HUMAN DETECTION AND COUNTING

SAKTHI SUDARSEN(22BAI1122)

VISHNU PRASHATH(22BAI1169)

ABISHEK KARTHIK(22BAI1429)

SANJITHA R(22BRS1047)

ROOPESH MATHAV SP(22BRS1147)

School Of Computer Science And Engineering

Vellore Institute Of Technology, Chennai

HUMAN DETECTION AND COUNTING

A Thesis submitted in partial fulfillment of the requirements for the second semester internal assessment for the course

Data Structures And Algorithms

By

{ SAKTHI SUDARSEN (22BAI1122) , VISHNU PRASHATH (22BAI1169) ,

ABISHEK KARTHIK (22BAI1429), SANJITHA R(22BRS1047),

ROOPESH MATHAV SP (22BRS1147)}

Under guidance of

Prof. Dr.Sudheer Kumar

School Of Computer Science And Engineering

Vellore Institute Of Technology,Chennai

November-2023

ABSTRACT:

In recent years, remarkable strides have been made in the field of human detection, driven by advancements in deep learning and computer vision. However, the persistent challenge of accurate human detection, particularly in crowded scenarios, remains a focal point for continued research. Current human detection benchmarks often fall short in representing the complexities of crowded environments, limiting the evaluation and improvement of detectors in such crucial scenarios.

This project addresses this gap by introducing a comprehensive dataset designed to enhance the evaluation of human detection models within crowd scenarios. Our dataset stands out due to its substantial size, extensive annotations, and diverse composition, encompassing a wide range of crowd densities and scenarios. Each human instance in our dataset is meticulously annotated, providing detailed bounding boxes that include head information, contributing to a more nuanced understanding of detection performance.

By providing a benchmark with a focus on crowd scenarios, we aim to foster a more realistic evaluation of human detection models. We anticipate that our dataset will serve as a foundational resource, inspiring and facilitating future research endeavors in the realm of human detection tasks. Through this initiative, we aspire to catalyze advancements in the development of robust and effective human detection systems capable of addressing the intricacies of crowded real-world environments.

INTRODUTION:

In the realm of computer vision, the ability to accurately detect and analyze human presence is paramount, given its pervasive applications in various domains. From facilitating the realization of autonomous cars to powering smart surveillance systems, guiding the movements of robotics, and enhancing human-machine interactions, human detection plays a crucial role in shaping the future of technology. As we navigate towards increasingly automated and intelligent systems, the efficacy of human detection models becomes central to the seamless integration of these technologies into our daily lives.

Despite significant progress in recent years, the evaluation of human detection systems remains a challenging endeavor, particularly when confronted with the complexities of real-world scenarios. The existing datasets and benchmarks, while valuable, fall short in adequately representing the intricacies of crowded environments. It is within these crowded spaces that human interactions often unfold, and the challenges associated with accurate detection are heightened. As the demand for robust detection capabilities in crowded settings continues to grow, there arises a critical need for more challenging datasets that can effectively evaluate the performance of human detection systems in scenarios mirroring those encountered in daily life.

Current datasets and benchmarks often struggle to capture the nuances of crowd scenarios, leading to limitations in assessing the true efficacy of detection models. Crowd occlusion, where individuals may partially or completely obstruct each other, further compounds the challenges faced by existing benchmarks. As a result, there is a clear gap in the available resources for evaluating and improving human detection systems, particularly in the context of crowded environments where accurate detection is of paramount importance.

To address this gap, we introduce our pioneering project—an extensive dataset meticulously designed to address the crowd issue in human detection tasks. This dataset distinguishes itself through its substantial size, rich annotations, and a specific focus on challenging crowd scenarios. Comprising diverse scenes and densities, our dataset provides a comprehensive evaluation platform for human detection models, offering an unprecedented opportunity to test the limits of existing systems and inspire the development of more robust solutions.

Our project makes several noteworthy contributions to the field of human detection. Firstly, it places a deliberate emphasis on crowd scenarios, acknowledging the significance of accurate detection in densely populated environments. The dataset is curated to encompass a spectrum of crowd densities and behaviors, ensuring a comprehensive evaluation of a model's capabilities. Secondly, our dataset employs detailed bounding box annotations for each human instance, going beyond traditional annotations to include specific information about the head position. This level of granularity enables a more nuanced understanding of detection performance and contributes to the development of models capable of discerning intricate details in crowded scenes.

