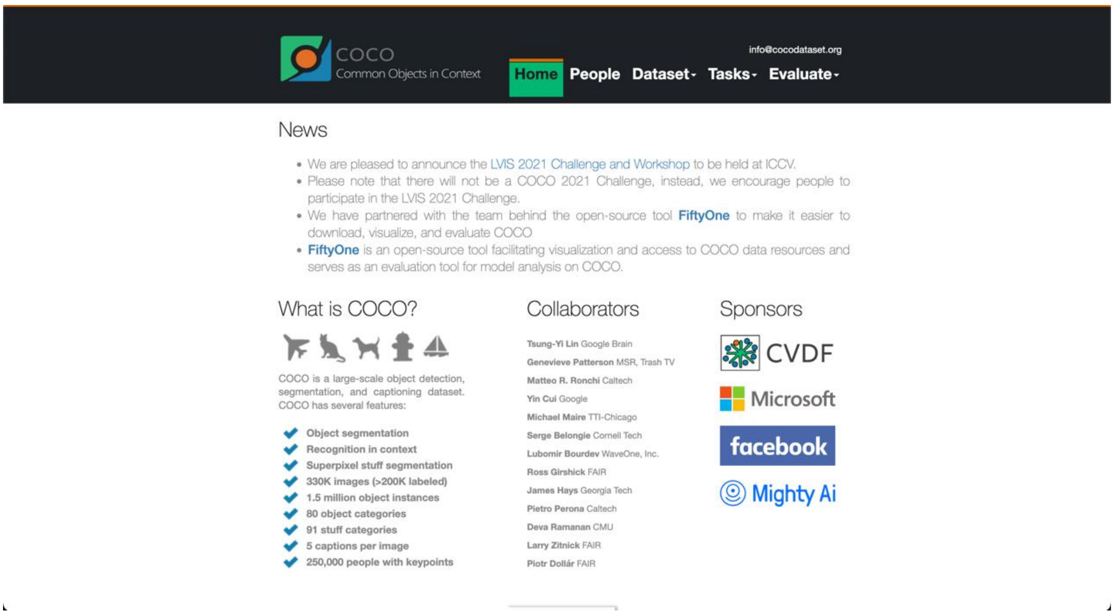

Figure 4 COCO Dataset Website

Dataset Processing

The original COCO dataset contains as many as 160,000 images. However, a large portion of these photos is irrelevant to people. Since our model only needs to recognize people, we extracted photos with labels related to people from the original dataset for training.

```
import json
import shutil
import os

# 設置路徑
coco_annotations_path = '/Users/chenguanhong/Documents/yolov5_training/annotations/instances_train2017.json'
images_dir = '/Users/chenguanhong/Documents/yolov5_training/train2017/'
output_dir = 'coco_people/'

# 讀取標註文件
with open(coco_annotations_path, 'r') as f:
    data = json.load(f)

# 創建輸出目錄    (variable) output_dir: Literal['coco_people/']
if not os.path.e
    os.makedirs(output_dir)

# 提取只有人的圖片
for image_info in data['images']:
    image_id = image_info['id']
    image_annotations = [ann for ann in data['annotations'] if ann['image_id'] == image_id]

    if all(ann['category_id'] == 1 for ann in image_annotations) and len(image_annotations) > 0:
        image_filename = image_info['file_name']
        input_image_path = os.path.join(images_dir, image_filename)
        output_image_path = os.path.join(output_dir, image_filename)
        shutil.copy(input_image_path, output_image_path)
```

Figure 5 A program for extracting photos labeled as 'person

The benefit of this approach is a reduction in the dataset from the initial 160,000 photos to 6,000 photos, substantially reducing the model's training time.

For model training, in addition to providing photos, we need to provide .txt files indicating the positions of people in each photo. This is to train the model to recognize the position of people in the training photos. The content format of the record file is: person's category in the COCO dataset, x-coordinate of the person in the photo, y-coordinate of the person in the photo, width of the person, and height of the person.

```
000000001464.txt
0 0.666406 0.382904 0.251563 0.302108
```

Figure 6 The content of the log file

Model Creation

YOLOv5 officially provides four different model versions, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5x has the highest accuracy but the slowest processing speed. Considering the need for fast processing speed in real-time video recognition, YOLOv5s was chosen for training and subsequent recognition.

