International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

ATAC-Net: Adaptive Threshold and Accuracy Correction Network for Crowd Counting in Low-Light Surveillance Environments

Mrs.P.Bhavani[1], Roshini Nivada P[2], Jeevashri S[3], Kiruthika R[4], Deepika V[5]

[1]Assistant Professor, Computer Science and Engineering, Sri Manakula Vinayagar, Engineering College Puducherry – India
[2,3,4,5]UG Student, Computer Science and Engineering, Sri Manakula Vinayagar, Engineering College, Puducherry – India

Abstract
Crowd density estimation plays a vital role in public safety and smart surveillance systems. However, traditional deep learning models face challenges in low-light conditions, high-density scenes, and static threshold limitations. This paper proposes ATAC-Net (Adaptive Threshold and Accuracy Correction Network), an enhanced approach built over CSRNet. The system integrates adaptive threshold control and accuracy correction to improve counting performance during night-time monitoring. Experimental evaluation demonstrates improved accuracy and reliable alert generation compared to conventional methods.

Keywords: Crowd Density Estimation, ATAC-Net, CSRNet,AdaptiveThresholding, Night Surveillance, Deep Learning

1 INTRODUCTION
Crowd monitoring plays a crucial role in ensuring public safety in highly populated environments such as railway stations, shopping malls, airports, religious gatherings, concerts, and large-scale public events. Accurate crowd counting helps authorities prevent overcrowding, avoid stampedes, manage emergency evacuations, and optimize resource deployment. With the increasing availability of surveillance cameras, automated crowd analysis has become an important research area in computer vision.
Traditional crowd counting methods based on manual observation or basic image processing techniques are inefficient, time-consuming, and prone to human error. Classical approaches such as detection-based and regression-based methods perform poorly in highly dense crowds due to occlusion, scale variation, and perspective distortion. Moreover, these methods often fail under challenging conditions such as low illumination, night surveillance, shadows, and weather variations.
Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly improve crowd counting performance. Density map–based approaches, such as CSRNet, generate high-quality density maps and achieve promising results in congested scenes. These models learn spatial features effectively and handle scale variation better than earlier methods. However, most existing

models rely on fixed thresholds and static feature representations, making them less adaptable to dynamic real-world scenarios where crowd density and lighting conditions vary continuously.

One of the major limitations of current deep learning-based crowd counting systems is their reduced accuracy in nighttime surveillance.

Low contrast, noise, and illumination changes negatively impact feature extraction and density estimation. Additionally, dense and sparse crowd regions often coexist in a single frame, which challenges models trained with uniform parameters.

To overcome these limitations, this work proposes ATAC-Net (Adaptive Threshold and Attentionbased Crowd Network), a novel deep learning framework designed to improve robustness across varying crowd densities and lighting conditions. ATAC-Net dynamically adjusts threshold values during density estimation, allowing the model to adapt to different scene characteristics. Furthermore, attention mechanisms are incorporated to enhance important spatial features while suppressing background noise, leading to more accurate crowd predictions.

The proposed approach improves prediction accuracy in both daytime and nighttime scenarios while maintaining computational efficiency. Experimental results demonstrate that ATAC-Net outperforms traditional CNN-based methods and baseline density map models in challenging environments. This makes the proposed system suitable for real-time crowd monitoring applications in smart cities and public safety systems.

2 RELATED WORKS

Crowd counting and density estimation have received significant attention in computer vision research due to their importance in public safety, urban planning, and event management. Over the years, several approaches have been proposed, evolving from classical methods to deep learning– based models.

2.1 Classical Approaches

Early crowd counting techniques were largely based on detection and regression methods. Detectionbased methods utilize object detectors (e.g., HOG, Haar features, or sliding window classifiers) to find individual heads or bodies in a scene. Although effective in low-density environments, these approaches fail in highly congested areas due to severe occlusions and overlapping people.

