CONVOLUTIONAL NEURAL NETWORK FOR HUMAN DETECTION

Mr.S.VIGNESH KUMAR M.E.

Assistant Professor
Department of Electronics and Communication Engineering
Chennai Institute of Technology (Autonomous)

SHEIK RASHATH D

Department of Electronics and Communication Engineering
Chennai Institute of Technology (Autonomous)

SAINARESH S

Department of Electronics and Communication Engineering Chennai Institute of Technology (Autonomous)

***

ABSTRACT -*This paper investigates YOLOv8, a deep learning model used for real-time human detection. Using its efficient architecture, YOLOv8 provides quick and precise object identification. Pre-trained models are utilized to perform initial human recognition inside a larger object classification. Frameworks such as PyTorch and Ultralytics make implementation easier. Roboflow is used to train the model on a special human detection dataset with over 1000 annotated photos, potentially outperforming earlier methods in terms of accuracy. Preprocessing processes and model output (bounding boxes with confidence ratings) are examined in the paper and methods for filtering detections and visualizing results are investigated.The paper also presents a method for detecting humans using YOLOv8 which includes data collection (human and human-like objects), annotation (labeling of objects), training, and optional fine-tuning. We describe YOLOv8's real-time applications such as video surveillance, emphasizing its potential for counting pedestriansand activity monitoring. Integration with multi-object tracking methods is being investigated for improved functionality.The study emphasizes the importance of fine-tuning pre-trained models using human-centric datasets in increasing accuracy. It closes by emphasizing YOLOv8's capabilities as a powerful and customizable solution for real-time human identification in a variety of circumstances.*

Keywords—Machine Learning, Deep Learning, Convolutional Neural Network (CNN),You Only Look Once(YOLO)

1. INTRODUCTION

Object identification in computer vision extends beyond merely detecting image content. It aims to replicate human visual perception, allowing machines to not only detect but also precisely locate objects in an image. This goes beyond simply identifying "what" an image contains and tries to answer the queries "what" and "where" by recognizing items and their exact locations with bounding boxes.

Object detection goes beyond basic image content understanding. It addresses two key goals by combining computer vision and machine learning approaches, frequently using deep learning models for improved accuracy. The first challenge is identifying things, similar to how humans see a scene. This requires teaching computers to detect certain objects inside an image, which can range from cars and humans to animals and household goods. However, identification is insufficient. Object detection demands localization, which identifies the actual locations of these objects inside the image. This is commonly accomplished by creating bounding boxes around the observed items. By combining these two tasks, object detection provides a more comprehensive grasp of the visual world within an image.YOLO (You Only Look Once) has been a game changer in object detection since its birth in 2015, thanks to its efficient single-stage architecture. Unlike approaches that require numerous passes of an image, YOLO analyzes the entire image in one step, making it extremely rapid and well-suited for real-time applications. YOLO's progress through its different versions, from YOLOv1 to the most recent YOLOv8, has been characterized by an unwavering commitment to improving accuracy and efficiency.

YOLOv1, the first version, introduced the single-stage detection technique, which achieved real-time rates but had lower accuracy than certain multi-stage detectors. In 2016, YOLOv2 achieved considerable improvements in accuracy while preserving real-time performance. This version included revolutionary approaches such as batch normalization and anchor boxes, which dramatically improved accuracy metrics.

In 2018, YOLOv3 was updated with a more complicated architecture that achieved a commendable balance of speed and accuracy. This version was critical in establishing YOLO's status as a premier object identification technology.

The third edition, YOLOv4 in 2020, represented another watershed moment by introducing new variations adapted to unique needs. YOLOv4-Tiny prioritized speed, catering to applications that required real-time performance, whereas YOLOv4-X valued accuracy, providing a solution for cases where precision was crucial. This adaptability enabled YOLO to meet a wider range of deployment requirements. Ultralytics introduced YOLOv5 in the same year, with the goal of simplifying the model and coding while keeping high accuracy. This iteration focused developer usability, reducing the deployment process and making YOLO more accessible to a larger audience.