Model Training

Firstly, the YOLOv5 official model should be downloaded from GitHub.

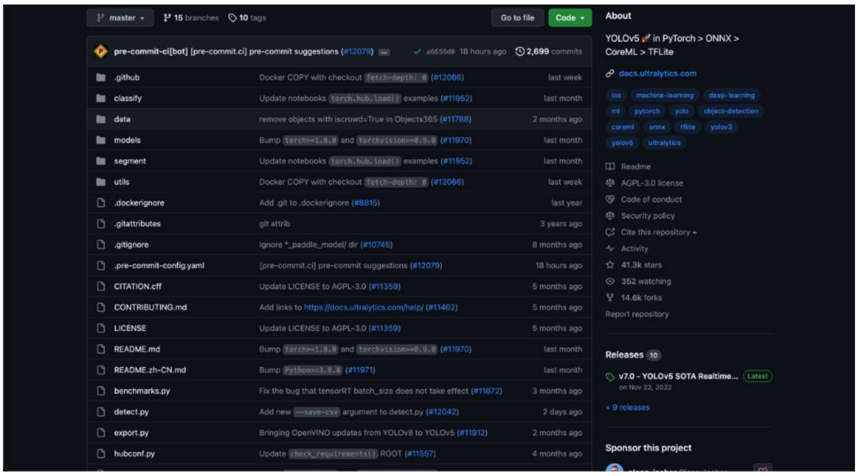

Figure 7 Official YOLOv5 GitHub

Next, download the required dataset from the COCO dataset official website. For this purpose, we selected the 2017 Train Image as the primary dataset because, compared to earlier versions, the 2017 training set contains more photos.

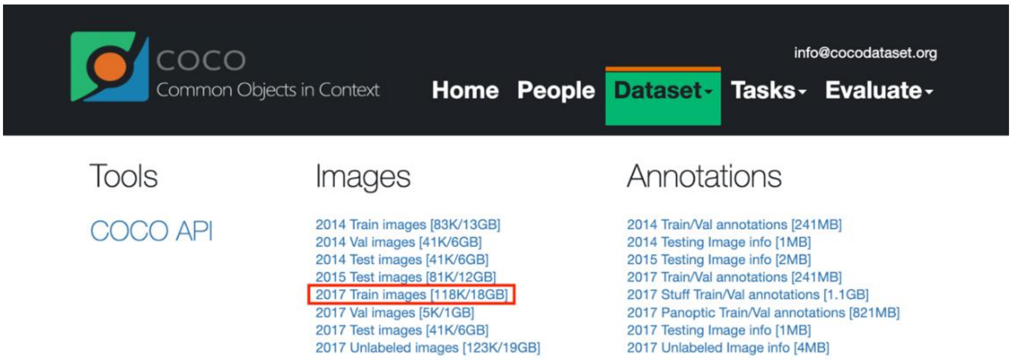

Figure 8 2017 Dataset

The downloaded training set folder should be placed together with the model files. The 'images' folder contains the training photo set, and the 'labels' folder contains the labels for people.

The command to execute the training program is:

python3 train.py -- img 640 --batch 16 --epochs 100 --data

/Users/chenguanhong/Documents/yolov5_training/dataset.yaml --cfg yolov5s.yaml --weights yolov5s.pt

This includes specifying the training image size as 640x640, the batch size as 16, and the number of iterations as 100. Further training command parameters are illustrated in the figure below.

