A Survey of Crowd Counting and Density Estimation Techniques in Computer Vision

www.surveyx.cn

Abstract

This survey paper explores the interdisciplinary field of crowd counting and density estimation, emphasizing the use of convolutional neural networks (CNNs) within computer vision frameworks. It addresses the challenges of occlusion, scale variation, and privacy preservation in real-time crowd monitoring, crucial for public safety, event management, and urban planning. The survey systematically reviews current advancements, including scale-aware and multi-scale architectures, attention mechanisms, and the integration of CNNs with transformer architectures. These innovations enhance the accuracy of density maps and crowd estimates, even in complex, densely populated environments. Privacy-preserving frameworks and occlusion handling techniques are also examined, highlighting strategies to balance data utility with privacy and improve accuracy in occluded settings. Additionally, the paper discusses the trade-offs between speed and accuracy in real-time applications and presents algorithmic innovations that optimize performance. The practical applications of these techniques in public safety, event management, and urban planning are explored, demonstrating their impact on enhancing crowd control and resource allocation. The paper concludes by identifying limitations and suggesting future research directions to further refine these methodologies, emphasizing the need for robust algorithms and comprehensive datasets to advance the field of crowd analysis.

1 Introduction

1.1 Significance of Crowd Counting and Density Estimation

Crowd counting and density estimation are pivotal in computer vision, significantly influencing public safety and urban planning by providing accurate population density assessments in crowded settings [1, 2]. These techniques are essential for managing densely populated areas, ensuring safety, and facilitating strategic urban planning as large gatherings and urbanization trends escalate [3].

In public safety, crowd counting supplies vital data for video surveillance, traffic management, and emergency response, where real-time information is crucial for informed decision-making, particularly within smart city infrastructures [4]. Beyond numerical data, insights gained from crowd counting inform crowd dynamics, essential for applications in retail and crowd management, thereby enhancing public safety by preventing hazardous situations like stampedes [1].

Crowd counting also finds relevance in non-traditional applications, such as biological studies and during global events like the COVID-19 pandemic, where monitoring crowd sizes was critical for enforcing health and safety regulations [5]. The challenge of accurately counting individuals in unconstrained scenes, often characterized by occlusion, pose variations, and background clutter, underscores the need for robust algorithms that provide precise estimates under diverse conditions [3]. Recent advancements in machine learning, particularly deep learning models, have significantly progressed this area, reinforcing the importance of crowd counting and density estimation in modern urban environments [4].

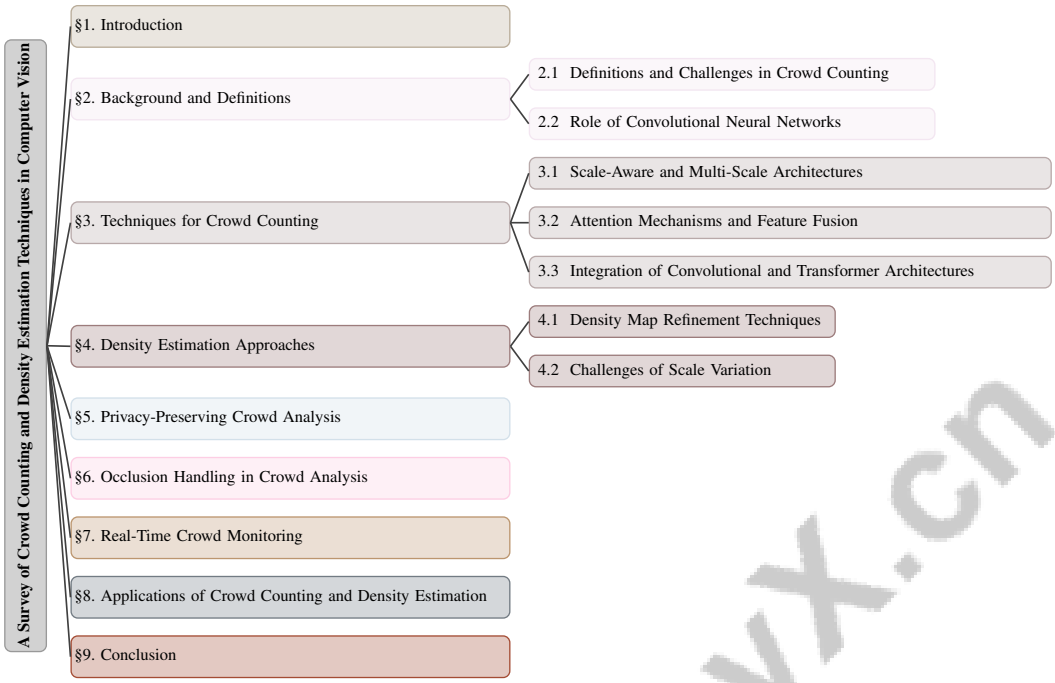


Figure 1: chapter structure

1.2 Challenges in Crowd Analysis

Crowd analysis in computer vision faces several significant challenges that impact the effectiveness of crowd counting and density estimation techniques. Occlusion is a primary obstacle, where individuals in densely populated environments are often obscured, complicating accurate counting efforts [1]. This issue is intensified by the reliance on bounding box annotations in detection-based methods, which become impractical in dense crowd scenarios [4]. Additionally, dramatic scale variations due to perspective effects introduce further complexity, leading to significant variability in the size and appearance of individuals across different images [2].

The inadequacy of existing detection and regression methods in managing both high and low-density scenes presents another challenge [1]. These methods often fail to effectively model and utilize global context information, crucial for accurate crowd density estimation. Furthermore, the extreme density variances in crowd scenes hinder the predictive capabilities of single convolutional neural network (CNN) models, resulting in inaccuracies in crowd estimation.

Privacy concerns complicate crowd analysis, necessitating innovative strategies to balance data utility with privacy preservation. The substantial computational resource requirements of current methods limit their scalability and deployment in real-time applications. The challenges of occlusion, scale variation, and privacy concerns underscore the urgent need for advanced techniques that can enhance the robustness and applicability of crowd counting methodologies. Recent studies indicate that traditional models often struggle with background inaccuracies, leading to substantial errors—up to 49

1.3 Structure of the Survey

This survey is meticulously structured to provide a comprehensive analysis of the latest advancements and ongoing challenges in crowd counting and density estimation through computer vision techniques. It reviews state-of-the-art methods, particularly those leveraging deep learning, and highlights their applications across various domains such as public safety, traffic monitoring, and urban planning. Additionally, the survey addresses significant obstacles in the field, including occlusions, non-uniform densities, and variations in scale and perspective, while categorizing impactful contributions by model architecture and evaluation metrics. This resource serves both novice and experienced researchers navigating the evolving landscape of crowd counting technologies [6, 7, 8, 9, 10].

The paper begins with an introduction that emphasizes the significance of these techniques in various domains, followed by a discussion of inherent challenges in crowd analysis, including occlusion, scale variation, and privacy concerns.

The second section, Background and Definitions, elaborates on key concepts such as crowd counting, density estimation, privacy-preserving analysis, occlusion handling, and real-time monitoring, emphasizing the role of convolutional neural networks (CNNs) in these tasks. This section sets the stage for understanding the technical complexities and terminologies used throughout the survey.

In the Techniques for Crowd Counting section, we explore various methodologies employed for crowd counting, focusing predominantly on CNN-based approaches. This includes an examination of scale-aware and multi-scale architectures, attention mechanisms, feature fusion techniques, and the integration of convolutional and transformer architectures for enhanced performance [10].

The subsequent section on Density Estimation Approaches discusses methods for estimating crowd density, addressing challenges such as scale variation and techniques used to refine density maps for improved accuracy. It provides insights into innovative strategies developed to overcome these challenges, including the pan-density crowd counting perspective [11].

Privacy-Preserving Crowd Analysis examines the balance between data utility and privacy, highlighting innovative frameworks and methods for ensuring privacy in crowd analysis, a critical consideration in contemporary applications.

Occlusion Handling in Crowd Analysis addresses the specific challenges posed by occlusions in crowd scenes. This review explores advanced techniques designed to enhance visibility and accuracy in occluded environments by integrating multi-modal and multi-view information, focusing on the synergistic use of optical and thermal images for improved crowd counting analysis, while addressing the complexities and advantages of multimodal approaches over traditional monomodal models [12, 5, 13].