In 2022, YOLOv7 concentrated on improving accuracy even more while maintaining the model's portability and suitability for real-time applications. This iteration carried on improving the ratio of speed to precision, guaranteeing YOLO's competitiveness in a market that was changing quickly.
YOLOv8, the most recent version, continues the history of its forebears by providing even quicker processing speeds without sacrificing high detection accuracy. By utilizing deep learning techniques and processing power breakthroughs to enable real-time applications and encourage the creation of increasingly complex computer vision tasks, YOLO has pushed the frontiers of object recognition with each iteration.

YOLO's ongoing evolution highlights the computer vision field's unwavering pursuit of innovation. We can see how computers are becoming more and more capable of "seeing" and interpreting the environment around them, which is opening the door to ground-breaking developments in the field of object detection, by comprehending the fundamentals of object detection and the effectiveness of techniques like YOLO.

2. LITERATURE REVIEW

"[1] Z. Li, Z. Liu and X. Wang, "On-Board Real-Time Pedestrian Detection for Micro Unmanned Aerial Vehicles Based on YOLO-v8," 2023 2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), Jiuzhaigou, China, 2023

"[2] The paper "Overview of YOLO Object Detection Algorithm" authored by Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, Bo Ma published (Science Direct 2022),YOLO's evolution, compares it with CNNs, discusses target recognition methods, feature selection, and their application in diverse fields.

"[3] The paper"CSL-YOLO: A New Lightweight Object Detection System for Edge Computing,"authored byYu-Ming Zhang, Chun-Chieh Lee, Jun-Wei Hsieh, Kuo-Chin Fan featured in (arxivin 2021),introduces the CSL Module, employing Depthwise Convolution to generate redundant features efficiently, reducing computation costs. Experiments on MS-COCO demonstrate CSL-Module's capability to approximate Convolution-3x3. Utilizing CSL-Module, CSL-YOLO achieves improved detection performance with reduced FLOPs and parameters compared to Tiny-YOLOv4.

[4]Dr. Muhammad Hussain'swork titled "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," featured in (MDPI 2023),examines the evolution of YOLO variants, emphasizing their alignment with industrial surface defect detection needs for speed, accuracy, and edge device deployment. It delves into architectural advancements from original YOLO to YOLO-v8, showcasing their evolution in meeting industrial demands.

[5] "A Practise For Object Detection Using Yolo Algorithm"applied on IJSRCSEIT in 2021, Object detection, vital for various applications like surveillance and autonomous driving, utilizes algorithms like YOLOv3. YOLO processes entire images with a single neural network, predicting bounding boxes and probabilities for objects. Its progressive learning enhances accuracy over time, making it ideal for real-time scenarios.

[6] The paper "Performance Validation of Yolo Variants for Object Detection," authored by Kaiyue Liu1 , Haitong Tang2 , Shuang He3 , Qin Yu4 , Yulong Xiong5 , Nizhuan Wang presented on (ACM Digital Library 2021), outlines deep learning object detection, particularly YOLO's one-stage, speedy approach, its strong generalization, and recent breakthroughs, offering valuable insights for related research.

[7] "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments" authored by wei fang 1,2, (member, ieee), lin wang 1, and peiming ren 1 Liu,published in the IEEE journal of selected topics in applied earth observations and remote sensing, vol. 17, 2017 introduces Tinier-YOLO, derived from Tiny-YOLO-V3, aiming for improved object detection on embedded devices. It employs SqueezeNet's fire module and introduces dense connections for enhanced feature propagation, achieving higher accuracy and real-time performance..

[8] "You Only Look Once: Unified, Real-Time Object Detection" authored by Joseph Redmon∗ , Santosh Divvala, Ross Girshick , Ali Farhadi presented on CVF 2015 - YOLO revolutionizes object detection with a unified regression approach, achieving remarkable speed and strong generalization across domains.

[9] "YOLO9000: Better, Faster, Stronger" by Joseph Redmon, Ali Farhadi published in the arXiv: Computer Vision and Pattern Recognition in 2016, introduces advancements in real-time object detection, surpassing state-of-the-art methods like Faster RCNN and SSD. Its joint training approach facilitates predictions for unlabeled classes, validated on ImageNet. With over 9000 object categories, YOLO9000 excels in versatility and efficiency.