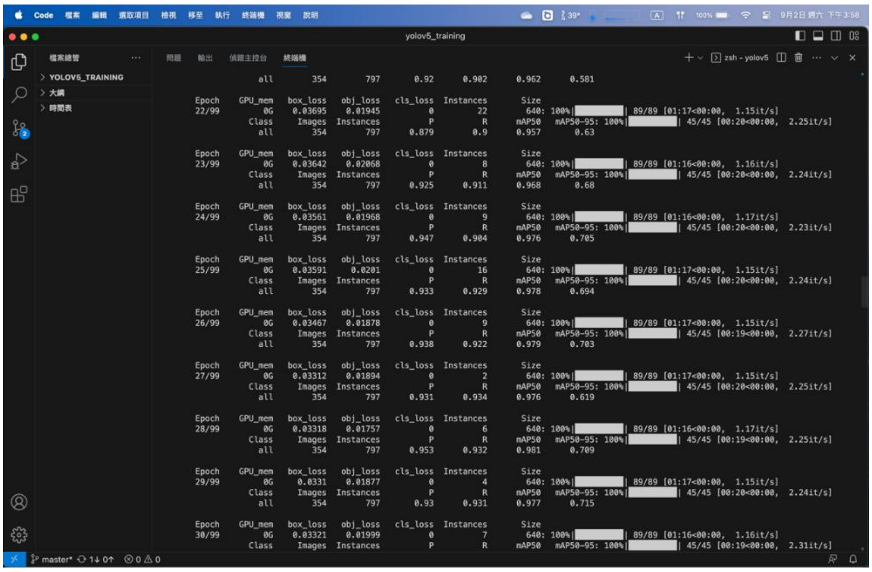

Figure 9 Training Model Process 1

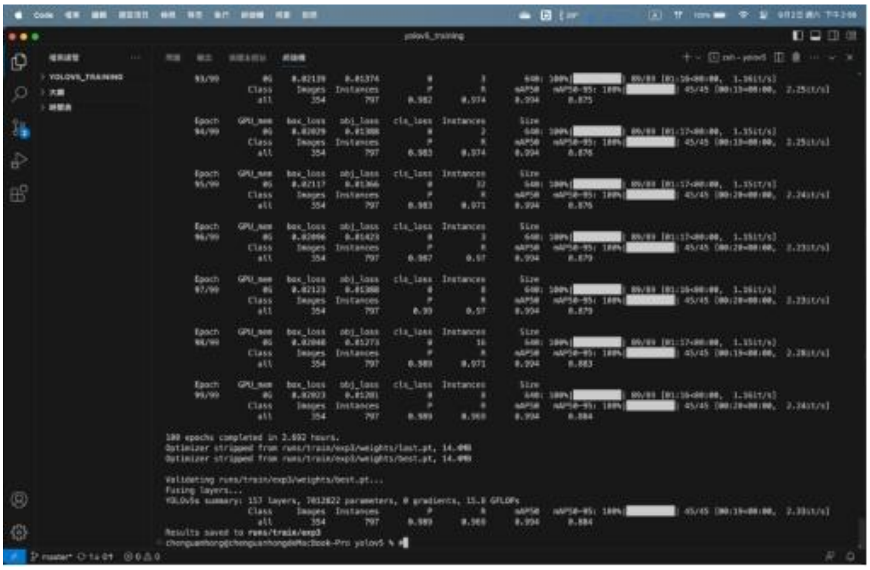

Figure 10 Training Model Process 2

Upon completion, a folder will be generated containing the weight file from this training for subsequent testing. It also contains data charts visualizing the training process.

The 'box_loss' represents the mean GIoU loss function value, where a smaller value indicates higher accuracy in bounding box predictions.

The 'obj_loss' indicates the mean objective detection loss value; the smaller the value, the more accurate the object detection.