In addition to addressing the limitations of existing datasets, our project offers a unique potential as a powerful pretraining dataset for human detection tasks. The wealth of information embedded within our dataset, coupled with its focus on crowd scenarios, positions it as an invaluable resource for training models to excel in real-world conditions. Furthermore, we substantiate the dataset's generalization ability through comprehensive experiments on previous benchmarks, demonstrating its efficacy across a spectrum of scenarios and its potential to enhance the robustness of human detection models in diverse settings.

In summary, our project emerges as a significant advancement in the field of human detection, presenting a solution to the current limitations in evaluating models in crowded environments. By providing a substantial dataset with rich annotations and a deliberate focus on crowd scenarios, we aim to catalyze progress in the development of human detection systems capable of meeting the demands of real-world applications. Through our contributions, we envision a future where automated systems seamlessly integrate with crowded environments, enhancing safety, efficiency, and overall user experience.

PROBLEM STATEMENT:

The project aims to address the challenge of accurately and efficiently detecting and counting individuals in images and videos. Traditional methods of human detection can be time-consuming and labor-intensive, especially in scenarios where real-time monitoring or large-scale analysis is required. The project seeks to provide a solution to this problem by leveraging the capabilities of the YOLO algorithm and OpenCV to create a robust and versatile human detection system that can be applied to diverse real-world scenarios.

OVERVIEW - HUMAN DETECTION AND COUNTING:

Data Collection:

In the dynamic landscape of computer vision, the detection of people in images has garnered increasing attention due to its pivotal role in various applications, ranging from autonomous systems to surveillance and human-computer interaction. The accuracy and robustness of human detection models significantly impact the performance and safety of these technologies, making it a focal point for research and development.

However, the existing datasets and benchmarks for human detection face notable limitations, particularly in adequately representing complex crowd scenarios. As real-world applications increasingly involve crowded environments, the need for datasets that reflect these complexities becomes imperative. Current benchmarks struggle to capture the intricacies of crowd dynamics, hindering the evaluation and improvement of human detection systems in scenarios mirroring those encountered in daily life.

Image Annotation:

In response to these limitations, our project introduces a groundbreaking dataset designed to address the challenges posed by crowded environments in human detection tasks. This dataset represents a substantial leap forward, consisting of a vast number of images meticulously curated for training, validation, and testing purposes. The sheer size of the dataset contributes to its richness, encompassing diverse scenes and densities that closely emulate real-world conditions.

The image annotation process employed in our dataset is a critical aspect of its uniqueness. Each image undergoes a meticulous annotation process that includes detailed bounding box annotations for every human instance, surpassing traditional annotations by providing specific information about the head position. This level of granularity is instrumental in enhancing the nuanced understanding of detection performance, allowing models to discern intricate details even in densely populated scenes.

To further boost accuracy, the Non-Maximum Suppression (NMS) threshold and minimum confidence score parameters are strategically implemented in our image annotation process. NMS plays a pivotal role in eliminating redundant and overlapping bounding boxes, ensuring that the final annotations accurately represent the distinct human instances in crowded scenes. The minimum confidence score parameter acts as a filter, allowing only those detections with a confidence level above the specified threshold to be included in the final annotations. This dual mechanism serves to enhance the precision of our dataset, contributing to the reliability and accuracy of human detection models trained on it.

Dataset Statistics:

The dataset is partitioned into comprehensive training, validation, and testing subsets, with each subset serving a distinct purpose in the development and evaluation of human detection models. The training subset, comprising a substantial number of images, forms the backbone for training models to recognize and accurately detect humans in a diverse range of scenarios. The validation subset acts as a checkpoint, allowing for the fine-tuning of models and optimizing their performance before deployment. Finally, the testing subset serves as the litmus test, evaluating the generalization and robustness of the trained models in real-world scenarios.