Real-Time Crowd Monitoring focuses on the requirements and challenges inherent in real-time applications, evaluating the computational efficiency of various algorithms and exploring trade-offs between speed and accuracy.

Finally, the Applications of Crowd Counting and Density Estimation section discusses the practical applications of these techniques across various domains, providing examples of real-world implementations and their impact on public safety, event management, and urban planning.

The survey concludes with a comprehensive summary of key findings in crowd analysis, emphasizing recent advancements such as the development of novel crowd counting networks and large-scale datasets, identifying ongoing challenges like background errors in predictions, and proposing future research directions that include exploring multimodal data integration and improving model generalization to enhance the accuracy and effectiveness of crowd analysis techniques [6, 14, 5, 15, 10].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions and Challenges in Crowd Counting

Crowd counting involves estimating the number of individuals in a scene, typically through generating density maps from images or video frames [16]. This is crucial for public safety, urban planning, and event management, where precise density estimates inform strategic decisions [17]. Key challenges include scale variation and occlusion, which undermine the accuracy of current methods [18]. Scale variation, due to perspective effects and varying distances, complicates the task for traditional CNNs, which struggle with scale invariance [2, 19]. Although multi-column and multi-scale architectures attempt to address this, they often fail to capture the full range of scales in complex scenes [1].

Occlusion, prevalent in densely populated environments, further complicates accurate counting as individuals are often obscured by others or environmental features [4, 3]. Techniques such as multi-view data integration and feature fusion offer additional perspectives to enhance accuracy in these scenarios [5]. However, reliance on foreground extraction and hand-crafted features remains a limitation, as these are susceptible to errors from occlusions and scale variations [17]. The limited

receptive field of convolutional kernels also restricts the capture of global patterns essential for accurate analysis [19].

The scarcity of annotated datasets poses another challenge, risking overfitting and reducing generalizability to diverse environments [1]. This underscores the need for innovative approaches leveraging unlabeled data and advanced architectures to enhance reliability and accuracy in practical applications [17]. Developing robust algorithms to address these challenges is crucial for advancing crowd counting and its applications.

2.2 Role of Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are pivotal in crowd counting and density estimation, addressing challenges such as scale variation and occlusion through hierarchical feature extraction [1]. CNN-based approaches have largely supplanted traditional methods due to their superior handling of variability in crowd scenes [17]. Recent advances in CNN architectures, like the Multi-Column CNN (MCNN) and Hydra-CNN, focus on multi-scale feature extraction to tackle scale variation, though they sometimes lack accuracy across varying densities [1]. The Scale-Aware Crowd Counting Network (SACCN) enhances this by integrating regional and semantic attention with an asymmetric multi-scale module, improving scale management [4].

Further advancements include integrating CNNs with transformer architectures, which incorporate global context for better geometric and semantic information capture. Networks like the Joint CNN-Transformer Network (JCTNet) combine CNNs for feature extraction with transformers for global context modeling, enhancing performance across diverse scenarios and addressing limitations in capturing long-range dependencies [4]. Attention mechanisms refine CNN capabilities by focusing on relevant image regions, improving density estimation accuracy. Techniques such as the Fourier-Guided Attention (FGA) module, which merges frequency domain processing with spatial and channel attention, exemplify this progression, enhancing detail discernment in crowded scenes [17].

Despite these advancements, the need for extensive labeled data remains a significant hurdle, as data acquisition is labor-intensive [17]. CNN-based methods also face challenges in high-density scenes due to small target sizes and occlusion, while regression-based methods may overestimate in low-density areas [1]. To mitigate these issues, novel detection networks like PSDDN utilize point-level annotations for predicting bounding boxes and counts, offering a more efficient and accurate alternative [4].

3 Techniques for Crowd Counting

Category	Feature	Method
Scale-Aware and Multi-Scale Architectures	Scale Normalization Techniques	L2SM[18]
Attention Mechanisms and Feature Fusion	Hybrid Learning Approaches	IRAST[17]
Integration of Convolutional and Transformer Architectures	Architectural Synergy	PBL-CCNN[20], Switch-CNN[21], CCT[22], CN[23], FGCC[24], DC[25], PFDNet[2]

Table 1: This table presents a comprehensive summary of contemporary methodologies in crowd counting, categorized into three primary areas: Scale-Aware and Multi-Scale Architectures, Attention Mechanisms and Feature Fusion, and Integration of Convolutional and Transformer Architectures. Each category is associated with specific features and methods, highlighting the innovative approaches and techniques utilized to address challenges such as scale variation and feature integration in crowd counting applications. The references provided offer insights into the recent advancements and applications of these methodologies.

Accurate crowd counting requires addressing challenges such as scale variation across diverse environments. This section examines the role of scale-aware and multi-scale architectures in enhancing crowd counting methodologies, which accommodate varying object sizes and densities to improve density estimation precision. As illustrated in **??**, the hierarchical categorization of advanced techniques in crowd counting underscores the significance of these architectures. The figure highlights not only the roles of scale-aware and multi-scale architectures but also the contributions of attention mechanisms and feature fusion. Furthermore, it details the integration of convolutional and transformer architectures, demonstrating the architectural approaches, methods, and recent developments in each category. Table 2 provides a detailed overview of the key methodologies and techniques

employed in crowd counting, categorizing them into distinct areas based on their architectural and functional characteristics. This visual representation emphasizes the importance of addressing scale variation, feature integration, and hybrid model advancements to enhance accuracy and robustness in crowd counting applications.

3.1 Scale-Aware and Multi-Scale Architectures

Addressing scale variation in crowd counting necessitates architectures that manage varying object scales effectively. Multi-column networks and scale-aware designs have significantly improved estimation accuracy by tackling these variations. Multi-modal architectures, employing early, late, and deep fusion strategies, effectively integrate diverse data sources to address scale variations [5]. The Adaptive Fusion Image Pyramid Counting (AFIPC) method and SegCrowdNet exemplify approaches that adapt to different image resolutions and integrate segmentation attention, respectively, to enhance accuracy [26, 27].

Structured models like the Scale-Aware Crowd Counting Network (SACCN) and the Top-Down Feedback Convolutional Neural Network (TDF-CNN) employ asymmetric multi-scale modules and feedback mechanisms to manage scale diversity [19, 3]. The L2SM model further addresses scale variation through patch-level density maps normalized with an online center learning strategy [18].

Innovative methods such as Multi-View Fusion Networks (MVFN) and the Fourier-Guided Attention (FGA) network enhance scale management by integrating multiple camera views and employing Fast Fourier Transformations with attention mechanisms, respectively [28, 29]. DecideNet and PSDDN offer frameworks combining detection and regression, and keypoint-based detection, respectively, to manage scale variation effectively [1, 4].

Recent studies highlight the importance of scale-aware and multi-scale architectures in improving crowd counting techniques. The Scale Aggregation Network (SANet) uses an encoder-decoder structure for multi-scale feature extraction, while scale-aware attention networks dynamically focus on relevant scales, enhancing accuracy and applicability in crowd analysis [30, 31].

As illustrated in Figure 2, the hierarchical categorization of scale-aware and multi-scale architectures in crowd counting highlights three main categories: multi-modal architectures, structured models, and innovative methods. Each category showcases specific techniques and models that address scale variation challenges, such as fusion strategies, attention mechanisms, and adaptive frameworks.

