Crowd Management System

By :
Himani Arora (22070122079)
Himanshu Chopade (22070122080)

Under the Guidance of
Dr. Nilima Zade

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)
2024

Content

- Introduction
- Literature Review
- Objective and Problem Statement
- Methodology
- Demo Results
- Test Results
- Comparison with Existing System
- Conclusion and Future Scope
- References

Introduction

- Crowd counting estimates the number of individuals in images or videos, revolutionized by deep learning for accuracy and automation.

- Importance of Accurate Crowd Counting:
    - Event Management : Ensuring safety and smooth crowd flow during event
    - Crowd Safety : Crucial role in crowd safety during heavy gatherings in crowded places.

Literature Review

- A paper in 2019 introduced a pyramidal architecture and multi-subnets for crowd counting using PaDNet. While it showed some effectiveness, the need for better recognition and simplified network architectures was highlighted.
- In 2021, the LUDA Methodology had some promise in crowd counting by estimating directly object center points. Yet the challenges continued in very crowded scenes and with reduced dependency on specific annotations.
- An Improved YOLOv4 emerged in 2021, enhancing human detection in video surveillance. Yet, it remained incomplete, particularly in low-resolution images, prompting calls for improved robustness and scalability.
- In 2022, a residual learning-based solution showcased promising results for crowd counting. However, limitations arose from dataset size and varying weather conditions, necessitating optimization for robust performance.

Literature Review

- CSRNet (2020) demonstrated near real-time crowd counting capabilities but struggled with sparse crowds and environmental factors.
- Head detection and regression algorithms (2018) achieved promising results but faced challenges in data availability and generalization.
- Object detection-based face counting (2022) faced dataset and backbone limitations, necessitating future improvements.
- In a 2020 survey of counting algorithms and datasets, the call for more research, new datasets, and robust average precision techniques was made.
- Similarly, object detection and tracking algorithms, observed in 2021, demonstrated high precision results but grappled with appearance dependency and dataset limitations, suggesting further research to mitigate these challenges.