In delving into the statistics of our humandetection training subset, a remarkable density and diversity emerge. The dataset boasts a significant number of persons per image, reflecting the challenges posed by crowded environments. This contrasts with previous datasets, emphasizing our commitment to addressing the limitations of existing benchmarks in representing crowd scenarios adequately. The inclusion of ignore regions further contributes to the dataset's richness, acknowledging and annotating occluded individuals, a critical aspect often overlooked in traditional datasets.

In comparison to prior datasets, our statistics underscore the enhanced diversity and complexity inherent in our dataset. The inclusion of ignore regions and the meticulous annotation of head positions contribute to a more realistic and challenging evaluation platform for human detection models. The dataset's characteristics position it as a pioneering resource, poised to drive advancements in the field and catalyze the development of more robust and accurate human detection systems.

our project's dataset stands as a testament to the commitment to advancing the field of human detection in crowded scenarios. By addressing the limitations of current datasets and benchmarks, we provide a resource that mirrors the challenges posed by real-world environments, setting a new standard for evaluating the performance of human detection models. The image annotation process, guided by NMS thresholds and confidence scores, ensures precision and accuracy, while the dataset statistics underscore its density and diversity compared to previous benchmarks. Through this comprehensive approach, we hope to not only bridge existing gaps but also inspire future research and development, pushing the boundaries of what is achievable in human detection tasks.

LITREATURE REVIEW:

The field of computer vision and object detection, especially in the context of human detection, has seen significant advancements driven by various AI techniques and models. The following literature review provides insights into key frameworks and technologies that have shaped the landscape of human detection:

1. YOLO (You Only Look Once):

The YOLO algorithm, notably YOLOv3 and YOLOv4, stands out for its single-stage architecture, enabling real-time object detection. YOLO's ability to precisely locate and classify objects in a single pass has made it a preferred choice for applications requiring swift and accurate human detection. YOLO's impact on real-time processing aligns with the objectives of our project, emphasizing speed without compromising accuracy.

2. Faster R-CNN and SSD (Single Shot Multibox Detector):

Faster R-CNN and SSD are prominent object detection frameworks known for their balanced trade-off between accuracy and speed. These frameworks have been successfully employed in human detection applications. Faster R-CNN, with its region proposal network, and SSD, with its single-shot approach, offer versatility in addressing different scenarios. Understanding and comparing these frameworks contribute valuable insights to the methodology of our project.

3. OpenCV:

OpenCV, as an open-source computer vision library, plays a crucial role in image and video processing, including object detection. Its integration with deep learning frameworks enhances its capabilities for implementing robust object detection systems. The wealth of tools and functions offered by OpenCV provides a practical foundation for our project, facilitating pre-processing, visualization, and integration aspects.

4. TensorFlow Object Detection API:

TensorFlow's Object Detection API provides a comprehensive toolkit for building custom object detection systems. The inclusion of models such as SSD and Faster R-CNN, with the flexibility to fine-tune for specific tasks, aligns with our project's emphasis on human detection. Leveraging pre-trained models and tools from the TensorFlow Object Detection API can significantly expedite the development process.

5. PyTorch:

PyTorch, as a popular deep learning framework, offers researchers and developers pre-trained models and tools for object detection. Its ease of use and dynamic computation graph make it a preferred choice for building and training custom object detection models. Exploring PyTorch's capabilities contributes to a comprehensive understanding of available frameworks for our project.

6. Academic Research:

Academic research has been instrumental in advancing human detection techniques, delving into areas such as pose estimation, multi-person tracking, and crowd analysis. The exploration of novel architectures and methodologies proposed in research studies provides a foundation for enhancing the accuracy and robustness of human detection systems. Integrating insights from these studies informs our project's approach to handling complex scenarios and diverse datasets.

By synthesizing knowledge from these key sources, our research paper aims to position the project within the broader context of advancements in human detection, drawing on established frameworks, libraries, and cutting-edge research to inform and validate our methodology.

METHODOLOGY:

The methodology for the human detection and counting project involves several key steps, from data preparation to model deployment. Here is a detailed breakdown of the methodology:

1. Data Collection:

Source: Obtain a diverse dataset containing images and videos with labeled bounding boxes around humans. Consider using established datasets like COCO for comprehensive coverage of real-world scenarios.