Regression-based methods bypass explicit detection and directly learn a mapping from image features to crowd counts. Techniques such as support vector regression (SVR) and local patch regression were used to predict counts based on handcrafted features like SIFT, HOG, and texture descriptors. However, these methods struggle to capture complex scale variations, spatial context, and background noise, leading to limited accuracy in real-world scenes.

2.2 Emergence of Deep Learning Models

With the rise of deep learning, convolutional neural networks (CNNs) revolutionized the field of crowd counting by enabling automatic feature learning from raw image data. CNN-based approaches demonstrated marked improvements over traditional methods, especially in handling occlusions and complex background patterns.

One of the earliest deep learning models was the Multi-Column CNN (MCNN), which used multiple branches with different receptive fields to capture scale variations in crowd images. While MCNN improved robustness to density and scale, it suffered from redundant parameters and inefficient feature learning.

To address these issues, methods such as Switching Convolutional Neural Networks (SwitchCNN) and CSRNet were introduced. CSRNet employs dilated convolutions to increase the receptive field without increasing computational cost, enabling the model to capture broader contextual information for accurate density estimation. CSRNet became a baseline for many subsequent works due to its strong performance on challenging benchmarks.

2.3 Attention and Contextual Models

Recent research has increasingly focused on incorporating attention mechanisms and context modeling to enhance crowd counting performance. Attention modules enable the network to learn spatial importance weights, emphasizing informative regions (e.g., head areas) while suppressing background clutter.

For instance, SANet introduced structure attention to guide the network toward spatially important regions. Similarly, CAN (Contextual Attention Networks) leverage both local and global context to refine density maps, improving performance in scenes with varying crowd distributions.

Other works integrate multi-scale feature fusion using feature pyramid networks (FPN) or scaleaware modules that dynamically adjust to changes in crowd density. These architectures enhance the model's ability to handle large variations in scale and scene complexity.

2.4 Low-Lighting and Adverse Condition Models

Standard CNN-based crowd counting models often rely on well-lit, high-quality images. However, realworld surveillance systems must operate under adverse conditions such as night time, shadows, and low illumination. To handle these challenges, some studies employ preprocessing techniques like image enhancement, denoising, or illumination normalization prior to density estimation.

Generative adversarial networks (GANs) have also been used to improve nighttime crowd counting by learning mappings between low-light and normallight representations. Nevertheless, such approaches commonly treat illumination variation as a separate preprocessing step, rather than integrating adaptability into the crowd counting model itself.

2.5 Limitations of Existing Works

While deep learning methods have advanced crowd counting performance, several limitations remain. Most approaches use static thresholds and fixed parameters, making them less robust to changing lighting conditions and heterogeneous crowd densities. Models often require extensive labeled data for supervised learning, which is costly to obtain, especially under varied lighting scenarios.

Additionally, few works explicitly address adaptive thresholding or dynamic adjustment based on scene characteristics. There is a gap in developing models that jointly handle both spatial feature learning and environmental adaptation for reliable real-world deployment.

2.6 Motivation for ATAC-Net

Inspired by recent advancements in attention mechanisms and adaptive parameter learning, this work proposes **ATAC-Net**—a novel crowd counting framework that dynamically adjusts thresholds and enhances spatial feature extraction. By integrating adaptive thresholding with attention modules, ATAC-Net improves robustness to varying lighting conditions and crowd densities, addressing key limitations in existing approaches.