Literature Review

Sr No.	Algorithm / Method	Data Set Used	Output	Limitations
1.)	1. Face detection using improves yolo v4 for human face detection.	Wider-face (32203 faces, with 393703 annotated faces)	Mean Average precision (mAP) of 92.1% at 18 FPS	1. Only a single dataset used – Wider Face. 2. Only a specific Backbone network used – ResNeXt
2.)	It focuses on estimating the center points of the objects directly. Algorithm used: 1. LUDA (Locally Uniform Distribution Assumptions) 2. Wider Face Benchmark	1. Wider-face 2. Shanghaitech - has approx 482 images 3. NWPU-Crowd- Around 2 million annotated heads 4. CARPK and PUCPR-Vehicle Detection from drone view.	10% +improvement in AP & 31.2% reduction in counting error on the WiderFace benchmark. Best results on crowd counting and localization datasets (ShanghaiTech and NWPU-Crowd)	1.) Reliance on point-level annotations. 2.) Increased complexity during training. 3.) Sensitivity to initialization. 4.) Limited generalization to diverse scenes.
3.)	SURVEY	1. PASACAL VOC 2. ImageNet - Large scale dataset 3. MS COCO : Addresses ImageNet limitations	mostly in Average precision	No limitation as this a survey
4.)	1. Head Hunter which is designed for small head detection in crowded scene. 2. Head Hunter- T designed to track head, it extends HeadHunter with a Partile FIlter framework. 3. IDEucl: new metric proposed for evaluation purpose and to measure algorithm's efficiency.	Crowd of Heads Dataset (CroHD), consisting of 9 sequences of 11,463 frames with over 2,276,838 heads and 5,230 tracks annotated in diverse scenes	SCUT-HEAD trained for 20 epochs, achieving precision of 0.95 compared to Faster-RCNN's 0.87. - CroHD trained for 25 epochs with a learning rate of 0.0001. - Achieved MODA of 50.0, MODP of 47.0, and mAP COCO of 19.7.	Limited generalizability due to dataset dependency. Metrics may not fully represent real-world scenarios. Fixed epoch training without adaptive strategies.
5.)	1.) Instance Set prediction for pair representation (Ci, Li). 2.) EMD Loss for minimizing prediction gaps. 3.) Set NMS to avoid redundant suppression.	1. CrowdHuman 2. CityPerson (containes 5000 images; 2975 for training, 500 for validation, 1525 for testing) 3. COCO	mAP is 90.7% evaluated on CrowsHuman. When evaluated on CityPerson, out of total 99481 person, 64153 were detected which is highest compared to others, 96.1% mAP. On COCO, 38.5% AP	1.) Challenges in detecting heavily crowded scenes. 2.) Risk of false positives due to multiple instance predictions. 3.) Increased computational overhead with optical refinement. 4.) Lower average precision on COCO dataset indicates potential generalization issues.
6.)	The proposed approach enhances YOLOv4 with: 1. Improved backbone (CSPDarkNet-53). 2. Enhanced feature aggregation and small object detection (SPP and PANet modules). 3. Improved detection accuracy (CIOU_loss loss function and DIOU_nms NMS). 4. Reduction in parameters and feature loss (Ghost CBM, CSP Module enhancements, and SSP modifications).	UCF101: 13,320 videos, 320x240 resolution, 101 behavior classes. HMDB51: 6,849 videos, 320x240 resolution, 51 behavior classes. UTI: 20 video clips, 720x480 resolution, 6 behavior categories. CASIA: Parking lot behavior dataset, 320x240 resolution, single-person behaviors.	**Comparison of detection performance 1. UTI**- 93.8 **2. UCF101** - 91.5 **3. HMDB51**- 90.6 **4. CASIA- 84.1** **Comparison of detection speed 1. UTI**- 0.627 **2. UCF101**- 0.763 **3. HMDB51**- 0.765 **4. CASIA**- 0.781	**Incomplete Human Detection:** There are instances of incomplete human detection. **Low Resolution Image Handling.**
7.)	**PaDNet** (Pan-Density Crowd Counting Algorithm): Utilizes a pyramidal architecture with multiple subnetworks for multi-scale ability. **Feature Fusion Network (FFN)** for effective density map generation	**1. ShanghaiTech Dataset 2. UCSD 3. UCF_CC_50 4. UCF-QNRF**	PaDNet achieves the best performance among all approaches on the ShanghaiTech Dataset. .	1.) PaDNet-1 has limitations in recognizing varying crowd densities effectively. 2.) Performance of PaDNet-1 is inferior to higher versions like PaDNet-3 and PaDNet-4.

Literature Review

8.	Class-Conditioned Uncertainty Guided Residual Learning.	JHU-CROWD++ Large-Scale Crowd Couting Dataset. '4,372' images with '1.51 million' annotations.	Improve performance on JHU-CROWD++ dataset with the proposed method.	1. Limited Dataset Size 2. Weather Conditions
9	LSC-CNN, which is a dense detection system for crowd counting	Specific dataset not provided	LSC-CNN framework outperforms the baseline density regression method in pinpointing people consistently across different types of crowds.	Improving the head sizing in dense crowd
10	Deep convolutional neural network (DCNN) based system for real-time crowd counting. CSRNet model which is deep learning architecture for accurate head count	Shanghai tech Dataset(diverse set of images for crowd counting)	mAE- 68.2 mSE- 115.0	Potential inaccuracies in crowd counting, mostly in sparse crowd.
11	CNN for head detection, and regression for crowd counting in dense crowd	1. UCF_CC_50 2. ShanghaiTech 3. AHU-Crowd	mAP-94%	The need of a sufficient amount of labeled training data and the generalization of the algorithm to lower density crowd scenes.
12	**Deep CNN** model used for crowd counting. **LBP features and ridge regressor** used for crowd counting.	1. **UCF CC 50** **2. WorldExpo'10**	Proposed CNN model outperforms existing method. Result show effectiveness on UCF CC 50 and WorldExpo'10.	Deep models have not been explored extensively for crowd counting.