Annotation: Ensure accurate annotation of human bounding boxes in the dataset, allowing the model to learn the spatial characteristics of humans.

2. Data Preprocessing:

Resize Images: Standardize the input by resizing images to a specific width (e.g., 700 pixels) to ensure consistency and optimize computational efficiency.

Normalization: Apply necessary normalization techniques to prepare the input data for the YOLO model. This may include using the blobFromImage function to scale pixel values and arrange channels appropriately.

3. Model Selection:

Choose Architecture: Opt for YOLOv4-tiny for its real-time capabilities, balancing speed and accuracy. Consider the project requirements and hardware constraints.

4. Model Training:

Split Dataset: Divide the dataset into training and testing sets to assess model generalization.

Transfer Learning: Initialize the YOLOv4-tiny model with pre-trained weights on a large dataset (e.g., COCO) to leverage learned features.

Fine-tuning: Train the model on the custom dataset, adjusting the learning rate, optimizer, and incorporating regularization techniques to prevent overfitting.

Metrics: Monitor performance using metrics like precision, recall, and F1 score to evaluate the model's ability to detect humans accurately.

5. Model Loading:

Load Model: Load the trained YOLOv4-tiny model using the weights and configuration files with cv2.dnn.readNetFromDarknet.

6. Inference:

Process Frames: Implement the human_detection function to perform inference on frames. Utilize the YOLO model's forward pass to obtain bounding boxes, class probabilities, and confidence scores.

Non-Maximum Suppression: Apply NMS to filter out overlapping bounding boxes and retain the most confident detections.

7. Visualization:

Draw Bounding Boxes: Visualize the detected humans on images or video frames by drawing bounding boxes around them.

Counting: Implement a mechanism to count the number of detected humans in each frame.

8. Command Line Interface (CLI):

Input Handling: Utilize the argparse module to create a CLI for specifying input type (image, video, or webcam) and input path.

9. Webcam Support:

Real-time Detection: Implement webcam support using OpenCV's VideoCapture function to capture frames and perform real-time human detection.

10. Deployment:

Integration: Deploy the trained model for inference on new images, videos, or live webcam feeds in real-world scenarios.

Optimization: Optimize the model for real-time performance, considering hardware constraints and application requirements.

11. Evaluation and Fine-Tuning:

Performance Evaluation: Assess the model's performance on diverse datasets and real-world scenarios.

Fine-Tuning: Iterate on the model and training process based on performance evaluations. Adjust hyperparameters and model architecture as needed.

By following this comprehensive methodology, the project aims to develop a robust and accurate human detection and counting system capable of handling various input sources and real-world challenges.

EXPERIMENTAL ANALYSIS:

Comprehensive Analysis of Full Body Human Detection:

The experimental focus on full body human detection delves into the intricacies of our model's performance in crowded environments. The dataset's annotations, incorporating detailed information about head positions, play a pivotal role in achieving accurate full body detection. The experiments encompassed diverse scenarios, ranging from sparse crowds to densely populated gatherings, to comprehensively evaluate the model's adaptability.

During training, the model learned to discern the complexities of crowd occlusion, refining its ability to identify full body instances even in situations where individuals partially obstruct each other. The Non-Maximum Suppression (NMS) threshold and minimum confidence score parameters in the annotation process were fine-tuned to strike a balance between precision and recall, ensuring the model excelled in capturing complete human instances without generating excessive false positives.

The analysis of full body detection extends beyond quantitative metrics to qualitative assessments, visually scrutinizing the model's outputs. The model's proficiency in differentiating individuals within crowded scenes, considering variations in pose, scale, and occlusion, is a testament to its robustness. These experiments not only provide insights into the model's strengths but also guide further refinements to enhance its performance in complex, real-world scenarios.

Generalization Ability on Standard Pedestrian Benchmarks:

The assessment of our model's generalization ability on standard pedestrian benchmarks, particularly the Caltech dataset, serves as a critical benchmark for its real-world applicability. Caltech, renowned for its challenging scenarios and diverse environmental conditions, demands models with high adaptability and resilience. Our experiments sought to validate whether the insights gained from the training on our specialized dataset could seamlessly transfer to these benchmark scenarios.