[10] The paper "Overview of YOLO Object Detection Al," authored by Chengjuan Wan, Yuxuan Pang, Shanzhen Lan published in International Journal of Computing and Information in Vol 1, No. 2, 20, surveys various methods of deepfake detection based on machine learning. The authors review existing techniques and highlight the advancements in leveraging machine learning for identifying manipulated content.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

3. PROPOSED SYSTEM

3.1 YOLO_v8:

Description:

YOLO v8, a cutting-edge convolutional neural network, is used in this research to precisely detect humans and items that resemble people.The model will be trained to precisely pinpoint items through the collection of a wide range of datasets and careful annotation.Its versatility and real-time processing capabilities make it more applicable to autonomous systems, crowd analysis, and security systems.

3.2 Training Details:

• Yolov8 uses a single-stage detection method, which means that for object detection and localization, it evaluates the entire image at once.

• By examining the labeled bounding boxes in your own dataset, the model gains the ability to recognize individuals and items that resemble persons during training.

• In order to improve the training process for your particular dataset, hyperparameter tuning—which involves altering variables such as learning rate, batch size, and optimizer—will probably be necessary.

Benefits:

- **Real-time Processing:** YOLOv8 can detect and locate objects in an image in a single pass since it uses the same single-stage detection method as earlier YOLO versions. When compared to multi-stage detectors, this results in faster processing times, which makes it appropriate for real-time applications.
- **Accuracy:** YOLOv8 retains a high degree of detection accuracy in spite of its emphasis on speed. This is probably because of featuressuch as the model architecture's use of cutting-edge methodologies and its capability for pre-training on big datasets.
- **Increased Efficiency:** YOLOv8 improves upon the innovations of preceding YOLO versions, possibly providing increased efficiency over earlier models in terms of computing power and memory utilization. When deploying on devices with limited resources, this can be quite important.
- **Adaptability:** Through user-defined configurations, the YOLOv8 framework provides training features. This enables you to customize the training procedure to your unique dataset and intended results.
- **Potential for Fine-tuning:** By utilizing pre-trained weights, YOLOv8 enables you to fine-tune the model for human and human-like object detection on your own dataset. The accuracy of the model can be further improved for your particular case with this fine-tuning.
- **Open-source Availability:** YOLOv8 is available for development and investigation by a larger community due to its open-source nature. This encourages cooperation and creativity in the object detecting field.

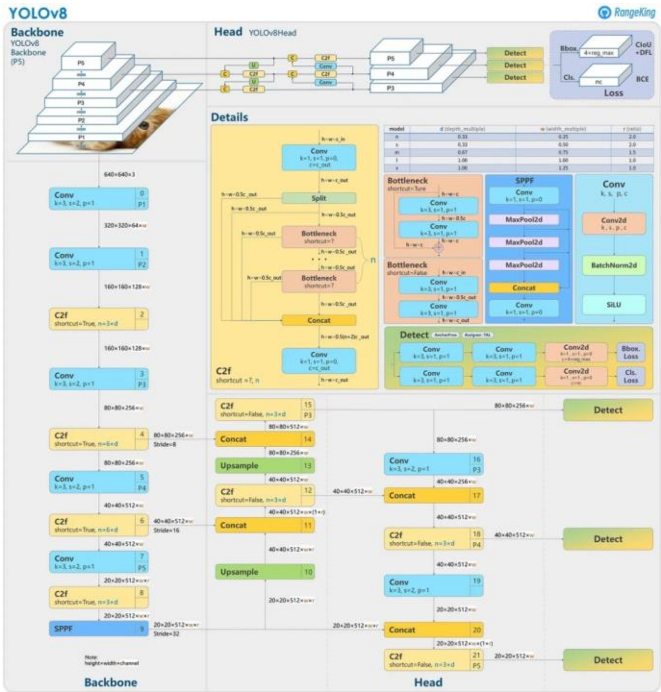

Figure 1. The architecture of YOLOv8 released by Ultralytics, the company behind YOLOv5

3.3 Flow Of Execution:

1. Data Acquisition:

- We Compiled a varied collection of pictures with people and objects that resemble people. This diversity ought to include a range of elements, such as:
oPositions (profile, frontal, etc.)
o Lighting circumstances (bright, dim, etc.)
o Backgrounds (cluttered, outside, indoors, etc.)
o The sizes of objects (whole body, partial view, etc.)
- The performance of the model will be greatly impacted by the quantity and variety of your dataset.