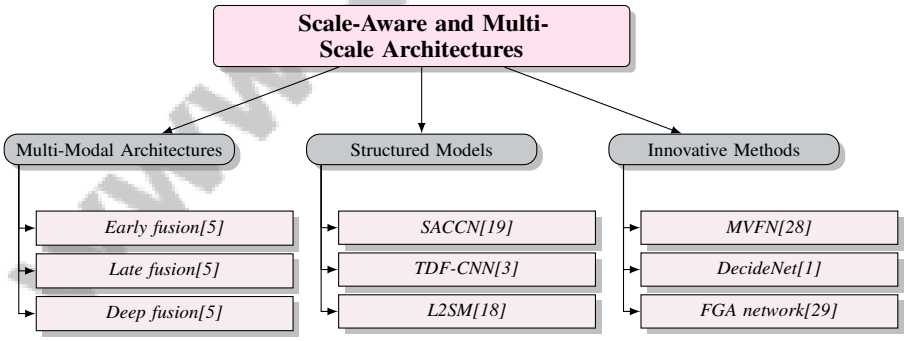


Figure 2: This figure illustrates the hierarchical categorization of scale-aware and multi-scale architectures in crowd counting, highlighting three main categories: multi-modal architectures, structured models, and innovative methods. Each category showcases specific techniques and models that address scale variation challenges, such as fusion strategies, attention mechanisms, and adaptive frameworks.

3.2 Attention Mechanisms and Feature Fusion

Attention mechanisms and feature fusion techniques are crucial for refining crowd counting models by emphasizing relevant features and scales, thus improving accuracy and robustness. These methodologies address challenges like scale variation and occlusion in crowded environments. Attention models identify head locations, while scale-aware attention networks focus on global and local scales

for enhanced density estimation. Content-aware methods further improve accuracy by incorporating advanced segmentation and filtering [32, 31, 33].

The Multi-Scale Convolutional Neural Network (MSCNN) captures diverse scale information through multiple filters, complemented by attention mechanisms, as demonstrated in AFIPC [34, 26]. The FGA network integrates frequency domain and spatial information, enhancing crowd pattern recognition, while regional and semantic attention modules focus on relevant image areas for effective density estimation [29, 19]. DecideNet uses attention mechanisms for crowd density map estimation through detection and regression modules [1].

Feature fusion techniques enhance crowd counting models by integrating features from different modalities and scales, improving management of scale variations and occlusions. Semi-supervised approaches using binary segmentation tasks as surrogate targets exemplify innovative training methods for feature extractors [17].

Recent advancements emphasize the contributions of attention mechanisms and feature fusion in enhancing model robustness against occlusions and complex backgrounds. Models like the Multi-faceted Attention Network and FusionCount dynamically focus on critical instances and integrate multi-scale features for accurate crowd density estimation, achieving significant improvements in accuracy and reliability [35, 32, 13, 36, 37].

3.3 Integration of Convolutional and Transformer Architectures

Integrating convolutional and transformer architectures has emerged as a promising approach to enhance crowd counting performance by leveraging the strengths of both methodologies. CNNs excel at extracting local features and recognizing spatial patterns, while transformers capture long-range dependencies and global context, making them suitable for modeling complex interactions within images. This hybrid approach improves performance across various datasets by combining detailed local feature extraction with global contextual understanding [15, 38].

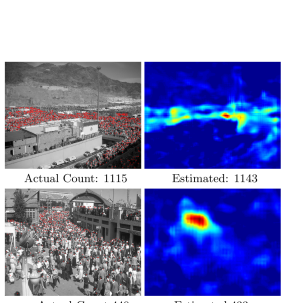
CCTrans exemplifies this integration by employing a pyramid vision transformer to capture global context, alongside a regression head utilizing multi-scale dilated convolutions for effective crowd density regression [22]. This underscores transformers' efficacy in managing complex spatial relationships within crowd scenes.

Research highlights the potential of transformer-based approaches to enhance performance, with context tokens improving feature representation and crowd counting accuracy [10, 39]. Hybrid models, like those combining density map estimation with semantic segmentation, leverage both local and global features to improve precision [24].

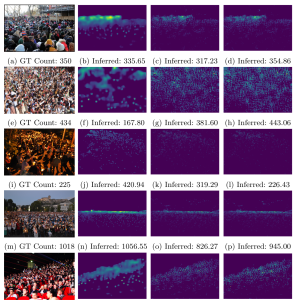
Innovations such as the Gradient Fusion approach demonstrate the potential of integrated architectures, allowing effective training without computationally expensive density maps during inference [25]. The Perspective-guided Fractional-Dilation Network (PFDNet) integrates perspective information with advanced architectures, dynamically adjusting convolutional receptive fields based on perspective to enhance accuracy [2].

These advancements highlight the significant potential of integrating convolutional and transformer architectures in crowd counting. By effectively combining CNNs' local feature extraction capabilities with transformers' global context modeling strengths, hybrid models offer enhanced accuracy and robustness in handling complex crowd scenes. Future research is expected to further integrate and refine innovative architectures, leading to significant enhancements in accuracy and efficiency through advanced methodologies and large-scale, diverse datasets [40, 6, 5, 15, 10].

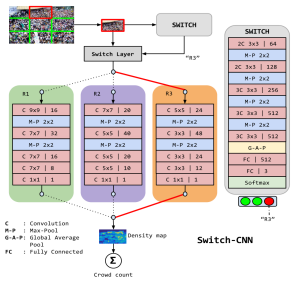
As shown in Figure 3, the integration of convolutional and transformer architectures has emerged as a powerful approach in crowd counting, offering enhanced accuracy and adaptability. The "Counting Crowds: A Comparison of Actual and Estimated Counts" illustrates the effectiveness of machine learning models in accurately estimating crowd sizes by comparing actual crowd counts with model-generated heat maps, where color intensity indicates density. The "Counting Crowd Size: A Comparative Study of Different Methods" highlights a comparative analysis of various crowd counting techniques, demonstrating variance in accuracy across methods applied to a dataset of crowd images with sizes ranging from 350 to 1018 people. Lastly, the "Switch-CNN: A Deep Learning Approach for Crowd Counting" delves into the architecture of a specialized deep learning model that utilizes convolutional layers and a unique Switch layer to dynamically adjust feature maps according



(a) Counting Crowds: A Comparison of Actual and Estimated Counts[23]



(b) Counting Crowd Size: A Comparative Study of Different Methods[20]



(c) Switch-CNN: A Deep Learning Approach for Crowd Counting[21]

Figure 3: Examples of Integration of Convolutional and Transformer Architectures

to crowd density. Together, these examples underscore the potential of integrating convolutional and transformer architectures to improve the precision and reliability of crowd counting applications [23, 20, 21].

Feature	Scale-Aware and Multi-Scale Architectures	Attention Mechanisms and Feature Fusion	Integration of Convolutional and Transformer Architectures
Scale Management	Multi-scale Modules	Attention Networks	Global Context
Architecture Type	Multi-column Networks	Multi-scale Cnn	Hybrid Models
Integration Strategy	Fusion Strategies	Feature Fusion	Local-global Integration

Table 2: This table provides a comparative analysis of key methodologies in crowd counting, focusing on scale-aware and multi-scale architectures, attention mechanisms, and the integration of convolutional and transformer architectures. It categorizes these methodologies by their scale management strategies, architecture types, and integration strategies, highlighting their unique approaches to addressing scale variation and feature integration challenges.

4 Density Estimation Approaches

4.1 Density Map Refinement Techniques

Refining density maps is crucial for enhancing crowd counting accuracy, particularly in complex environments. Techniques like the Deep Structured Scale Integration Network (DSSINet) improve feature representation by producing side output density maps at various layers, leading to more accurate crowd distribution estimates [41]. Models integrating CNNs with Transformers, such as JCTNet, optimize local and global feature capture while reducing model complexity compared to pure Transformers [38]. The Fourier-Guided Attention (FGA) network further refines density maps by integrating local and global features, enhancing pattern recognition in crowded scenes [29]. Similarly, DecideNet's adaptive estimation techniques achieve state-of-the-art results on challenging benchmarks [1].

The Multi-Scale Convolutional Neural Network (MSCNN) addresses perspective-induced distortions by learning scale-relevant density maps, thus improving accuracy [34]. Wide-area crowd counting methods, which align features from multiple camera views, further enhance accuracy by incorporating diverse perspectives [28]. Advanced refinement techniques are vital for improving algorithms, particularly in high-density distributions and varying crowd dynamics. Content-aware density map generation and adaptive learning frameworks bolster robustness, facilitating reliable applications in safety monitoring and event planning [42, 18, 33, 43, 44]. Integrating multi-scale features, attention mechanisms, and global context modeling enhances precision and reliability across diverse scenarios.