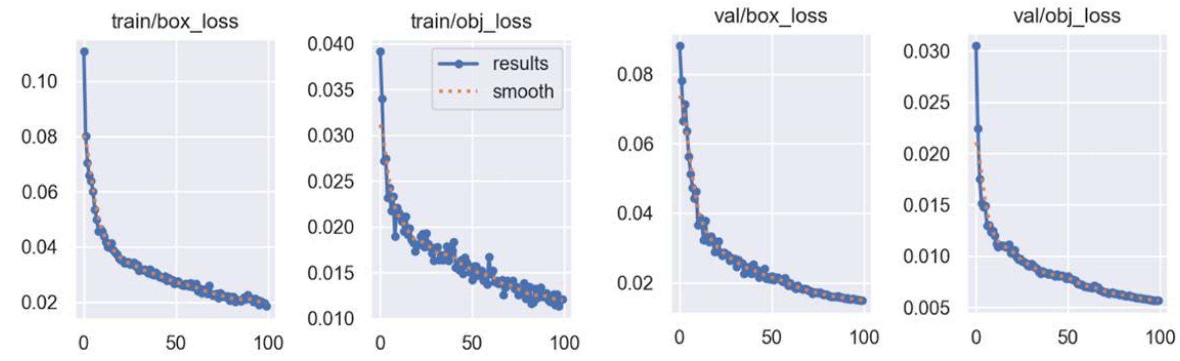

Figure 11 Training Results

Model Testing

The official detect.py is used for detection. Since we need to display the total detected people count, we made some additions to detect.py. The following command is then entered:

python3 detect.py --weights runs/train/exp2/weights/best.pt —source/Users/chenguanhong/Desktop/1.jpg

The --weights parameter indicates the weight file generated from model training, while the --source parameter indicates the path of the image to be tested. Setting the --source parameter to 0 allows real-time recognition.

```
    if int(cls) == 0 and conf > 0.01:
        num_people_in_image += 1

total_people += num_people_in_image

font = cv2.FONT_HERSHEY_SIMPLEX
font_scale = 3
font_thickness = 3
cv2.putText(im0, f"People in Image: {num_people_in_image}", (10, 70), font, font_scale, (0, 255, 0), font_thickness)
```

Figure 12 Added content to detect.py

```
chenguanhong@chenguanhongdeMacBook-Pro yolov5 % python3 detect.py --weights runs/train/exp5/weights/best.pt --source /Users/chenguanhong/Desktop/1.jpg
detect: weights=['runs/train/exp5/weights/best.pt'], source=/Users/chenguanhong/Desktop/1.jpg, data=data/coco128.yaml, imgsz=[640, 640], conf_thres=0.25, iou_thres=0.45, max_det=1000, device=, view_img=False, save_txt=False, save_conf=False, save_crop=False, nosave=False, classes=None, agnostic_nms=False, augment=False, visualize=False, update=False, project=runs/detect, name=exp, exist_ok=False, line_thickness=3, hide_labels=False, hide_conf=False, half=False, dnn=False, vid_stride=1
YOLOv5 🚀 v7.0-212-g9974d51 Python-3.11.3 torch-2.0.1 CPU

Fusing layers...
YOLOv5s summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
image 1/1 /Users/chenguanhong/Desktop/1.jpg: 480x640 6 persons, 64.8ms
Speed: 1.0ms pre-process, 64.8ms inference, 5.8ms NMS per image at shape (1, 3, 640, 640)
Results saved to runs/detect/exp
```