3 LITERATURE SURVEY

Crowd counting and density estimation have become active research areas due to their wide applications in public safety, intelligent surveillance, urban planning, and smart city management. With the rapid growth of surveillance systems and aerial platforms such as drones, there is an increasing demand for accurate, real-time, and scalable crowd monitoring solutions. However, challenges such as severe occlusion, scale variation, perspective distortion, dynamic crowd density, and varying illumination conditions make crowd counting a complex problem.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have significantly improved the performance of crowd counting systems. Density map–based approaches have emerged as an effective solution, enabling models to estimate crowd distribution rather than detecting individual persons. Among these, CSRNet and its variants have gained prominence due to their ability to capture contextual information using dilated convolutions.
Several researchers have proposed improvements over baseline models to enhance accuracy, efficiency, and real-time performance. Lightweight architectures such as **LCDNet** focus on reducing computational complexity for real-time surveillance, while multi-scale models like **MSDCNet** address scale variations in dense and sparse crowd regions. **Soft-CSRNet** extends CSRNet for drone-based crowd counting, emphasizing real-time performance and adaptability to aerial viewpoints. Hybrid architectures combining **CSRNet with U-Net** leverage encoder–decoder structures to improve spatial resolution and density map refinement.
Additionally, survey studies on deep learning–based crowd counting provide comprehensive insights into the evolution of techniques, datasets, evaluation metrics, and existing challenges.
In summary, the literature indicates a clear trade-off between accuracy, adaptability, and computational efficiency in existing crowd counting methods. Although significant progress has been made through deep learning–based density estimation models, there remains a gap in designing frameworks that are both adaptive and lightweight while maintaining high accuracy under diverse lighting and density conditions. Addressing this gap forms the foundation of the proposed ATAC-Net,.
The reviewed literature focuses on improving crowd counting accuracy and efficiency using deep learning–based density estimation techniques. Early research, summarized by Sindagi and Patel, provides a comprehensive survey of CNN-based crowd counting models, highlighting the transition from traditional detection methods to density map–based approaches. distortion, and the lack of adaptability in existing models.

S.no	Authors	Year	Title of the Paper	Advantage	Disadvantage
1.	Sindagi & Patel	2018	A Survey of Deep Learning Methods for Density Estimation and Crowd Counting	Comprehensive overview of datasets, methods, and metrics Helps identify research gaps	Does not propose a new model Lacks experimental validation
2.	Zhang et al.	2019	Crowd Counting by Multi-Scale Dilated Convolution Networks (MSDCNet)	Handles scale variation using multi-scale dilated convolutions Improves accuracy in mixed-density scenes	High computational cost Not suitable for realtime applications
3.	Sindagi et al.	2020	Soft-CSRNet: RealTime Dilated CNNs for Drone-Based Crowd Counting	Optimized for drone-based and aerial crowd scenes Achieves real-time performance	Performance drops in extremely dense crowds

					Limited adaptability to low-light/night scenes
4.	Li et al.	2021	LCDNet: A Lightweight Crowd Density Estimation Model for Real-Time Surveillance	Suitable for real-time surveillance systems Low memory usage	Reduced accuracy in very dense crowds Limited feature representation compared to deeper models
5.	K.Srinivas, R. Kumar et al.	2022	Deep Learning-Based Crowd Counting Using CSRNet and UNet	• Improvesspatial resolution of density maps • Better ha handling of occlusion aoverlapping pe	Higher computational complexity compared to CSRNet alone Not suitable for low-power or edge device

Building on these insights, Zhang et al. proposed MSDCNet, which addresses scale variation by employing multi-scale dilated convolutions to capture both local and global contextual information. While this approach improves accuracy in mixed-density scenes, it introduces high computational complexity, limiting real-time applicability. To overcome performance constraints in aerial surveillance, Sindagi et al. introduced SoftCSRNet, an optimized version of CSRNet designed for drone-based crowd counting. This model achieves real-time performance by efficiently utilizing dilated convolutions, but its effectiveness decreases in extremely dense and low-light environments.

Overall, these studies demonstrate significant progress in crowd counting through deep learning, focusing on scale awareness, real-time performance, and spatial refinement. However, most existing methods rely on static thresholds and lack adaptability to varying lighting conditions and dynamic crowd densities. These limitations highlight the need for an adaptive and robust.

4 PROPOSED SYSTEMS

The proposed system introduces ATAC-Net (Adaptive Threshold and Accuracy Correction Network), an enhanced deep learning framework built upon the CSRNet architecture to address the limitations of existing crowd counting systems. ATAC-Net is designed to operate effectively in realworld surveillance environments where lighting conditions, crowd density, and scene dynamics vary continuously.