Figure 4 illustrates the hierarchical categorization of density map refinement techniques in crowd counting. It highlights three primary categories: Feature Integration, Adaptive Estimation, and Content-aware Generation. Each category encompasses specific methods that enhance crowd counting accuracy by improving feature representation, adapting to varying densities, and generating content-aware density maps. The figure further exemplifies various density estimation and refinement techniques, including the transition from raw images to density heatmaps, the analysis of crowd

density in a stadium through multiple perspectives, and a comparison of traditional and modern approaches to crowd counting methodologies, which address scale and occlusion challenges [42, 11, 45].

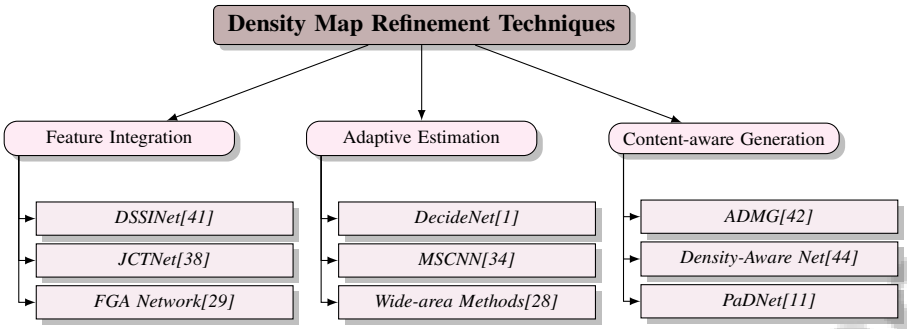


Figure 4: This figure illustrates the hierarchical categorization of density map refinement techniques in crowd counting. It highlights three primary categories: Feature Integration, Adaptive Estimation, and Content-aware Generation. Each category encompasses specific methods that enhance crowd counting accuracy by improving feature representation, adapting to varying densities, and generating content-aware density maps.

4.2 Challenges of Scale Variation

Scale variation poses a significant challenge in crowd density estimation, as individuals appear in different sizes due to perspective effects. This variability complicates tasks for traditional CNNs, which struggle with scale invariance across diverse scenes [46]. The long-tailed distribution of pixel values in dense regions exacerbates this issue, affecting density prediction accuracy [46].

Innovative methods such as FusionCount enhance accuracy by allowing pixel-wise scale selection [35]. CCTrans leverages transformers to capture local and global features, enabling accurate estimation despite scale variations [22]. Adaptive methods like the Top-Down Feedback Convolutional Neural Network (TDF-CNN) improve accuracy through feedback mechanisms [3], while the Perspective-guided Fractional-Dilation Network (PFDNet) utilizes fractional dilation rates to manage scale variation [2].

Pixel-wise soft gating nets and new relative local counting losses in multi-scale networks enhance density map optimization [47]. TEDnet's capability to produce high-quality maps with improved localization precision underscores the importance of advanced designs in addressing scale variation [48]. However, challenges persist, especially in errors for unlabeled data, impacting agent allocation accuracy [49]. Methods like L2SM centralize patches to similar density levels, enabling a single model to learn from diverse patterns [18].

Addressing scale variation is crucial for accurate density estimation, as different image regions exhibit varying crowd densities. CNNs excel in low-density areas but struggle in high-density regions, while transformers perform well in dense settings but face challenges in sparse areas. Advanced methods, including the CNN and Transformer Adaptive Selection Network (CTASNet), dynamically select suitable models for each density context. Techniques like content-aware mapping and scale aggregation networks enhance precision, benefiting applications in public safety and event planning [30, 50, 18, 33, 51]. By integrating multi-scale feature extraction, attention mechanisms, and adaptive learning, contemporary models have made significant strides in overcoming scale variation challenges, resulting in more reliable and precise crowd analysis.

5 Privacy-Preserving Crowd Analysis

5.1 Balancing Data Utility and Privacy

Balancing data utility and privacy in crowd analysis requires innovative strategies to maintain model efficacy while safeguarding individual privacy. The integration of high-level contextual information,

as demonstrated by the Top-Down Feedback Convolutional Neural Network (TDF-CNN), enhances crowd counting accuracy and addresses privacy concerns in densely populated settings [3].

Current datasets often contain biases that skew crowd analysis results [5]. Addressing these biases is crucial for developing privacy-preserving methods that generalize effectively across diverse scenarios. Incorporating multimodal data sources and varying environmental conditions in dataset creation aids in developing robust models that comply with privacy constraints.

Advanced methodologies, such as iterative distillation techniques, refine model performance by learning from aggregate data representations, maintaining high accuracy while protecting privacy. The Bound Tightening Network employs certification mechanisms to ensure reliable predictions without compromising privacy, while content-aware density maps and uncertainty estimation enhance accuracy and reduce reliance on extensive human annotation [52, 33, 53].

As illustrated in Figure 5, key strategies, considerations, and optimizations in balancing data utility and privacy for crowd analysis are highlighted, emphasizing innovative methods, addressing dataset biases, and optimizing model architectures. Deploying lightweight models for resource-constrained environments demonstrates the potential for real-time processing while preserving privacy, emphasizing the need for optimized architectures in privacy-preserving crowd analysis. Techniques like content-aware density maps are vital for safety, event planning, and consumer behavior analysis. Multi-modal models utilizing optical and thermal images show promise for improved predictions, despite increased complexity. Innovations such as the Pyramid Scale Network (PSNet) address scale limitations and feature similarity, paving the way for efficient crowd counting methodologies [54, 5, 33].

Achieving a balance between data utility and privacy necessitates integrating contextual information, managing dataset biases, and optimizing model architectures while considering background, camera angle, and human density. This comprehensive approach is essential for enhancing crowd counting algorithm accuracy and reliability, as evidenced by advancements in multimodal data utilization and techniques like conditional diffusion models, which improve density estimation and mitigate noise in training data [55, 40, 5, 33]. These strategies collectively contribute to effective and secure crowd monitoring solutions, addressing the dual imperatives of utility and privacy in modern applications.

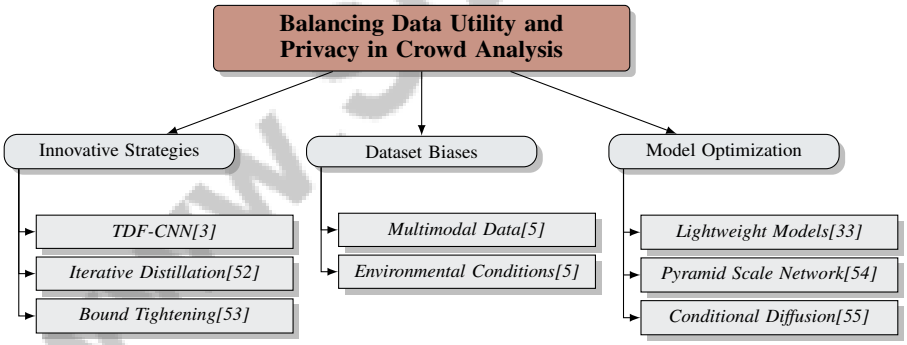


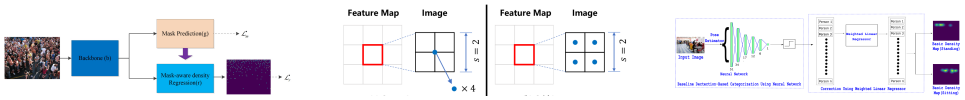
Figure 5: This figure illustrates the key strategies, considerations, and optimizations in balancing data utility and privacy for crowd analysis, highlighting innovative methods, addressing dataset biases, and optimizing model architectures.

5.2 Innovative Frameworks for Privacy Preservation

Innovative frameworks for privacy preservation in crowd analysis are crucial for addressing data security concerns while ensuring model effectiveness. A notable approach integrates secure multi-party computation with machine learning, enabling collaborative model training across parties without exposing sensitive data, thus preserving privacy and enhancing performance [56]. This method ensures data confidentiality, preventing unauthorized access and breaches.

Future research may enhance architectures like TEDnet, which shows promise in density estimation tasks. Incorporating attention mechanisms into TEDnet could improve performance in complex scenarios by focusing on relevant features and scale variations [57]. This would boost crowd counting accuracy and ensure effective integration of privacy-preserving techniques.