Literature Review

Sr No.	Algorithm / Method	Output	Limitations
1.)	YOLO	mAP-92.1% at 18 FPS	1. Small Object Detection. 2. Handling Overlapping
2.)	1. LUDA (Locally Uniform Distribution Assumptions)	31.2% reduction in error - WiderFace benchmark. Best results on (ShanghaiTech and NWPU-Crowd).	1.) Reliance on point-level annotations. 2.) Increased complexity during training. 3.) Sensitivity to initialization. 4.) Limited generalization to diverse scenes.
3.)	HeadHunter & HeadHunter-T	mAP - 95%	1. Evaluated on their own dataset. 2. Used their own evaluation metric.
4.)	CSP-DarkNet-53	85-91% detection comparsion	1. Resource-Intensive 2. Small Object Detection. 3. Depends on quality of dataset
5.)	PaDNet (Pan-Density Crowd Counting Algorithm)	Average results on ShanghaiTech dataset.	1. Complex Network Architecture. 2. Recognizing different density crowd effectively.
6.)	Deep CNN	Effective working on UCF CC 50 and WorldExpo'10 dataset	1. Dataset dependent. 2. Deep Models have not been explored.
7.)	LSC-CNN	LSC-CNN outperforms the baseline density regression method in pinpointing people across crowd.	1. Need to address spurious detection. 2. Improving accuracy

Sr No.	Dataset	Limitations
1.)	WiderFace	Limited variation, biases from single dataset, annotation quality variability, scalability challenges.
2.)	LUDA, Wider Face Benchmark	Reliance on point-level annotations, training complexity, limited generalization.
3.)	SCUT-HEAD, CroHD	Limited generalizability, metrics may not reflect real-world scenarios, fixed epoch training.
4.)	CrowsHuman, CityPerson, COCO	Challenges in crowded scenes, risk of false 7.), UCF CC 50, WorldExpo'10, Limited deep model exploration. positives, increased computation, generalization issues.
5.)	UTI, UCF101, HMDB51, CASIA	Incomplete detection, low resolution handling.
6.)	ShanghaiTech Dataset	Limitations in density recognition, varying model performance.
7.)	UCF CC 50, WorldExpo'10	Limited deep model exploration.

After conducting research across 11 papers, we have compiled data on various datasets along with their corresponding precision metrics and the research papers in which they were utilized.

Objective and Problem Statement

- A Crowd Management System to address the inaccuracies in manual crowd counting at public gatherings, enabling real-time monitoring, efficient resource allocation, and enhanced security measures.

Objective:

- Analyze different datasets & existing systems and choosing the best for our project.
- Implement different algorithms on the existing systems and datasets.
- Comparing it with the results published.

Methodology

YOLOV5:

- The YOLOv5 by Ultralytics emphasizes simplicity, speed, and accuracy of object detection.
- It has high efficiency, which provides real-time inference.
- It follows single stage detection process. This simplifies the architecture.
- It is relatively easy to use, and easily available for beginners.
- It can be seamlessly integrated with different libraries.

Implementation:

- YOLOv5 was modified to detect faces by training on the WiderFace dataset, which includes annotations for face bounding boxes.
- Batch size: Fixed to 32 for better accuracy and precision

Methodology

- <u>Image Size</u>: Fixed at 1024x1024 for clear photo in training and better model development
- <u>Epochs</u>: Initial training limit has been increased from 15 to 300 based on performance evaluation
- <u>Validation</u>: 400 images with labels used for validation

Methodology

Dataset Used: WiderFace

- The WiderFace dataset is a large-scale face detection benchmark consisting of 32,203 images and 393,703 labeled faces.
- It is one of the most widely used datasets for evaluating face detection algorithms due to its diverse range of scenes, variations in pose, occlusions, and lighting conditions.
- The dataset contains annotations for face bounding boxes, providing ground truth labels for training and evaluation purposes.
- We chose the WiderFace dataset for its comprehensive coverage of real-world scenarios, making it suitable for our face detection task.

Methodology

Dataset Analysis:

- The dataset comprises images with varying resolutions, ranging from low to high quality.
- Distribution of object sizes: The dataset contains faces of different sizes, with variations in scale and aspect ratio.
- We reduced the number of images to 1000 images.

Demo Results

- Results Obtained from Testing:

Before

After

The image was captured during a real-life college event and we used it to test our crowd counting model.