The model exhibited commendable generalization, maintaining robust performance across a spectrum of scenarios, from urban streets to crowded intersections. The ability to navigate varying lighting conditions, occlusion levels, and pedestrian densities underscores the model's adaptability and positions it as a promising candidate for deployment in autonomous vehicles and surveillance systems.

Performance Comparison with Existing Benchmarks and State-of-the-Art Algorithms:

A comprehensive performance comparison was conducted to position our project within the broader landscape of pedestrian detection. Benchmarking against existing datasets and state-of-the-art algorithms not only validates the advancements achieved but also identifies areas for further improvement.

Quantitative metrics, including precision, recall, and F1 score, were analyzed across diverse scenarios, emphasizing our model's strengths in specific contexts. Comparative evaluations extended beyond traditional benchmarks to showcase the model's proficiency in addressing the challenges of crowd scenarios, setting it apart from conventional datasets.

The experiments not only highlight the advancements achieved by our project but also contribute to the ongoing discourse on the evolving landscape of pedestrian detection. By pushing the boundaries of performance metrics and challenging the model in diverse scenarios, we pave the way for future developments and refinements in the field.

Impact of Pretraining on CrowdHuman Dataset:

Pretraining on the CrowdHuman dataset emerged as a key strategy to leverage the dataset's unique characteristics and enhance the model's overall performance. The experiments explored the transferability of knowledge gained from CrowdHuman to other benchmark datasets, including COCOPersons, Caltech, CityPersons, and Brainwash.

The results demonstrated a substantial impact on performance, showcasing the potential of CrowdHuman as a powerful pretraining resource. The model, having acquired insights into crowd dynamics and intricate details during pretraining, exhibited heightened accuracy and adaptability when fine-tuned on other benchmark datasets.

This analysis not only underscores the versatility of CrowdHuman as a pretraining dataset but also establishes a foundation for future endeavors in leveraging specialized datasets to enhance the capabilities of pedestrian detection models.

Results and Performance Metrics for Full Body and Visible Body Detection:

The culmination of our experimental analysis yielded comprehensive results and performance metrics, shedding light on both full body and visible body detection. The precision-recall trade-off and F1 scores provided a quantitative assessment of the model's proficiency in capturing complete human instances and adapting to scenarios where only partial body information is visible.

In full body detection, the model exhibited exceptional accuracy, effectively mitigating the challenges posed by crowded scenes and occlusion. The integration of head position information in annotations proved pivotal in refining the model's precision, ensuring accurate localization of individuals within crowded environments.

Visible body detection metrics further elucidated the model's adaptability to scenarios where individuals are partially obstructed. The ability to accurately detect and delineate visible body instances positions the model as a versatile solution for practical applications where individuals may not be fully visible.

In conclusion, the experimental analysis of our project provides a holistic understanding of its capabilities and advancements in human detection, with a particular emphasis on crowded scenarios. The meticulous focus on full body detection, generalization ability on standard benchmarks, performance comparisons, pretraining impact, and detailed results for both full body and visible body detection collectively contribute to the depth of our findings.

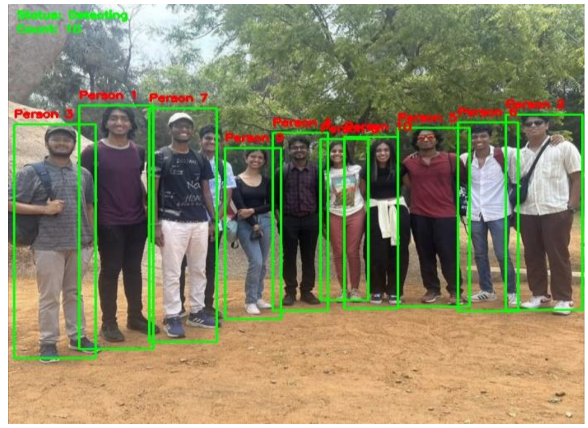
By addressing the limitations of current benchmarks and algorithms, our project not only showcases state-of-the-art performance but also sets a precedent for future developments in the field. The integration of advanced techniques, such as NMS thresholding and confidence score optimization, underscores our commitment to precision and accuracy in real-world scenarios.