Recent research emphasizes balancing data utility—such as accurate crowd counting and density estimation for safety, urban planning, and consumer behavior analysis—with robust privacy preservation, addressing challenges like backdoor attacks and data limitations in existing datasets [11, 33, 58, 59]. By leveraging secure computation techniques and refining model architectures, researchers can develop robust crowd analysis solutions that meet privacy constraints while maintaining high accuracy and reliability across diverse applications.



(a) Mask-aware density regression for crowd counting[60]

(b) Feature Map and Image Comparison[61]

(c) Baseline Detection-Based Categorization Using Neural Network[62]

Figure 6: Examples of Innovative Frameworks for Privacy Preservation

As illustrated in Figure 6, innovative frameworks in privacy-preserving crowd analysis tackle the challenge of maintaining individual privacy while accurately analyzing crowd dynamics. The "Mask-aware density regression for crowd counting" utilizes a specialized flowchart to integrate mask predictions with a density regression network, enhancing accuracy while protecting individual identities. The "Feature Map and Image Comparison" analyzes two representations of the same image—a feature map in a 2x2 grid and a resized image in a 4x4 grid—highlighting how varying resolutions impact analysis. The "Baseline Detection-Based Categorization Using Neural Network" framework categorizes crowd densities based on standing and sitting poses, generating a corrected density map through input images processed by a pose estimator and neural network. These frameworks represent cutting-edge advancements in privacy-preserving crowd analysis, showcasing the potential to balance data utility with privacy concerns [60, 61, 62].

6 Occlusion Handling in Crowd Analysis

6.1 Techniques for Improving Accuracy in Occluded Environments

Addressing occlusion is crucial for precise crowd analysis, especially where individuals are obscured by others or environmental features. Recent advancements, such as attention models focusing on head locations, have effectively tackled challenges posed by complex backgrounds and scale variations. These models generate probability maps indicating likely head positions, enhancing crowd density estimation and mitigating annotation errors, thus improving predictions in densely populated settings [32, 50]. Innovative approaches like MTCNet employ multi-task learning to capture high-level scale information, aiding in discerning individuals in occluded environments [63]. Similarly, CrowdNet's integration of deep and shallow networks captures individuals at multiple scales, combining local features with broader contextual information [23].

Despite these advancements, challenges remain, particularly in dynamic scenes where rapid movement can violate conservation constraints [64]. Models must adapt to such changes to ensure accurate detection despite occlusions. Attention-based methods like CCCNet focus on relevant image regions to enhance detection, though they may struggle in extremely dense crowds [62]. While methods like HoMG are effective, they can falter with occlusions, leading to missed detections in crowded scenes [65]. This underscores the need for continuous innovation in occlusion handling techniques for reliable detection in highly occluded environments.

Advancements in techniques for occluded environments are vital for crowd analysis progression, given the complexities of varied backgrounds, scale variations, and non-uniform distributions. Recent innovations, such as attention models and Bound Tightening Networks, have shown promise in improving robustness and reducing errors in crowd counting. New datasets like JHU-CROWD, encompassing diverse conditions, highlight the necessity for sophisticated methods to effectively analyze crowd dynamics in real-world scenarios [32, 14, 6, 53]. By integrating multi-task learning, attention mechanisms, and hybrid architectures, contemporary models can significantly enhance performance in challenging scenarios, ensuring accurate and reliable crowd counting across diverse environments.

6.2 Leveraging Multi-Modal and Multi-View Information

Utilizing multi-modal and multi-view data offers significant opportunities for improving occlusion handling in crowd analysis by providing diverse perspectives and complementary information that enhance crowd counting and density estimation precision. Research indicates that integrating optical and thermal images results in superior predictive performance, although the mechanisms through which multimodal models extract enriched features remain partially understood. Multi-view approaches expand the field of view beyond a single camera, effectively capturing occluded individuals and improving counting accuracy. Recent advancements, such as the cross-view cross-scene multi-view paradigm, enable models to adapt to varying scenes and camera configurations, addressing challenges associated with non-correspondence errors. Attention mechanisms focused on head locations further enhance robustness against complex backgrounds and scale variations, improving outcomes in crowd analysis applications [12, 32, 5, 66, 28].

Multi-modal data, encompassing sensory inputs like RGB images, depth maps, and thermal images, significantly improve model robustness by providing diverse information that compensates for single data source limitations. Multi-view techniques integrate data from multiple camera angles to generate a comprehensive understanding of the crowd scene, reducing occlusion challenges; individuals hidden in one view may be visible in another. This approach employs techniques such as perspective-aware convolutional networks and cross-view information fusion to adaptively combine insights from various angles, ensuring a holistic scene understanding and mitigating perspective distortion and background noise [14, 67, 13, 66].

The W-Net architecture exemplifies multi-modal approaches' potential, maintaining high structural similarity in density maps while achieving faster convergence, crucial for addressing occlusion challenges [68]. However, methods like the Weighted VLAD approach may struggle in scenarios with severe occlusions or rapidly changing backgrounds, affecting crowd estimate accuracy [69].

Despite multi-modal and multi-view techniques' advantages, challenges remain, particularly in extremely sparse crowd conditions or significant occlusions. For instance, the SOFA-Net method, while effective, may face difficulties under such conditions [70]. Similarly, single-column networks like SCNet may encounter challenges in extremely dense scenes where occlusions are prevalent [71]. DecideNet, combining detection and regression, may also struggle in dense scenes where occlusion is pronounced, potentially leading to inaccurate counts if detection confidence is low [1].

The integration of multi-modal and multi-view data represents a significant advancement in occlusion handling for crowd analysis. By incorporating perspective information into the crowd density regression process, these innovative approaches enhance crowd counting and density estimation accuracy and reliability, particularly in challenging situations marked by substantial occlusions and perspective distortions. The proposed perspective-aware convolutional neural network (PACNN) addresses difficulties associated with varying person scales in images and employs multi-scale perspective maps to enhance density predictions' robustness, yielding more reliable results in complex crowd scenarios [72, 67].

7 Real-Time Crowd Monitoring

7.1 Trade-offs Between Speed and Accuracy

Achieving a balance between speed and accuracy in real-time crowd monitoring is a critical challenge, necessitating improvements in computational efficiency without sacrificing precision. This challenge arises from the need to analyze dynamic environments where occlusions, scale variations, and complex backgrounds impact the accuracy of crowd density estimations. Recent advancements in deep learning, such as attention models and uncertainty estimation, have facilitated better localization and density predictions, enhancing situational awareness while reducing reliance on extensive human annotation for model training in new domains [6, 73, 52, 32, 10]. Developing lightweight models that operate at high frame rates is essential for real-time applications, especially in environments demanding quick decision-making.

Density-based clustering methods, as illustrated by [74], demonstrate effective real-time crowd flow detection by reducing computational overhead, thus enabling rapid processing without the need

for resource-intensive tracking techniques. This efficiency is crucial for maintaining high-speed operations while ensuring accurate crowd flow analysis.

The Scene Invariant Crowd Segmentation method achieves an average processing time of 18 ms per frame, allowing real-time analysis at over 50 frames per second [65]. This performance highlights the importance of optimizing algorithms to achieve both speed and accuracy, enabling systems to swiftly respond to changes in crowd dynamics.

However, these advancements involve inherent trade-offs. Prioritizing speed may reduce analysis granularity or the ability to accurately handle complex scenarios. Rapid processing methods may struggle with nuanced crowd behaviors or environments with significant occlusions, leading to high error rates in background regions—accounting for 18-49

The trade-offs between speed and accuracy in real-time crowd monitoring require innovative algorithmic solutions that harmonize computational efficiency with precise analysis. By leveraging sophisticated techniques and refining model architectures, researchers can develop advanced crowd counting systems that meet the stringent demands of real-time applications while ensuring timely responses and precise insights into crowd dynamics. This includes the use of multimodal data, such as optical and thermal images, to enhance predictive capabilities, although the complexities introduced by these multimodal approaches—such as increased inference time and memory requirements—necessitate a thorough examination of their advantages over traditional monomodal models. Current research underscores the need for effective dataset criteria to clarify performance differences between these model types, ultimately guiding future advancements in automated crowd counting for improved situational awareness in public spaces [5, 10].

