Crowd Counting via Lightweight Neural Networks: A Literature Review

Jing-an Cheng[1†], Wenzhe Zhai[1†], Qilei Li[2,3], Mingliang Gao[1*]

1 School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China.
2 Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, 430079, China.
3 National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan, 430079, China.

E-mail: mlgao@sdut.edu.cn (*Corresponding author: Mingliang Gao)

[†]Jing-an Cheng and Wenzhe Zhai contributed equally.

Abstract. Crowd counting refers to the task of estimating the number of people within a specific area, which provides insights into crowd dynamics and distribution. This task has widespread applications across various domains, including public safety, urban planning, and traffic management. Recent advances in deep learning, particularly the convolutional neural networks and the Transformer architecture, as well as multimodal pre-trained models, such as CLIP and SAM, have substantially improved counting accuracy. However, existing models are mainly designed in high-complexity architecture, which incur heavy computational and data requirements, and hinder real-time deployment and increase the cost. Consequently, lightweight crowd-counting approaches have emerged as a key research direction, which aims to reduce parameters and inference time while retaining high performance. In this survey, we summarize publicly available datasets, evaluation metrics, latest lightweight architectures, and evaluate representative models on benchmark datasets to inspire future research in crowd counting.

Keywords: Lightweight network, Crowd counting, Literature review, Density estimation, Performance evaluation.

1. Introduction

Crowd counting is a key research area within computer vision. It aims at estimating and analyzing crowd density and distribution across designated spaces [1, 2, 3, 4, 5]. This research has widespread application in public safety, urban planning, and traffic management fields [6, 7, 8, 9, 10]. For example, to address crowd counting under public security demands, Guo *et al.* [11] proposed a scale region recognition network. It is tailored to identify human bodies of varying scales and incorporates scale-level

awareness and object region recognition modules to distinguish body regions and reduce distractions from background [5]. Chen *et al.* [10] introduced a frequency pyramid-based model aimed at object counting for urban development and traffic systems. A pyramid attention and hybrid feature pyramid modules were proposed to enhance detection precision and reduce the effects of background complexity [10].

Early crowd counting methods relied on classical computer vision techniques, which can be categorized into detection-based approaches [12, 13, 14, 15, 16, 17] and regression-based approaches [18, 19, 20]. Detection-based methods detect the individual by scanning the image with sliding windows and applying classifiers to check for the presence of human targets. For example, the Viola-Jones detector [12] uses Haar features and a cascade of AdaBoost classifiers for fast face detection. The Deformable Part Models [15] improve robustness to pose changes and occlusions using deformable parts. These methods perform well in sparse scenes. However, they often fail in dense crowds due to frequent occlusions and overlapping boxes. In contrast, regression-based approaches skip individual detection. They directly map an image or image patch to a count or density value. Early studies [18, 19] extracted hand-crafted features, such as HOG [20], and used linear or support vector regression to estimate numbers. These methods handle occlusion better in dense scenes, but their accuracy is limited due to the hand-crafted features. Overall, the detection-based and regression-based methods often struggled with accuracy in high-density, complex scenes [21]. The rapid progress of deep learning has since introduced convolutional neural networks (CNNs) as the leading approach for crowd counting, with CNNs proving highly effective at feature extraction, particularly in dense and complex crowd scenes [22, 23, 24, 25, 26, 27].

While deep networks like CNNs have advanced significantly in performance, they typically require extensive layers and parameters, resulting in high computational costs [28, 29, 30, 31]. High-complexity models [32, 33, 34, 35, 11] achieve high accuracy but impose significant computational and memory overhead, which restricts their application in real-world settings, especially in resource-limited environments. With the growth of IoT (Internet of Things) and edge computing, this issue has become even more pressing [36, 37, 38, 39, 40]. Consequently, reducing computational demands while maintaining model accuracy has become a pivotal challenge in crowd counting research [41, 42]. To address this challenge, researchers have increasingly focused on designing lightweight networks [43, 44, 45, 46]. These networks aim to decrease the number of parameters and computational demands to achieve greater efficiency and resource utilization while retaining model performance. Current lightweight methods can be categorized into three main types.