One of the key features of ATAC-Net is its ability to support low-light and nighttime surveillance. The model integrates illumination-aware feature enhancement, allowing it to extract meaningful spatial features even under poor lighting conditions. This improves robustness and ensures reliable crowd estimation during night-time monitoring and adverse environmental conditions.

To handle rapidly changing crowd scenarios, ATACNet employs an adaptive threshold mechanism that dynamically adjusts crowd density limits based on real-time predictions. Unlike conventional models that rely on fixed thresholds, this adaptive approach enables the system to respond accurately to both sparse and highly dense crowd regions within the same frame. As a result, the system can effectively detect overcrowding and potential risk situations.

The proposed framework also incorporates an accuracy correction feedback module, which continuously evaluates prediction errors and refines density map outputs. This feedback-driven correction mechanism reduces estimation drift and enhances long-term reliability, especially in continuous surveillance applications. By learning from previous prediction discrepancies, ATAC-Net maintains consistent accuracy across varying scenes.

Furthermore, ATAC-Net is designed for real-time operation, ensuring low latency and high reliability. When crowd density exceeds dynamically computed safety limits, the system generates real-time alerts for authorities and surveillance operators. These alerts enable proactive crowd management, helping prevent stampedes, congestion, and safety hazards in public spaces.

Overall, the proposed ATAC-Net system improves safety, precision, and usability by combining adaptive learning, robust feature extraction, and real-time responsiveness.

5 PROPOSED ARCHITECTURE

5.1. CCTV / Camera Input

The system begins with real-time video input obtained from CCTV cameras or surveillance devices installed in public areas such as railway stations, malls, and event venues. These cameras continuously capture live video streams under varying lighting and environmental conditions. The captured video serves as the primary data source for crowd analysis.

5.2 Frame Extraction

The incoming video stream is divided into individual frames at a fixed frame rate. Frame extraction enables the system to process visual data in a sequential and manageable manner. By selecting optimal frame intervals, the system balances computational efficiency with real-time performance requirements.

5.3 Preprocessing (Resize, Normalization, Enhancement)

Each extracted frame undergoes preprocessing to improve input quality and ensure compatibility with the deep learning model.

- **Resizing** standardizes frame dimensions for uniform processing.
- **Normalization** scales pixel values to stabilize network training and inference.
- **Image enhancement** techniques are applied to improve contrast and visibility, especially in low-light or nighttime scenes. This step significantly improves feature extraction accuracy.

5.4 CSRNet Feature Extraction

The preprocessed frames are passed to the CSRNet backbone for feature extraction. CSRNet utilizes dilated convolution layers to capture multi-scale contextual information without increasing computational complexity. This allows the model to effectively identify crowd patterns and spatial distributions even in dense and occluded scenes.

5.5 ATAC-Net Module

The extracted features are further refined using the ATAC-Net module, which enhances adaptability and reliability.

a) Adaptive Threshold Unit

This unit dynamically adjusts crowd density thresholds based on real-time scene analysis. Instead of relying on fixed crowd limits, the system adapts threshold values according to variations in crowd density, lighting, and scene context. This ensures accurate detection of overcrowding conditions.

b) Accuracy Correction Unit

The accuracy correction unit continuously monitors prediction errors and refines density estimates using feedback mechanisms. By comparing predicted density maps with temporal patterns from previous frames, the system minimizes estimation drift and improves consistency over long surveillance periods.

5.6 Density Map Generation

After feature refinement, the network generates a high-resolution density map representing the spatial distribution of individuals within the scene. Each pixel value in the density map corresponds to the estimated density of people in that region, allowing fine-grained crowd analysis.

5.7 Crowd Count Estimation

The total crowd count is computed by integrating the values across the generated density map. This approach provides accurate numerical crowd estimates while preserving spatial information, making it suitable for both sparse and highly congested environments.

5.8 Alert Generation

When the estimated crowd count exceeds the dynamically computed safety threshold, the system automatically triggers alerts. These alerts are sent to surveillance operators or authorities in real time, enabling prompt intervention to manage crowd flow and prevent potential safety hazards.