7.2 Algorithmic Approaches and Innovations

Innovative algorithmic approaches are vital for enhancing real-time crowd monitoring systems, enabling efficient data processing while maintaining high accuracy. Hierarchical Pyramid Mapping (HPM), with its user-friendly graphical interface, significantly boosts real-time monitoring capabilities by facilitating seamless interaction and rapid data interpretation [75].

The development of advanced loss functions, such as Bayesian Loss, further refines the precision of crowd counting models by addressing uncertainties in crowd density estimation [75]. This approach enhances the robustness of monitoring systems, ensuring accurate estimates even in diverse scenarios with varying densities and occlusions.

The integration of sophisticated deep learning techniques, particularly convolutional neural networks (CNNs) and transformer architectures, has significantly transformed real-time crowd monitoring by improving accuracy and efficiency in crowd counting and density estimation. This evolution facilitates the combination of local and non-local features, leading to substantial performance improvements across challenging scenarios, including occlusions and varying crowd densities. Additionally, the development of compact models capable of near real-time speeds has expanded the applicability of these techniques in practical situations such as video surveillance and public safety [9, 76, 15]. These hybrid models excel in extracting both local and global features, adapting to dynamic environments and rapidly changing crowd dynamics. By capturing intricate spatial patterns and long-range dependencies, these models provide comprehensive insights into crowd behavior, enhancing system responsiveness and accuracy.

These algorithmic innovations underscore the importance of ongoing advancements in real-time crowd monitoring. By incorporating intuitive user interfaces, sophisticated loss functions, and cutting-edge deep learning methodologies, modern crowd analysis systems are designed to deliver accurate and reliable insights into crowd dynamics. This capability is crucial for real-time applications in complex environments, such as large public gatherings, where effective crowd counting and density estimation can significantly enhance safety and management. Recent advancements in deep learning have notably improved system performance, addressing challenges related to model architecture, computational efficiency, and accuracy in crowd localization. Innovative approaches like Composition Loss have been developed to simultaneously tackle counting, density map estimation, and individual localization, leveraging extensive datasets such as the UCF-QNRF, which includes over 1.25 million annotated individuals in diverse settings [77, 10].

8 Applications of Crowd Counting and Density Estimation

8.1 Public Safety and Surveillance

Crowd counting and density estimation are integral to enhancing public safety and surveillance by providing crucial insights into crowd dynamics, thereby facilitating effective management of densely populated areas. Accurate crowd size and distribution assessments are essential to prevent overcrowding and ensure safety in public spaces. The UCF-QNRF dataset, noted for its diversity and high-resolution images, supports the development of robust models for complex scenarios [77]. In environments with inhomogeneous density distributions, the Density-Aware Network excels by addressing these complexities, which is vital for public safety applications to avert hazardous situations such as stampedes or bottlenecks [44].

Lightweight architectures like DRResNet are optimized for resource-limited environments, ensuring high performance in real-time surveillance systems [78]. These systems enable continuous monitoring and rapid responses to safety threats. The ZoomCount mechanism demonstrates superior accuracy across datasets, enhancing surveillance by providing reliable data for informed decision-making [79]. Integrating advanced techniques such as the Pan-Density Network (PaDNet) and Convolutional Neural Networks (CNNs) improves monitoring across varying densities by addressing challenges like perspective distortions and occlusions. These systems employ sophisticated algorithms to capture both global and local contextual features, contributing to incident prevention and enhancing urban planning, emergency response, and social security [45, 80, 9, 58, 81].

8.2 Event Management

In event management, crowd counting and density estimation are crucial for effective crowd control and enhancing attendee safety and experience. These techniques provide real-time insights into crowd dynamics, allowing proactive management of crowd flow and density. Machine learning models incorporating content-aware density mapping and attention mechanisms focused on head locations offer precise data on crowd size, density variations, and movement patterns, essential for optimizing event layouts and safety protocols [32, 49, 33, 58].

Advanced algorithms like Hierarchical Pyramid Mapping (HPM) and Bayesian Loss approaches enhance the precision of crowd counting systems in event settings [75]. These methods facilitate accurate crowd size and distribution estimations, optimizing resource allocation. The integration of deep learning techniques, including CNNs and transformer architectures, enables comprehensive spatial and temporal analyses of crowd behavior. These methodologies improve accuracy by combining local and non-local features to address challenges such as inter-occlusion and varying densities. Models like the Switching Convolutional Neural Network and CCTrans enhance predictions for civic agencies and planners during large gatherings [82, 15, 83].

Real-time monitoring systems with these advanced techniques enable rapid detection of potential issues, allowing swift intervention to mitigate risks. This proactive approach enhances attendee safety through accurate monitoring and management of crowd density, facilitating smooth movement throughout event venues and reducing congestion and hazards [84, 85, 32, 33, 14]. Integrating these technologies in event management is essential for enhancing safety and order during large gatherings, providing accurate insights into crowd size, density variations, and flow dynamics, crucial for effective safety measures, emergency response planning, and optimizing venue design [33, 58].

8.3 Urban Planning and Resource Allocation

Crowd analysis techniques, particularly crowd counting and density estimation, significantly impact urban planning and resource allocation strategies. These methodologies provide urban planners with insights into population distribution and movement patterns, enabling precise decision-making in urban space design and management, enhancing safety and optimizing resource allocation [6, 45, 33, 80, 81]. Advanced crowd counting models, especially those utilizing CNNs and transformer architectures, offer enhanced accuracy and scalability for large-scale urban analysis. These models facilitate real-time monitoring of pedestrian flow and congestion, allowing dynamic adjustments to urban infrastructure and services to meet fluctuating demands. Implementing density-aware networks in urban settings significantly improves population estimates, providing planners with reliable data to guide public amenities and transportation system development [44].

Integrating crowd analysis into urban planning supports intelligent transportation systems that adapt to varying traffic conditions and pedestrian volumes. Real-time crowd movement data enhances urban mobility through dynamic traffic signal optimization, efficient public transit rerouting, and effective public space management, alleviating congestion and improving accessibility [86, 6, 85, 81, 10]. Crowd analysis techniques are also essential for effective emergency management and disaster response in urban environments, enabling accurate monitoring and assessment of crowd dynamics crucial for public safety during large gatherings and unexpected events. Advanced methods, such as Depth Information Guided Crowd Counting and Confidence Guided Deep Residual Counting Networks, enhance precision in crowd density estimation by addressing challenges like occlusions and non-uniform distributions, supporting intelligent urban safety systems [6, 87, 33, 55, 15]. Providing accurate crowd density and movement estimates enables effective resource deployment during emergencies, ensuring timely and coordinated responses to incidents threatening public safety, particularly in densely populated areas.

Integrating advanced crowd analysis techniques such as Confidence Guided Deep Residual Counting Networks and semi-supervised learning frameworks signifies a transformative shift in urban management, enhancing public safety and optimizing infrastructure usage in response to rapid urbanization and increasing crowd gatherings [6, 85]. By delivering detailed insights into population dynamics and infrastructure needs, these techniques empower urban planners to create resilient, efficient, and sustainable urban environments that meet their populations' diverse needs.

9 Conclusion

9.1 Addressing Limitations and Future Directions

Despite notable progress in crowd analysis, several challenges remain that need to be addressed to improve the robustness and adaptability of models in diverse environments. Current techniques, like MSCNN, often falter in extremely dense crowds due to prevalent occlusions, highlighting the need for optimized architectures and hybrid models that effectively combine detection and regression approaches. Future research should focus on refining gating mechanisms, incorporating additional loss functions, and extending these methodologies to domains beyond crowd counting, as seen in the redesign of multi-scale neural networks.

Key areas for future exploration include enhancing model adaptability to varying crowd densities and developing lightweight architectures for efficient crowd counting. Integrating contextual information into feature fusion processes is crucial for improving performance amidst scale variations. Additionally, advancing the capabilities of models like TEDnet in challenging settings and exploring real-time applications in dynamic environments are promising research directions.