(i) **Lightweight Architecture-based Networks** [47, 48, 49, 50, 51]. These networks aim to build efficient neural architectures that lower both parameter counts and computational overhead. By refining the network structures, including layer reduction and the use of lightweight convolutions, these networks achieve a balance between performance and computational efficiency.

(ii) **Lightweight Module-based Networks** [52, 53, 54, 46, 55]. These networks build on existing models by incorporating specific modules, like lightweight attention mechanisms or multi-scale feature fusion, to improve feature extraction and adaptability. This strategy optimizes computational resources and enhances performance in complex, high-density crowd scenarios.

(iii) **Knowledge Distillation-based Networks** [56, 57, 58, 59, 60]. These networks transfer insights from a large teacher model to a compact student model. This approach enables the student model to achieve high accuracy while significantly reducing parameters and computational demands. During distillation, soft labels and feature representations from the teacher model serve as learning samples to guide the student model to better understand complex data distributions. Knowledge distillation is particularly useful for crowd counting in complex environments. It allows lightweight models to attain high counting accuracy and be suitable for deployment in resource-limited environments.

The three lightweight methods show strong adaptability and potential for use in resource-limited environments. This paper systematically reviews these methods and provides a detailed analysis of their features and suitability to support lightweight network research in crowd counting. The main contributions of this paper are as follows.

(i) This paper reviews mainstream lightweight crowd counting methods, which are categorized into three main types to provide researchers with a detailed technical framework.

(ii) This paper analyzes various lightweight models by comparing their performance in accuracy, efficiency, and resource utilization to provide practical insights for application.

(iii) This paper identifies the development potential of lightweight crowd counting methods based on current limitations and provides a clear direction for future research.

The structure of this paper is organized as follows. Section 1 provides an overview of the research background and current progress. Section 2 discusses basic knowledge of lightweight crowd counting with an overview of essential datasets and evaluation metrics that ground the later analysis. Section 3 reviews recent mainstream lightweight methods in crowd counting, with a focus on network structure optimization, lightweight module integration, and knowledge distillation. Section 4 presents a comparative analysis of mainstream lightweight models on common datasets and evaluates their accuracy and efficiency. Section 5 discusses existing challenges and future directions in lightweight crowd counting technologies to offer guidance for future research. Section 6 concludes the paper with a comprehensive summary.

2. Crowd Counting Datasets and Evaluation Protocols

2.1. Datasets

The field of crowd counting has witnessed the development of many benchmark datasets, each distinguished by its specific features. We present an overview of several widely adopted datasets, with their key characteristics summarized in Table 1.

Shanghai Tech[61] is one of the largest and most widely used crowd-counting datasets in recent years. It contains 1,198 images with a total of 330,165 labeled annotations. The dataset is divided into Part A and Part B to represent various crowd densities and scene complexities. Part A contains 300 training images and 182 testing images sourced from the internet and features dense crowd scenes. Part B includes 400 training images and 316 testing images taken on Shanghai streets, which reflect sparse crowd distributions. The dataset shows an imbalance in image density, with a higher prevalence of low-density images in both the training and testing sets. Additionally, the scale and perspective variations introduce challenges and opportunities for designing CNN-based network architectures.

UCF_CC_50 [62] represents the first challenging dataset created from publicly accessible web images. It consists of 50 images with various resolutions across a range of scenes, including concerts, protests, stadiums, and marathons, which showcase diverse density levels and perspective distortions.