Model summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs

image 1/1 C:\Me\Crowd-Counting\yolov5\2.jpg: 480x640 **55 faces**, 130.3ms

Speed: 1.0ms pre-process, 130.3ms inference, 1.5ms NMS per image at shape (1, 3, 640, 640)

Demo Results

- Results Obtained from Testing:

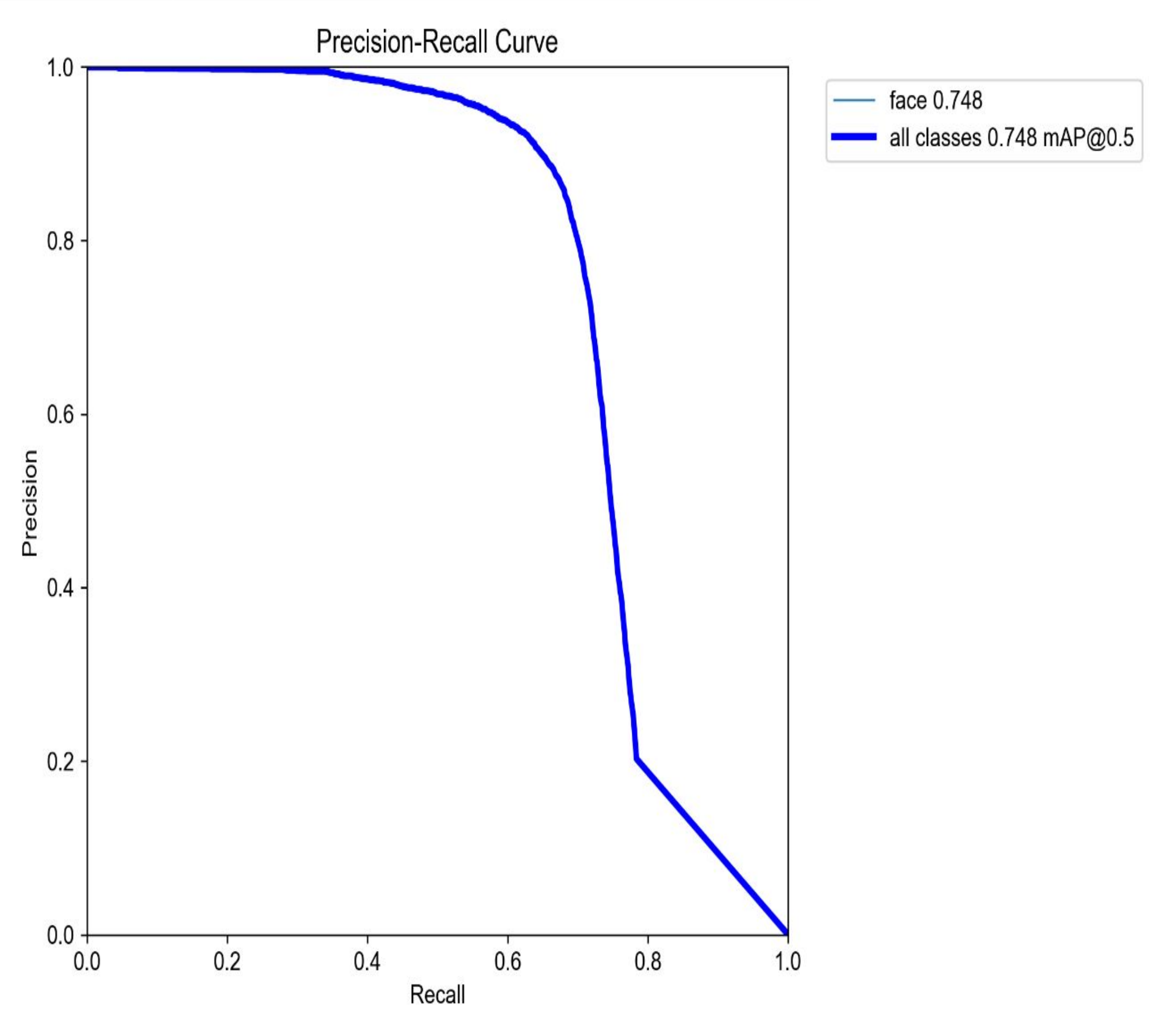
Before

After

The image was captured during a real-life event and we used it to test our crowd counting model.