Figure 13 Detection Process

Figure 14 Detection Results1

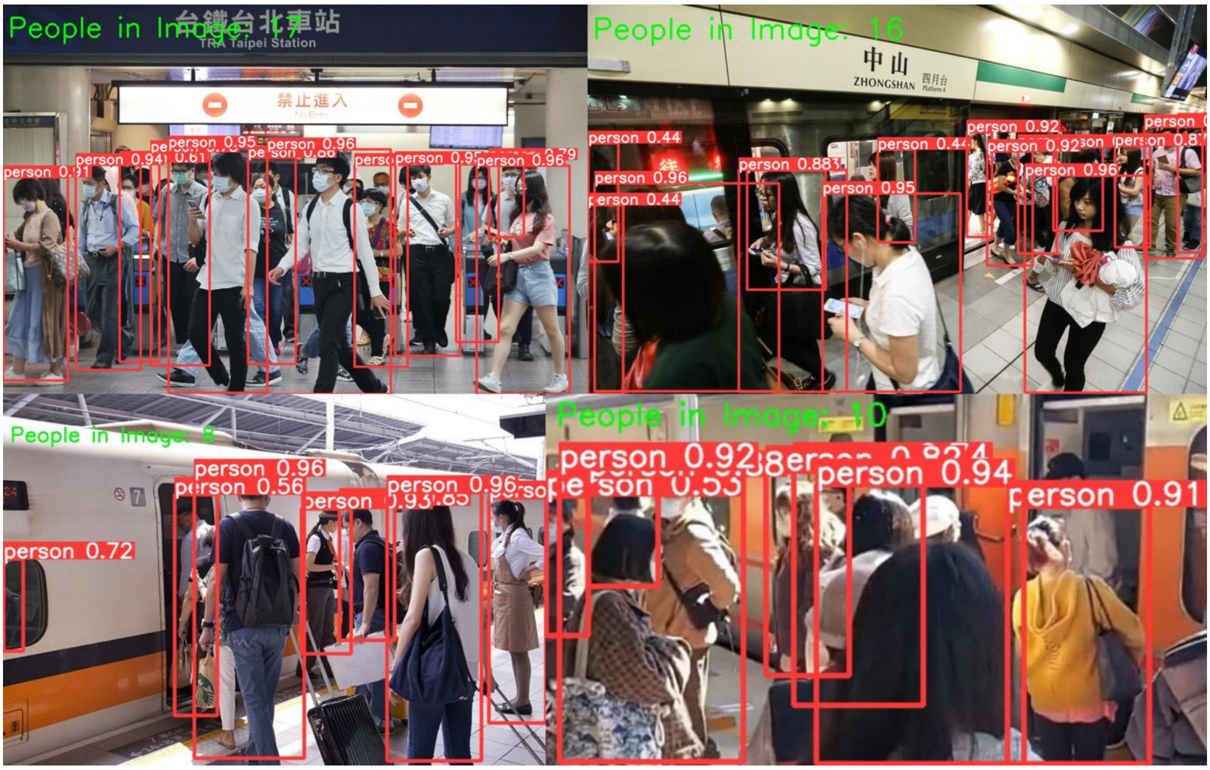

Figure 15 Detection Results 2

Figure 16 Detection Results3

The detection results are accurate. Even in the evening, as long as there is sufficient lighting, the model can still effectively recognize groups of people.

Results Analysis

To evaluate the performance of the model, we employed common performance evaluation methods such as the confusion matrix and ROC curve.

Model Performance Evaluation Methods

Confusion Matrix

The confusion matrix is a tool used to assess the performance of classification models. This matrix displays the correct and incorrect classifications made by the model across various categories. It is primarily divided into four parts:

True Positive (TP): The actual value is positive, and the software prediction is also positive. This is correct.

False Positive (FP): The actual value is negative, but the software predicts a positive result. This is incorrect. This represents a false alarm or Type I error.

True Negative (TN): The actual value is negative, and the software prediction is also negative. This is correct.

False Negative (FN): The actual value is positive, but the software predicts a negative result. This is incorrect. This represents a miss or Type II error.

Using these four values, we can calculate other essential evaluation metrics, such as precision and recall.

Precision: Also known as accuracy, it refers to how many of the data predicted to be Positive are genuinely Positive.

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

Recall: Also known as the hit rate, it means how much of the original Positive data was predicted correctly.

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

Taking the PR_AUC graph as an example, the larger the area under the PR curve, the better the performance of the model.