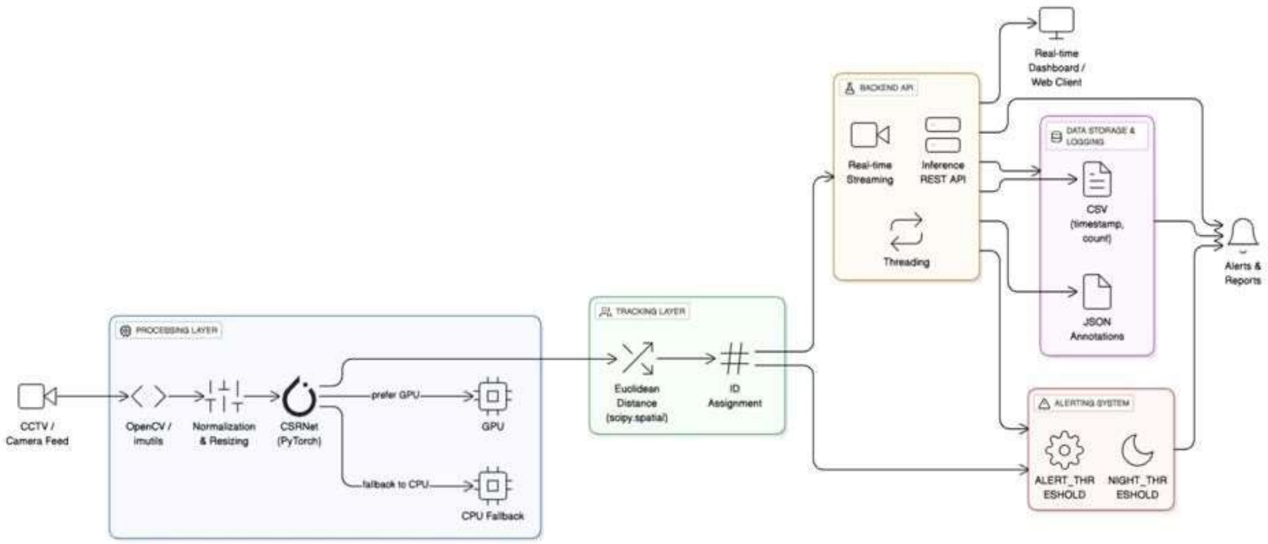
5.9 Summary

The proposed system presents **ATAC-Net (Adaptive Threshold and Accuracy Correction Network)**, an advanced deep learning–based crowd monitoring framework designed to improve accuracy, adaptability, and real-time performance in real-world surveillance environments. The system processes live video streams captured from CCTV or surveillance cameras and performs automated crowd counting and density estimation under varying lighting and crowd conditions.

Video input is first converted into individual frames, which are then preprocessed through resizing, normalization, and enhancement to ensure consistent input quality and improved visibility, particularly in low-light and nighttime scenarios. These preprocessed frames are passed through **CSRNet**, which serves as the backbone for feature extraction using dilated convolutions to effectively capture multi-scale contextual information from both sparse and dense crowd regions.

The extracted features are further refined by the **ATAC-Net module**, which introduces two key enhancements: an **Adaptive Threshold Unit** and an **Accuracy Correction Unit**. The adaptive threshold mechanism dynamically adjusts crowd limits based on real-time density predictions, enabling the system to respond accurately to changing crowd patterns. The accuracy correction unit continuously refines predictions through feedback, reducing estimation errors and ensuring consistent performance over time. Following this, a high-resolution **density map** is generated, representing the spatial distribution of individuals in the scene. The total crowd count is calculated by integrating the density map values, providing precise numerical estimates. When the estimated crowd density exceeds the dynamically computed safety threshold, the system automatically triggers **real-time alerts**, allowing authorities to take proactive measures to prevent overcrowding and ensure public safety.