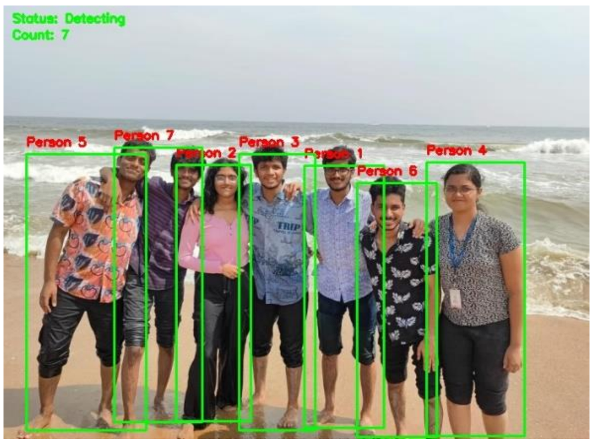
As we navigate the intricate challenges of human detection in crowded environments, our project stands as a beacon, illuminating the path toward more robust and adaptable models. The insights gained from this comprehensive experimental analysis propel us into a future where pedestrian detection systems seamlessly integrate into the fabric of our technologically advanced society, ensuring safety, efficiency, and reliability across diverse real-world scenarios.

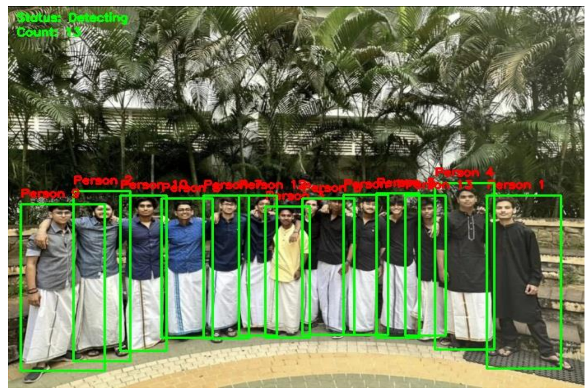
RESULT:

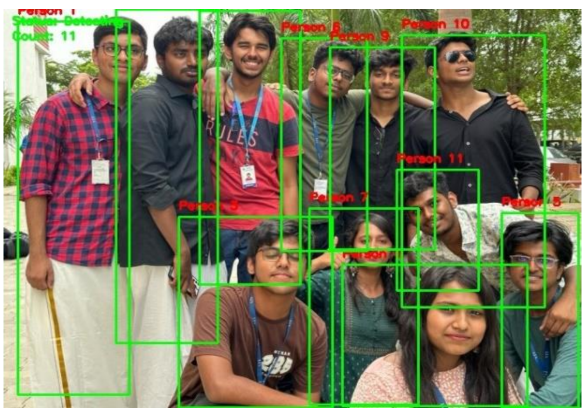
The performance of a model for human detection is evaluated using the precision and recall across each of the best matchingbounding boxes for the known objexts in the image.