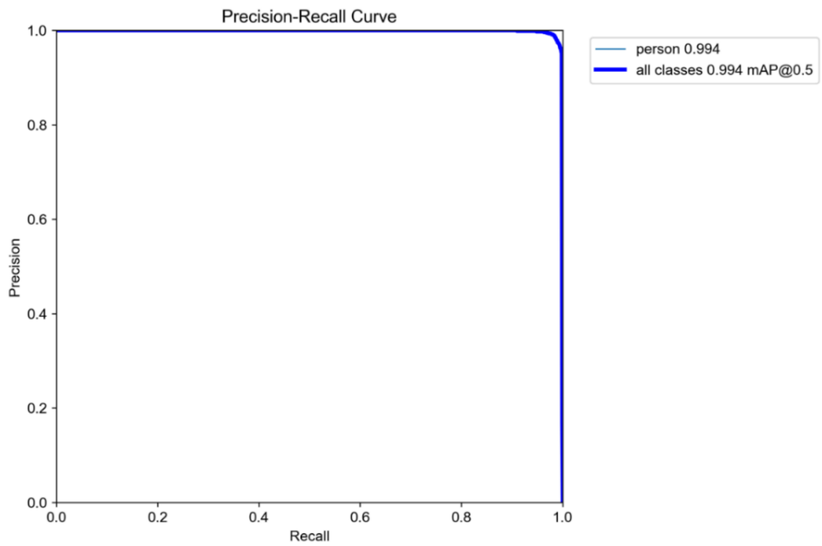

Figure 17 PR_AUC Curve

Conclusion

With advancements in technology, artificial intelligence is playing an increasingly crucial role in optimizing transportation strategies. This study introduces an efficient and accurate method for monitoring and assessing crowd density inside train compartments.

The research results demonstrate the superior performance of YOLOv5s in recognizing crowd density within train compartments. While it exhibits slightly lower precision compared to other versions, such as YOLOv5m, YOLOv5x, and YOLOv5l, its rapid detection speed makes it highly suitable for real-time applications. Future iterations of YOLO are anticipated to address limitations in low-light recognition capabilities, further enhancing its effectiveness.

Future research could focus on developing a congestion alert system that automatically issues warnings when crowd levels exceed specific thresholds. Additionally, exploring multi-angle real-time recognition techniques by deploying cameras strategically across compartments or platforms could capture multi-angular visual data. Advanced image fusion technologies could then generate comprehensive passenger distribution maps, improving recognition accuracy and reducing misjudgments caused by angle constraints.

From a societal and economic perspective, this technology offers substantial benefits. It can help transportation authorities better understand compartment usage, enabling optimized strategies that enhance passenger comfort and safety. This improvement in operational efficiency has the potential to lower costs and increase overall transportation effectiveness.

In the context of supply chain management, integrating deep learning technologies like YOLOv5s provides significant advantages. Real-time and accurate crowd density monitoring can facilitate more informed logistics decisions, optimizing the flow of goods and resources in high-density areas. This reduces shipping costs, minimizes delays, and improves overall supply chain efficiency.

Understanding crowd patterns and peak congestion times enables better asset utilization and resource allocation. For instance, recognizing peak travel times can help schedule more efficient routes for goods transportation, ensuring timely deliveries while reducing vehicle idling.

Moreover, the transparency offered by real-time crowd density monitoring aligns with the objectives of modern supply chain management, which emphasizes visibility and sustainability. Leveraging deep learning and multi-angle recognition technology can enhance demand forecasting accuracy, increase operational flexibility, and support sustainable practices by optimizing resource use and minimizing waste.

This study presents a promising framework for crowd density recognition, laying the groundwork for future research and applications. It underscores the potential of multi-angle recognition and deep learning technologies in improving the accuracy of crowd monitoring in train compartments while highlighting their broader implications for supply chain optimization.

References

Behrooz, H., & Hayeri, Y. M. (2022). Machine learning applications in surface transportation systems: A literature review. *Applied Sciences*, *12*(18), 9156.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527-1554.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*(7), 1097-1105.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.

Lu, K., Han, B., & Zhou, X. (2018). Smart urban transit systems: from integrated framework to interdisciplinary perspective. *Urban Rail Transit*, *4*, 49-67.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016*(6), 779-788.

Russell, S., & Norvig, P. (2022). Artificial Intelligence: A Modern Approach, 4th Global ed.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*(1), 85-117.

Yang, X., Xue, Q., Ding, M., Wu, J., & Gao, Z. (2021). Short-term prediction of passenger volume for urban rail systems: A deep learning approach based on smart-card data. *International Journal of Production Economics*, *231*, 107920.