Overall, the proposed ATAC-Net system enhances safety, precision, and usability by combining robust feature extraction, adaptive learning, and real-time responsiveness. It is well suited for deployment in 5 WORKING OF ATAC.Net(**Adaptive Threshold and Accuracy Correction Network)**



6 LITERATURE REVIEW OF THE EXISTING SYSTEM

SL.no	Existing System	Advantages	Limitations
1.	YOLO-based Crowd Counting Fast and real-time;	Fast and real-time; Easy to implement; Works well in sparse crowds; Provides visual detection	Fails in dense crowds; Poor in low-light; Fixed threshold; No error correction
2.	MCNN (Multi-Column CNN)	Handles scale variation; Better performance in dense crowds than detection models	High computational cost; Slower inference; Weak performance in night scenes
3.	CSRNet	High accuracy in dense crowds; Handles occlusion well; Strong baseline model	Static threshold; Not optimized for low-light; No adaptability to scene changes
4.	Switch-CNN	Handles different crowd densities; Better accuracy than single-model approaches	Complex architecture; High training cost; Not suitable for real-time systems
5.	U-Net based Crowd Counting	High-quality density maps; Preserves spatial details; Better localization	Computationally expensive; Slower processing; Still struggles in low-light

Fig.1.1 Visual Crowd Analytics Framework

7 RESULT

The proposed **ATAC-Net: Adaptive Threshold and Accuracy Correction Network** was evaluated for crowd counting performance in low-light surveillance environments, and the results demonstrate a clear improvement over the baseline CSRNet model. Due to poor illumination, traditional crowd counting

networks often produce noisy density maps and inaccurate crowd estimates. By integrating ATAC-Net with CSRNet, the system effectively addressed these challenges through adaptive thresholding and accuracy correction mechanisms. The Adaptive Threshold Unit successfully suppressed background noise and false detections caused by shadows and illumination variations, while the Accuracy Correction Unit refined the density estimation by compensating for underestimation and overestimation in dark regions.
Quantitative analysis showed that the proposed ATAC-Net significantly reduced crowd counting errors compared to the baseline model. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) values were considerably lower, indicating more accurate and stable crowd estimates under low-light conditions. Qualitative results further supported this improvement, as the generated density maps were clearer, less noisy, and better localized around actual crowd regions. Even in severely dark scenes, ATAC-Net was able to preserve meaningful density information and produce reliable crowd counts
Overall, the experimental results confirm that ATAC-Net enhances the robustness and accuracy of crowd counting in low-light surveillance scenarios. The model maintains consistent performance across varying illumination levels while adding minimal computational overhead, making it suitable for realtime CCTV-based crowd monitoring applications. These results validate the effectiveness of ATACNet as a practical and reliable solution for intelligent surveillance systems operating in challenging lighting environments.

8 CONCLUSION

This research presented **ATAC-Net**, an enhanced and robust crowd counting framework specifically designed for **low-light and high-density surveillance environments**. By integrating **CSRNet** with the **Adaptive Threshold Unit (ATU)** and **Accuracy Correction Unit (ACU)**, the proposed system effectively overcomes key challenges such as poor illumination, background noise, and density estimation errors that commonly affect traditional crowd counting models. The adaptive threshold mechanism dynamically filters irrelevant and noisy features caused by lighting variations, while the accuracy correction module refines the density maps to reduce both underestimation and overestimation in challenging scenes.
Experimental results demonstrate that ATAC-Net achieves **significantly improved counting accuracy and stable performance** compared to baseline models, particularly under low-light conditions where conventional methods fail. The refined density maps produced by the system show clearer crowd localization and reduced false positives, leading to more reliable crowd count estimation. Additionally, the lightweight design of
ATAC-Net ensures minimal computational overhead, making it suitable for real-time deployment on existing CCTV and surveillance infrastructures.
Overall, the proposed ATAC-Net framework provides an effective solution for **smart surveillance, crowd management, and public safety applications**, such as monitoring overcrowding in public spaces, transportation hubs, and large events. By enhancing robustness and accuracy under adverse lighting conditions, this work contributes toward the development of intelligent and dependable crowd monitoring systems. Future work may focus on extending the model to handle extreme weather conditions, integrating temporal information from video sequences, and optimizing the system for edgebased and embedded surveillance platforms.