2. Data Annotation:

- Draw bounding boxes around individuals and items that resemble persons in each image in your dataset.
- Annotation tools can assist speed up this process
- Bounding boxes are essentially rectangles drawn around the things you want the model to detect.

3.4 YOLOv8 Training:

- Train the model on your annotated dataset using the YOLOv8 framework.For effective training, the framework probably makes use of a deep learning library like PyTorch or TensorFlow.
- The model is trained by feeding it with the bounding boxes and annotated photos. Through learning to recognize relationships and patterns in the data, the model will be able to recognize items that are similar in new photographs.

3.5 Algorithm in YOLOv8 (Single-Stage Detection):

YOLOv8's one-stage method simplifies object detection into a seamless pipeline, revolutionizing the field. Instead of processing a picture in numerous passes, YOLOv8 processes the full image in one pass. Its convolutional layers precisely extract relevant characteristics, identifying colors, shapes, and edges. By utilizing these characteristics, the model forecasts bounding boxes that contain possible objects together with matching class probabilities that indicate the degree of confidence in the prediction. Non-maxima suppression (NMS) resolves overlapping bounding boxes and combines them into solitary, refined detections to guarantee correctness. This simplified approach perfectly captures the efficacy and efficiency of YOLOv8 in object detection tasks.

3.5.1 Fine-tuning (Optional):

• You can fine-tune the pre-trained YOLOv8 model on your particular dataset based on your needs. The model's capacity to identify individuals and objects that resemble humans in your specific surroundings is significantly improved by this fine-tuning.

3.5.2 Evaluation and Deployment:

• To determine the accuracy and generalizability of the trained model, its performance on an independent test dataset should be analyzed.
•If the model's performance meets the needs, it is used for real-time object recognition applications such as autonomous system navigation, crowd analysis, and security monitoring.

4. IMPLEMENTATION DETAILS :

4.1 Setting up the collab environment:

"!Pip install yolov8 torch torchvision" should be used to install necessary libraries, such as yolov8, torch, and torchvision"

4.2 Downloading YOLOv8 model weights:

Users can access the official repository for YOLOv8 variants by using the weights of the YOLOv8 model. They have a variety of options to select from, each catering to particular needs. For example, "yolov8n.pt" is suggested for devices with limited resources, "yolov8s.pt" strikes a balance between speed and precision, and "yolov8m.pt" provides greater accuracy but necessitates more computational power. Once the preferred version has been chosen, users can download the associated model weights file and add it to their Colab workspace for additional use.

4.3 Data preparation:

We implemented a set of steps for the dataset of photos that included humans and items that resembled people in the context of data preparation. The photos must first undergo pre-processing, such as resizing and RGB format conversion, in order to comply with YOLOv8 specifications. Consequently, annotation of the photographs is important, which involves drawing bounding boxes around items that resemble individuals. For this, annotation programs like LabelImg and VGG Image Annotator are commonly used. This step can be omitted, though, if a pre-defined dataset is being used for presentation.

4.4 Person Detection script

Initially, the notebook is filled with the YOLOv8 model and weights. Then, if a custom dataset is being used, a method for pre-processing and image loading is defined. An additional function for human detection is implemented, in which a test image path or the previously stated function is used to load the image. The YOLOv8 model is used for inference to provide detections, which are then processed using bounding boxes and confidence ratings to identify human and human-like objects. Alternatively, detections can be shown visually by utilizing libraries such as OpenCV to put bounding boxes and labels ("person") on the image.

5. RESULTS AND DISCUSSION

5.1 Training the system:

"Run the dataset and train the dataset repeatedly to improve "the focal loss graph by pulling the curve of gamma value first setting it to 1 and increasing it step by step until the confidence curve of the predicted result improves.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