The resilience of models such as the Scale-Aware Crowd Counting Network (SACCN) in extreme crowd conditions and the exploration of advanced attention mechanisms are vital research avenues. The limitations of methods like the Top-Down Feedback Convolutional Neural Network (TDF-CNN) in handling extreme occlusions or atypical crowd patterns underscore the necessity for improved feedback mechanisms and greater adaptability to diverse scenarios.

Further research should also aim to enhance self-training processes and investigate alternative surrogate tasks to boost performance in semi-supervised crowd counting models. Strengthening perspective estimation techniques, such as those used in the Perspective-guided Fractional-Dilation Network (PFDNet), and expanding their applications to other domains are advisable.

Developing new datasets that meet specific criteria is crucial for a deeper understanding of multimodal crowd counting. Improving the robustness of detection components in highly congested environments and refining attention mechanisms for better performance are promising directions. Additionally, reducing the supervision requirements in models like Point-in-Box-Out and exploring alternative weak annotations could significantly advance the field.

Addressing these limitations and exploring new research paths are essential for the advancement of crowd analysis techniques. By optimizing existing models, developing innovative learning approaches, and broadening applications to wider contexts, future research can greatly enhance the effectiveness and reliability of crowd counting and density estimation in real-world scenarios.

References

[1] Jiang Liu, Chenqiang Gao, Deyu Meng, and Alexander G. Hauptmann. Decidenet: Counting varying density crowds through attention guided detection and density estimation, 2018.

[2] Zhaoyi Yan, Ruimao Zhang, Hongzhi Zhang, Qingfu Zhang, and Wangmeng Zuo. Crowd counting via perspective-guided fractional-dilation convolution, 2021.

[3] Deepak Babu Sam and R. Venkatesh Babu. Top-down feedback for crowd counting convolutional neural network, 2018.

[4] Yuting Liu, Miaojing Shi, Qijun Zhao, and Xiaofang Wang. Point in, box out: Beyond counting persons in crowds, 2019.

[5] Martin Thißen and Elke Hergenröther. Why existing multimodal crowd counting datasets can lead to unfulfilled expectations in real-world applications, 2023.

[6] Vishwanath A Sindagi, Rajeev Yasarla, and Vishal M Patel. Pushing the frontiers of unconstrained crowd counting: New dataset and benchmark method. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1221–1231, 2019.

[7] Ankan Bansal and K. S. Venkatesh. People counting in high density crowds from still images, 2015.

[8] Viresh Ranjan, Hieu Le, and Minh Hoai. Iterative crowd counting. In *Proceedings of the European conference on computer vision (ECCV)*, pages 270–285, 2018.

[9] Vishwanath A. Sindagi and Vishal M. Patel. A survey of recent advances in cnn-based single image crowd counting and density estimation, 2017.

[10] Muhammad Asif Khan, Hamid Menouar, and Ridha Hamila. Revisiting crowd counting: State-of-the-art, trends, and future perspectives, 2022.

[11] Yukun Tian, Yiming Lei, Junping Zhang, and James Z. Wang. Padnet: Pan-density crowd counting, 2020.

[12] Haoliang Meng, Xiaopeng Hong, Chenhao Wang, Miao Shang, and Wangmeng Zuo. Multimodal crowd counting via a broker modality, 2024.

[13] Geng Chen and Peirong Guo. Enhanced information fusion network for crowd counting, 2021.

[14] Davide Modolo, Bing Shuai, Rahul Rama Varior, and Joseph Tighe. Understanding the impact of mistakes on background regions in crowd counting, 2020.

[15] Viresh Ranjan, Mubarak Shah, and Minh Hoai Nguyen. Crowd transformer network, 2019.

[16] Yi Wang, Junhui Hou, Xinyu Hou, and Lap-Pui Chau. A self-training approach for point-supervised object detection and counting in crowds, 2021.

[17] Yan Liu, Lingqiao Liu, Peng Wang, Pingping Zhang, and Yinjie Lei. Semi-supervised crowd counting via self-training on surrogate tasks, 2020.

[18] Chenfeng Xu, Kai Qiu, Jianlong Fu, Song Bai, Yongchao Xu, and Xiang Bai. Learn to scale: Generating multipolar normalized density maps for crowd counting, 2019.

[19] Qiaosi Yi, Yunxing Liu, Aiwen Jiang, Juncheng Li, Kangfu Mei, and Mingwen Wang. Scale-aware network with regional and semantic attentions for crowd counting under cluttered background, 2021.

[20] Javier Antonio Gonzalez-Trejo and Diego Alberto Mercado-Ravell. Dense crowds detection and counting with a lightweight architecture, 2020.

[21] Deepak Babu Sam, Shiv Surya, and R. Venkatesh Babu. Switching convolutional neural network for crowd counting, 2017.

[22] Ye Tian, Xiangxiang Chu, and Hongpeng Wang. Cctrans: Simplifying and improving crowd counting with transformer, 2021.

[23] Lokesh Boominathan, Srinivas S S Kruthiventi, and R. Venkatesh Babu. Crowdnet: A deep convolutional network for dense crowd counting, 2016.

[24] Jia Wan, Nikil Senthil Kumar, and Antoni B. Chan. Fine-grained crowd counting, 2020.

[25] Zhuojun Chen, Junhao Cheng, Yuchen Yuan, Dongping Liao, Yizhou Li, and Jiancheng Lv. Deep density-aware count regressor, 2020.

[26] Di Kang and Antoni Chan. Crowd counting by adaptively fusing predictions from an image pyramid, 2018.

[27] Jiwei Chen and Zengfu Wang. Crowd counting with segmentation attention convolutional neural network, 2022.

[28] Qi Zhang and Antoni B. Chan. Wide-area crowd counting: Multi-view fusion networks for counting in large scenes, 2022.

[29] Yashwardhan Chaudhuri, Ankit Kumar, Arun Balaji Buduru, and Adel Alshamrani. Fga: Fourier-guided attention network for crowd count estimation, 2024.

[30] Xinkun Cao, Zhipeng Wang, Yanyun Zhao, and Fei Su. Scale aggregation network for accurate and efficient crowd counting. In *Proceedings of the European conference on computer vision (ECCV)*, pages 734–750, 2018.

[31] Mohammad Asiful Hossain, Mehrdad Hosseinzadeh, Omit Chanda, and Yang Wang. Crowd counting using scale-aware attention networks, 2019.

[32] Youmei Zhang, Chunluan Zhou, Faliang Chang, and Alex C. Kot. Attention to head locations for crowd counting, 2018.

[33] Mahdi Maktabdar Oghaz, Anish R Khadka, Vasileios Argyriou, and Paolo Remagnino. Content-aware density map for crowd counting and density estimation, 2019.

[34] Lingke Zeng, Xiangmin Xu, Bolun Cai, Suo Qiu, and Tong Zhang. Multi-scale convolutional neural networks for crowd counting. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 465–469. IEEE, 2017.

[35] Yiming Ma, Victor Sanchez, and Tanaya Guha. Fusioncount: Efficient crowd counting via multiscale feature fusion, 2022.

[36] Usman Sajid and Guanghui Wang. Towards more effective prm-based crowd counting via a multi-resolution fusion and attention network, 2021.

[37] Hui Lin, Zhiheng Ma, Rongrong Ji, Yaowei Wang, and Xiaopeng Hong. Boosting crowd counting via multifaceted attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19628–19637, 2022.

[38] Fusen Wang, Kai Liu, Fei Long, Nong Sang, Xiaofeng Xia, and Jun Sang. Joint cnn and transformer network via weakly supervised learning for efficient crowd counting, 2022.

[39] Guolei Sun, Yun Liu, Thomas Probst, Danda Pani Paudel, Nikola Popovic, and Luc Van Gool. Rethinking global context in crowd counting, 2023.

[40] Yi Hou, Chengyang Li, Yuheng Lu, Liping Zhu, Yuan Li, Huizhu Jia, and Xiaodong Xie. Enhancing and dissecting crowd counting by synthetic data, 2022.

[41] Lingbo Liu, Zhilin Qiu, Guanbin Li, Shufan Liu, Wanli Ouyang, and Liang Lin. Crowd counting with deep structured scale integration network, 2019.