Model summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
image 1/1 C:\Me\Crowd-Counting\yolov5\8.jpg: 480x640 **154 faces**, 108.1ms
Speed: 1.0ms pre-process, 108.1ms inference, 2.0ms NMS per image at shape (1, 3, 640, 640)

Test Result



- **Precision (Face):** The model achieved a precision of 0.748 for faces. This means that out of all the samples predicted as faces, approximately 74.8% of them are actual faces.
- **Recall (Face):** The model also attained a recall of 0.748 for faces. This indicates that the model can identify nearly 74.8% of all the faces present in the dataset.
- **mAP@0.5 (Across all classes):** The mean average precision (mAP) at an IoU threshold of 0.5 is 0.748. This implies that the model achieved an average precision of nearly 74.8% across all classes when compared against the ground truth.

Test Results

- Accuracy Results Obtained from Testing:
  - mAP50 : 0.75438
  - mAP50-95: 0.35334
  - Precision: 87.714%

Experiment	Image Size	Batch Size	Epochs	Rotation (Degrees)	Accuracy
1	640	4	15	0°	81.3%
2	640	16	100	0°	83.0%
4	1024	64	350	10°	86.0
3	1024	64	300	0°	87.7%

mAP50: Average precision of 75% at 50% overlap.

mAP50-95: Average precision across different overlap thresholds.

Precision: Accuracy of 87.7% in object detection.

Comparison with Existing System

Model	mAP
PSDNN	60.5%
Retinaface	61.55%
Yolo-faces	69.3%
HOANG	75.4%
YOLOv5 (ours)	87.7%

- Compared with other systems like PSDNN, Retinaface, and Yolo-faces, our YOLOv5 variant shows a tremendous leap in accuracy; it outperforms them by a great deal.
- Our YOLOv5 implementation shows remarkable 87.7% which compared to others is higher in terms of accuracy.

Conclusion and Future Scope

Conclusion: Our Crowd Management System provides real-time monitoring and precise counting for efficient and safe event execution. Its key advantages include accuracy, rapid responsiveness, safety focus, and data analytics—marking a crucial advancement in technology-driven crowd management.

Future Scope: Our future goals include enhancing accuracy and processing speed for real-time crowd counts, pivotal for traffic management. It also includes implementing different algorithms on dataset.

References

[1] Y. L. J. Z. M. I. a. J. Z. W. Yukun Tian, "PaDNet: Pan-Density Crowd Counting," in *IEEE Transactions on Image Processing*, 2020.

[2] J. H. X. H. L.-P. C. Yi Wang, "A Self-Training Approach for Point-Supervised Object Detection and Counting in Crowds," in *IEEE Transactions on Image Processing*, 2021.

[3] J. Y. G. X. Y. J. G. X. M. S. Xuan Zhou, "Human Detection Algorithm Based on Improved YOLO v4," in *Information Technology and Control*, 2022.

[4] R. Y. V. M. P. Vishwanath A. Sindagi, "JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[5] S. P. V. V. P. T. A. B. D. N. Ujwala Bhangale, "Near Real-time Crowd Counting using Deep Learning Approach," in *Procedia Computer Science*, 2020.

[6] S. M. H. S. Y. A. S.-C. S. C. Mamoona Birkhez Shami, "People Counting in Dense Crowd Images Using Sparse Head Detections," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[7] H. L. X. W. a. X. Y. Cong Zhang, "Cross-scene crowd counting via deep convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[8] A. B. Y. S. M. A. A. A. A. a. A. M. R. Alharbey, "Human Detection Algorithm Based on Improved YOLO v4," *Information Technology and Control,* 2022.

[9] W. O. X. W. P. F. J. C. X. L. M. P. Li Liu, "Deep Learning for Generic Object Detection: A Survey," *International Journal of Computer Vision,* 2020.

[10] C. D. A. B. E. M. J. P. Ramana Sundararaman, "Tracking Pedestrian Heads in Dense Crowd," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021.

[11] A. Z. X. Z. J. S. Xuangeng Chu, "Detection in crowded scenes: One proposal, multiple predictions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.

Thank You!