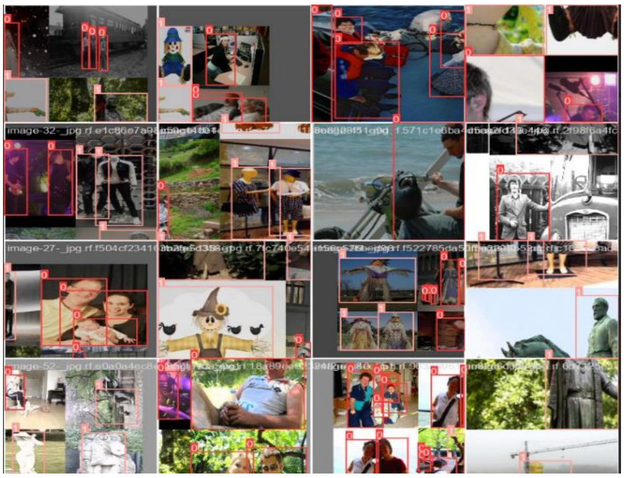

Fig 5.1: Train batch_0

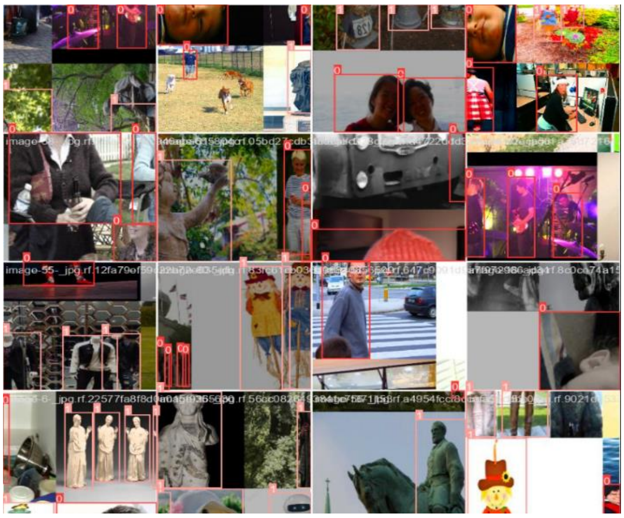

Fig 5.1.1 : Train batch_2

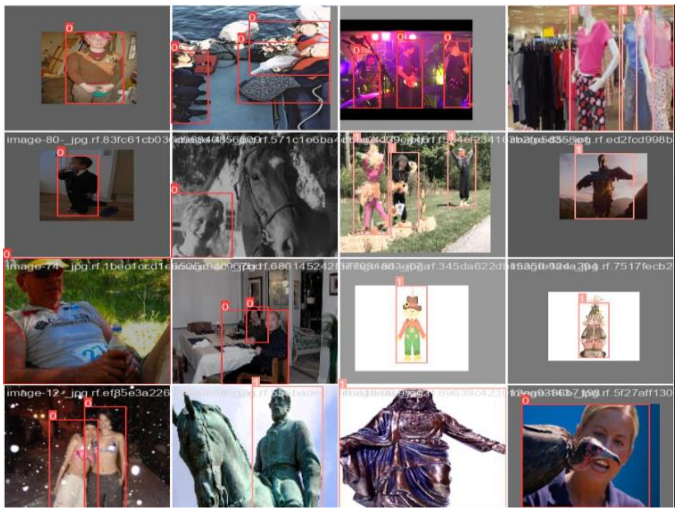

Fig 5.1.2: Train batch-150

5.2 Predicted results:

Fig 5.2 :Value batch_0

Fig5.2.1:Val bactch_2,predicted output by training the images and valuing it, which gives the output of detecting the human and human like objects from the given dataset.

5.3 Graph curve:

5.3.1 Confidence Curve:

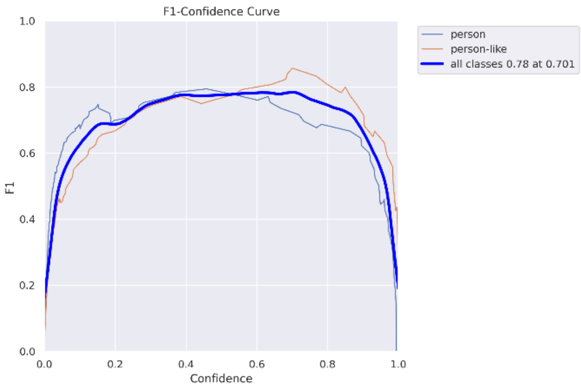

Fig 5.3.1:Confidence Curve:

5.3.2 Precision confidence curve:

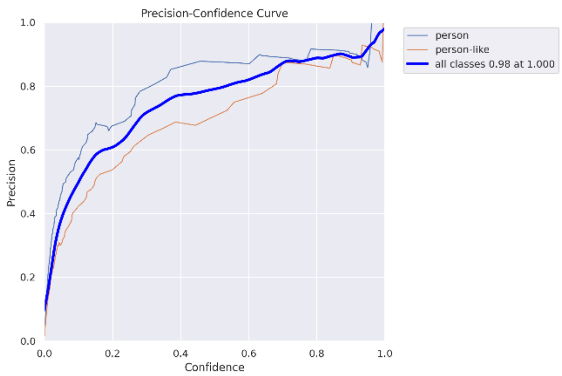

Fig 5.3.2:Precision confidence curve:

5.3.3 Precision recall curve

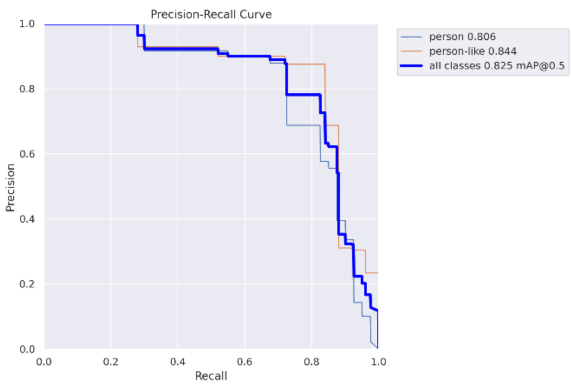

Fig 5.3.3Precision recall curve

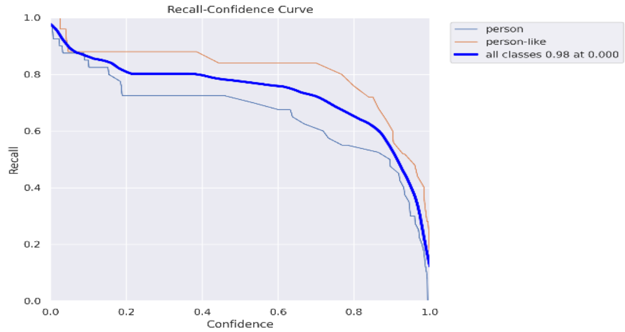

Fig 5.3.4:Recall confidence curve

5.4 Confusion Matrix Analysis:

Understanding Predictions:

Four categories are created by the confusion matrix, a useful tool for analyzing model predictions: true positives, true negatives, false positives, and false negatives. This graphic depiction highlights the model's strong points and potential areas for improvement. Analyzing the confusion matrix makes it easier to pinpoint the ensemble model's advantages and shortcomings.

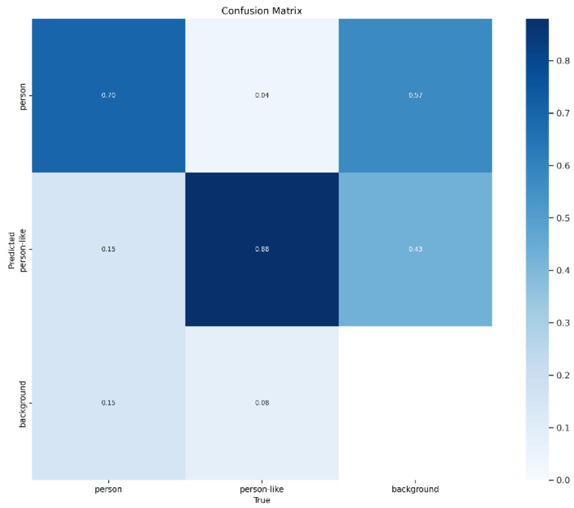

Fig 5.4:Confusion Matrix Analysis

6. CONCLUSION

In this research, we investigate real-time detection of human and human-like objects in photos using YOLOv8, a state-of-the-art deep learning model. Data collection, annotation, YOLOv8 training, and possible fine-tuning are important modules. For the purpose of training the model and influencing its performance and generalization, a varied dataset including a range of characteristics including stances, lighting situations, and backgrounds was essential. An important but time-consuming stage, bounding boxes around persons and person-like items were carefully added to every image in the collection. Model training was made easier by the YOLOv8 framework, which used a single-stage detection strategy for quicker processing times than multi-stage detectors. The model was able to recognize similar things in new photographs because it had learned to recognize patterns and correlations within the data during training.