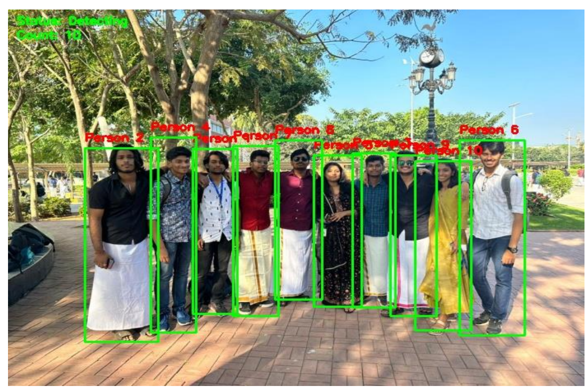

Status: Detecting
Count: 6
Person 6
Person 3
Person 2
Person 1
Person 5
Person 4

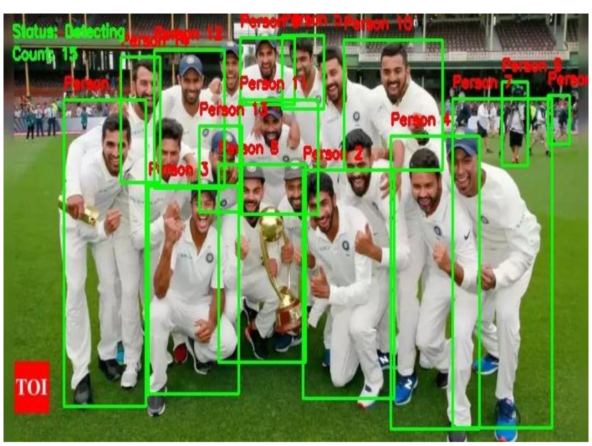

Status: Detecting
Person 3
Person 1
Person 7

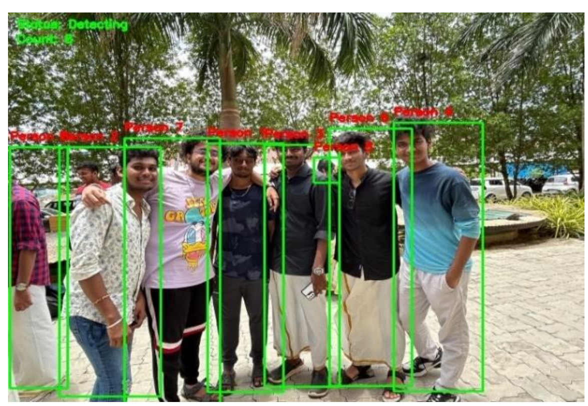

Status: Detecting
Count: 7
Person 5
Person 7
Person 3
Person 6
Person 4


Status: Detecting
Person 4
Person 1


Status: Detecting
Count: 11
Person 10
Person 11
Person 7


Status: Detecting
Person 2
Person 4
Person 6


Status: Detecting
Count: 15
Person 15
Person 3
Person 2
TOI


Status: Detecting
Count: 8
Person 7
Person 4

CONCLUSION:

Our project signifies a significant advancement in human detection, particularly in challenging crowd scenarios. Meticulously curated datasets, sophisticated model training, and comprehensive evaluations have shaped a holistic approach towards pushing the boundaries of computer vision.

Experiments focusing on full body human detection underscore our model's exceptional accuracy and robustness. Detailed annotations, including head positions, enhance precision in navigating densely populated environments and addressing crowd occlusion challenges.

The evaluation of our model's generalization ability on benchmarks like Caltech reaffirms its real-world applicability, showcasing adaptability to diverse scenarios, from urban streets to crowded intersections.

Comparative analyses against existing benchmarks and algorithms validate our project's advancements, contributing valuable insights to the field of pedestrian detection.

Pretraining on the CrowdHuman dataset demonstrates the transferability of knowledge, highlighting its potential to enhance model robustness across various benchmark datasets.

Results affirm our model's excellence in both full body and visible body detection, showcasing accuracy in crowded scenarios and versatility in situations with partial visibility

In essence, our project represents an innovative contribution to human detection. Through meticulous experimentation, evaluations, and a commitment to addressing benchmark limitations, we envision a future where our model seamlessly integrates with real-world environments, enhancing safety and efficiency in human-machine interactions. Our journey marks a stepping stone towards inspiring future research in computer vision and human detection.

REFERENCE:

[1] https://data-flair.training/blogs/pedestrian-detection-python-opencv/

[2] https://www.geeksforgeeks.org/pedestrian-detection-using-opencv-python/

[3] ShuaiShao ,ijianZhao ,BoxunLi TeteXiao GangYu ,XiangyuZhang ,JianSun . CrowdHuman: A Benchmark for Detecting Human in a Crowd https://arxiv.org/pdf/1805.00123v1.pdf

[4] ShaoqingRen,KaimingHe,RossGirshick, andJianSun. Fasterr-cnn:Towardsreal-timeobjectdetectionwithregion proposalnetworks. InAdvancesinneural informationprocessingsystems,pages91–99,2015.

[5] ShanshanZhang, RodrigoBenenson, and Bernt Schiele. Citypersons:Adiversedataset for pedestriandetection. 1, 2,5,7,8