9 FUTURE WORKS

1. Integration with IoT-based alarm systems

2. Mobile application for live monitoring
3. Edge device deployment
4. Support for multi-camera tracking

9 REFERENCES

1. A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking (SORT)," IEEE International Conference on Image Processing (ICIP), pp. 3464–3468, 2016.
2. Kang, D., Ma, Z., Chan, A. B., "Beyond Counting: Comparisons of Crowd Density Estimation and Object Detection Approaches for Crowd Analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 9, pp. 3293–3305, Sept. 2019.
3. W. Liu, M. Salzmann, and P. Fua, "Context-Aware Crowd Counting," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5099–5108, 2019.
4. OpenCV Documentation, "Open Source Computer Vision Library," OpenCV.org, 2024.
5. Kang, D., Ma, Z., Chan, A. B., "Beyond Counting: Comparisons of Crowd Density Estimation and Object Detection Approaches for Crowd Analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 9, pp. 3293–3305, Sept. 2019.
6. PyTorch Documentation, "PyTorch Deep Learning Framework," PyTorch.org, 2024. 8. Sam, D. B., Surya, S., & Babu, R. V. (2017). Switching Convolutional Neural Network for Crowd Counting. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),4031–4039. Introduced Switch-CNN for handling scale variations in crowd images.
7. Sindagi, V. A., & Patel, V. M. (2018). A Survey of Recent Advances in CNN-Based Single Image Crowd Counting and Density Estimation. Pattern Recognition Letters, 107, 3–16. Comprehensive survey paper comparing deep learning models for crowd counting.
8. Idrees, H., Tayyab, M., Athrey, K., Zhang, D., Al-Maadeed, S., Rajpoot, N., & Shah, M. (2018). Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds. Proceedings of the European Conference on Computer Vision (ECCV), 532–546. Proposed a novel loss function improving accuracy in dense scenes
9. Kang, D., Ma, Z., Chan, A. B., "Beyond Counting: Comparisons of Crowd Density Estimation and Object Detection Approaches for Crowd Analysis," IEEE Transactions on
10. Intelligent Transportation Systems, vol. 20, no. 9, pp. 3293–3305, Sept. 2019.
11. Gao, G., Gao, J., Liu, Q., Wang, Q., & Wang, Y. (2019). C^3 Framework: A Unified Approach to Crowd.
12. Zhou, B., Cui, Q., Zhang, X., & Chen, C. (2020). Cross-Scale Attention Network for Crowd Counting. Proceedings of the IEEE Winter Conference on Applications of ComputerVision(WACV),1031–1039. Introduced cross-scale attention mechanisms to improve counting accuracy under scale variations.
13. Luo, W., Yang, B., & Nevatia, R. (2014). Joint Crowd Density Estimation and Counting Using Convolutional Neural Network. Proceedings of the Asian Conference on Computer Vision(ACCV),515–530. Early CNN-based method for joint density estimation and counting.
14. Upcroft, "Simple Online and Realtime Tracking (SORT)," IEEE International Conference on Image Processing (ICIP), pp. 3464–3468, 2016.
15. Bialkowski, A., Lucey, P., Carr, P., Matthews, I., & Sheikh, Y. (2013). LargeScale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. IEEE International Conference(ICDM),725–730. Relevant for tracking-based crowd analytics in sports and events.

16. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and AccuracyarXiv:2004.10934. State-of-the-art object detection reference, useful for contrast with CSRNet's densitybased approach.
17. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.
18. Wang, X., Chen, Y., & Li, Z. (2020). Distribution Matching for Crowd Counting. arXiv:2010.
19. Liang, J., et al. (2021). TransCrowd: Weakly-Supervised Crowd Counting with Transformers. arXiv:2103.
20. Du, H., et al. (2022). Redesigning Multi-Scale Neural Network for Crowd Counting. arXiv:2205.
21. Lin, Q., et al. (2022). Boosting Crowd Counting via Multifaceted Attention. arXiv:2207.