[42] Jia Wan and Antoni Chan. Adaptive density map generation for crowd counting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1130–1139, 2019.

[43] Jia Wan and Antoni Chan. Modeling noisy annotations for crowd counting. *Advances in Neural Information Processing Systems*, 33:3386–3396, 2020.

[44] Hanhui Li, Xiangjian He, Hefeng Wu, Saeed Amirgholipour Kasmani, Ruomei Wang, Xiaonan Luo, and Liang Lin. Structured inhomogeneous density map learning for crowd counting, 2018.

[45] Guangshuai Gao, Junyu Gao, Qingjie Liu, Qi Wang, and Yunhong Wang. Cnn-based density estimation and crowd counting: A survey, 2020.

[46] Chenfeng Xu, Dingkang Liang, Yongchao Xu, Song Bai, Wei Zhan, Xiang Bai, and Masayoshi Tomizuka. Autoscale: Learning to scale for crowd counting and localization, 2021.

[47] Zhipeng Du, Miaojing Shi, Jiankang Deng, and Stefanos Zafeiriou. Redesigning multi-scale neural network for crowd counting, 2023.

[48] Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, and Ling Shao. Crowd counting and density estimation by trellis encoder-decoder network, 2019.

[49] Hui Lin, Zhiheng Ma, Xiaopeng Hong, Yaowei Wang, and Zhou Su. Semi-supervised crowd counting via density agency, 2022.

[50] Muhammad Asif Khan, Hamid Menouar, and Ridha Hamila. Crowd density estimation using imperfect labels, 2023.

[51] Yuehai Chen, Jing Yang, Badong Chen, and Shaoyi Du. Counting varying density crowds through density guided adaptive selection cnn and transformer estimation, 2022.

[52] Viresh Ranjan, Boyu Wang, Mubarak Shah, and Minh Hoai. Uncertainty estimation and sample selection for crowd counting, 2020.

[53] Qiming Wu. Bound tightening network for robust crowd counting, 2024.

[54] Junhao Cheng, Zhuojun Chen, XinYu Zhang, Yizhou Li, and Xiaoyuan Jing. Exploit the potential of multi-column architecture for crowd counting, 2020.

[55] Yasiru Ranasinghe, Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, and Vishal M. Patel. $crowddiff$: Multi-hypothesis crowd density estimation using diffusion models, 2024.

[56] Weizhe Liu, Mathieu Salzmann, and Pascal Fua. Context-aware crowd counting, 2019.

[57] Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, and Ling Shao. Crowd counting and density estimation by trellis encoder-decoder networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6133–6142, 2019.

[58] Yukun Tian, Yiming Lei, Junping Zhang, and James Z Wang. Padnet: Pan-density crowd counting. *IEEE Transactions on Image Processing*, 29:2714–2727, 2019.

[59] Yuhua Sun, Tailai Zhang, Xingjun Ma, Pan Zhou, Jian Lou, Zichuan Xu, Xing Di, Yu Cheng, and Lichao. Backdoor attacks on crowd counting, 2022.

[60] Shengqin Jiang, Xiaobo Lu, Yinjie Lei, and Lingqiao Liu. Mask-aware networks for crowd counting, 2019.

[61] Qingyu Song, Changan Wang, Zhengkai Jiang, Yabiao Wang, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Yang Wu. Rethinking counting and localization in crowds:a purely point-based framework, 2021.

[62] Sarkar Snigdha Sarathi Das, Syed Md. Mukit Rashid, and Mohammed Eunus Ali. Cccnet: An attention based deep learning framework for categorized crowd counting, 2019.

[63] Abhay Kumar, Nishant Jain, Suraj Tripathi, Chirag Singh, and Kamal Krishna. Mtcnet: Multi-task learning paradigm for crowd count estimation, 2019.

[64] Weizhe Liu, Mathieu Salzmann, and Pascal Fua. Counting people by estimating people flows, 2021.

[65] Parthipan Siva, Mohammad Javad Shafiee, Mike Jamieson, and Alexander Wong. Scene invariant crowd segmentation and counting using scale-normalized histogram of moving gradients (homg), 2016.

[66] Qi Zhang, Wei Lin, and Antoni B. Chan. Cross-view cross-scene multi-view crowd counting, 2022.

[67] Miaojing Shi, Zhaohui Yang, Chao Xu, and Qijun Chen. Revisiting perspective information for efficient crowd counting, 2019.

[68] Varun Kannadi Valloli and Kinal Mehta. W-net: Reinforced u-net for density map estimation, 2019.

[69] Biyun Sheng, Chunhua Shen, Guosheng Lin, Jun Li, Wankou Yang, and Changyin Sun. Crowd counting via weighted vlad on dense attribute feature maps, 2016.

[70] Haoran Duan, Shidong Wang, and Yu Guan. Sofa-net: Second-order and first-order attention network for crowd counting, 2020.

[71] Ze Wang, Zehao Xiao, Kai Xie, Qiang Qiu, Xiantong Zhen, and Xianbin Cao. In defense of single-column networks for crowd counting, 2018.

[72] Miaojing Shi, Zhaohui Yang, Chao Xu, and Qijun Chen. Revisiting perspective information for efficient crowd counting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7279–7288, 2019.

[73] Deepak Babu Sam, Skand Vishwanath Peri, Mukuntha Narayanan Sundararaman, Amogh Kamath, and R. Venkatesh Babu. Locate, size and count: Accurately resolving people in dense crowds via detection, 2020.

[74] Giovanna Castellano, Eugenio Cotardo, Corrado Mencar, and Gennaro Vessio. Density-based clustering with fully-convolutional networks for crowd flow detection from drones, 2023.

[75] Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Bayesian loss for crowd count estimation with point supervision, 2019.

[76] Xiaowen Shi, Xin Li, Caili Wu, Shuchen Kong, Jing Yang, and Liang He. A real-time deep network for crowd counting, 2020.

[77] Haroon Idrees, Muhmmad Tayyab, Kishan Athrey, Dong Zhang, Somaya Al-Maadeed, Nasir Rajpoot, and Mubarak Shah. Composition loss for counting, density map estimation and localization in dense crowds, 2018.

[78] Xinghao Ding, Zhirui Lin, Fujin He, Yu Wang, and Yue Huang. A deeply-recursive convolutional network for crowd counting, 2018.

[79] Usman Sajid, Hasan Sajid, Hongcheng Wang, and Guanghui Wang. Zoomcount: A zooming mechanism for crowd counting in static images, 2020.

[80] Guangshuai Gao, Junyu Gao, Qingjie Liu, Qi Wang, and Yunhong Wang. Cnn-based density estimation and crowd counting: A survey. *arXiv preprint arXiv:2003.12783*, 2020.

[81] Eric K. Tokuda, Yitzchak Lockerman, Gabriel B. A. Ferreira, Ethan Sorrelgreen, David Boyle, Roberto M. Cesar-Jr., and Claudio T. Silva. A new approach for pedestrian density estimation using moving sensors and computer vision, 2020.

[82] Deepak Babu Sam, Shiv Surya, and R Venkatesh Babu. Switching convolutional neural network for crowd counting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5744–5752, 2017.

[83] Ye Tian, Xiangxiang Chu, and Hongpeng Wang. Cctrans: Simplifying and improving crowd counting with transformer. *arXiv preprint arXiv:2109.14483*, 2021.

[84] Zhen Zhao, Miaojing Shi, Xiaoxiao Zhao, and Li Li. Active crowd counting with limited supervision, 2020.

[85] Haoran Duan, Fan Wan, Rui Sun, Zeyu Wang, Varun Ojha, Yu Guan, Hubert P. H. Shum, Bingzhang Hu, and Yang Long. Semi-supervised crowd counting from unlabeled data, 2024.

[86] Qi Wang, Junyu Gao, Wei Lin, and Yuan Yuan. Pixel-wise crowd understanding via synthetic data, 2020.

[87] Mingliang Xu, Zhaoyang Ge, Xiaoheng Jiang, Gaoge Cui, Pei Lv, Bing Zhou, and Changsheng Xu. Depth information guided crowd counting for complex crowd scenes. *Pattern Recognition Letters*, 125:563–569, 2019.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.