Accurate detection, adaptability, and real-time applicability are among the anticipated outcomes. The system's goal is to minimize false positives and negatives by precisely identifying and localizing human and human-like objects within photos. Because of its speed, YOLOv8 can process video streams or image sequences in real time, making it appropriate for applications that need to respond quickly. Furthermore, the framework provides user-defined configuration functions that enable the training process to be customized, enabling the model to be adjusted to different contexts outside of the original dataset. This versatility creates opportunities for a wide range of applications in many fields.

This project's successful completion offers potential uses in domains where real-time human detection is essential. Real-time warnings for unwanted breaches can enhance security system monitoring. Accurate person counting and tracking can be used with crowd analysis to improve resource allocation and public safety measures during events or in public areas. Through precise detection of humans and items that resemble human, autonomous systems can improve navigation and interaction with the environment while guaranteeing safe operation by preventing collisions. Overall, this work contributes to the improvement of computer vision jobs demanding real-time processing and flexibility to a variety of circumstances by investigating the potential of YOLOv8 for human and human-like object detection.

REFERENCES

1. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

2. A. Amerini et al., "Exposing Deepfakes Using Inconsistent Head Poses," IEEE Transactions on Information Forensics and Security, 2019.

3. Dewei Zhao, Faming Shao, Qiang Liu, Li Yang, Heng Zhang, Zihan Zhang, "A Small Object Detection Method for Drone-Captured Images Based on Improved YOLOv7", *Remote Sensing*, vol.16, no.6, pp.1002, 2024.

4. X. Yu, T. W. Kuan, Y. Zhang and T. Yan, "YOLO v5 for SDSB Distant Tiny Object Detection," *2022 10th International Conference on Orange Technology (ICOT)*, Shanghai, China, 2022,

5. T. W. Kuan, S. Chen, S. N. Luo, Y. Chen, J. F. Wang and C. Wang, "Perspective on SDSB Human Visual Knowledge and Intelligence for Happiness Campus", *2021 9th International Conference on Orange Technology (ICOT)*, pp. 1-4, 2021, December.

6. S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks", *arXiv 2015*.

7. T.Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal loss for dense object detection", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988, 22–29 October 2017.

8. A. Bochkovskiy, C.Y. Wang and H.Y.M. Liao, "Yolov4: Optimal speed and accuracy of object detection", *arXiv*, 2020.

9. Y. Wang, J. Liu, S. Yu, K. Wang, Z. Han and Y. Tang, "Underwater Object Detection based on YOLO-v3 network," *2021 IEEE International Conference on Unmanned Systems (ICUS)*, Beijing, China, 2021.

10. M. Sun et al., "Facial Action Transfer with Personalized Bilinear Regression," IEEE Transactions on Multimedia, 2020.

11. X Sun, P Wu and S C H Hoi, "Face detection using deep learning: An improved faster RCNN approach", *Neurocomputing*, vol. 299, pp. 42-50, 2018.

12. G Huo, Z Wu and J. Li, "Underwater Object Classification in Side-scan Sonar Images Using Deep Transfer Learning and Semi-synthetic Training Data", *IEEE Access*, pp. 1-1, 2020

13. Shuai Shao, Chaolong Liu, Juan Cheng, Jingjing Liu, "A Method Based on YOLOv8 for Sonar Image Object Detection", *2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp.1-5, 2023.

14. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," arXiv, 1703.10593.

15. J. Fan, J. Lee, I. Jung and Y. Lee, "Improvement of Object Detection Based on Faster R-CNN and YOLO," *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jeju, Korea (South), 2021.

16. Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel and F. Porikli, "Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework", *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1002-1014, April 2016.

17. M Prakash, M Janarthanan, D Devi, "Multiple Objects Identification for Autonomous Car using YOLO and CNN", *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp.597-601, 2023.

18. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *ArXiv*.

19. B. Ye, Y. Shi, H. Li, L. Li and S. Tong, "Face SSD: A Real-time Face Detector based on SSD," *2021 40th Chinese Control Conference (CCC)*, Shanghai, China, 2021.

20. Pei Wen, Cheng Sun, Shiwen Zhang, Yuansheng Luo, Haowei Huang, Jin Zhang, "EOPSA-FACE: An Encoder-Only DETR for Classroom Face Detection", *2022 10th International Conference on Information Systems and Computing Technology (ISCTech)*, pp.341-348, 2022.