UCF-QNRF [63] is a challenging collection of 1,535 high-resolution crowd images, with 1,201 used for training and 334 for testing. It contains approximately 1.25 million annotations, which cover diverse scenes with varying perspectives, densities, and lighting conditions. The average resolution of these images is $2,013 \times 2,902$ pixels. Additionally, the dataset includes authentic outdoor scenes from across the globe, capturing various elements such as buildings, vegetation, sky, and roads. This diversity is crucial for examining regional differences in population density.

NWPU-Crowd dataset [64] contains 5,109 images with a total of 2,133,238 annotated individuals, and an average resolution of $2,191 \times 3,209$ pixels. This dataset includes negative samples, which enhance model robustness during training. Compared to other datasets, it offers greater diversity in scale, density, and background. Additionally, it includes negative samples without any individuals, further enriching the dataset's overall diversity.

WorldExpo'10 dataset [65] is a comprehensive cross-scene crowd counting dataset from the 2010 Shanghai World Expo. It includes 1,132 annotated video sequences recorded by 108 surveillance cameras, comprising 3,920 frames at a resolution of 576×720 and annotations for 199,923 individuals. The training data covers 3,980 frames from 103 scenes, while the testing data consists of 600 frames from an additional 5 scenes. This diversity of scenes supports crowd counting research across varied environmental contexts.

JHU-CROWD++ [66] comprises 4,372 images from varied online scenes, with an average of 346 annotations per image, and a peak annotation count of 25,791. It

captures a range of weather conditions, such as rain, snow, and haze. It also offers comprehensive annotations at both the image and head levels, with each annotation including head position, size, occlusion level, and blur. These rich details enhance data quality for more efficient model training.

Table 1: Information of the datasets adopted for comparison.

Dataset	# Images	Train	Val	Test	Average resolution	Min	Max	Avg	Total
Part A [61]	482	300	-	182	589 × 868	33	3,139	501	241,677
Part B [61]	716	400	-	316	768 × 1,024	9	578	123	88,488
UCF_CC_50 [62]	50	40	-	10	-	94	4,543	1,280	63,705
UCF-QNRF [63]	1,535	1,201	-	334	2,013 × 2,902	49	12,865	815	1,251,642
NWPU-Crowd [64]	5,109	3,190	500	1,500	2,191 × 3,209	0	20,033	418	2,133,238
WorldExpo'10 [65]	3,920	3,380	-	600	576 × 720	-	-	-	199,923
JHU-CROWD++ [66]	4,372	2,272	500	1,600	910 × 1,430	0	25,791	346	1,515,005

2.2. Evaluation Protocols

In crowd counting, the two most commonly used criteria are Mean Absolute Error (MAE) and Mean Square Error (RMSE), which are defined as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|C_{I_i}^{pred} - C_{I_i}^{gt}\right|, \tag{1}$$

$$\mathrm{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|C_{I_i}^{pred} - C_{I_i}^{gt}\right|^2}, \tag{2}$$

where N is the number of the test image, $C_{I_i}^{pred}$and $C_{I_i}^{gt}$ represent the prediction results and ground truth, respectively. Especially, MAE determines the accuracy of the estimates, while RMSE indicates the robustness of the estimates.

To assess the performance of lightweight networks, four metrics, namely parameters(Params), floating-point operations (FLOPs), frames per second (FPS), and inference time [44, 52, 67, 68] are commonly adopted. Parameters reflect memory and storage needs, crucial for edge deployment. FLOPs estimate the number of operations required during forward passes. It represents computational demand. FPS indicates how efficiently the model processes real-time data, while inference time focuses on the latency for a single input. These indicators examine the model from multiple aspects, *i.e.,* size, speed, complexity, and delay. They form a foundational framework for lightweight model evaluation [49, 69, 48, 70]. The formulas are described as follows,

$$\mathrm{Params} = \sum_{i=1}^{L}\left(W_i + B_i\right), \tag{3}$$

$$\text{FLOPs}_{\text{conv}} = 2 \times C_{\text{in}} \times H_{\text{out}} \times W_{\text{out}} \times K_h \times K_w \times C_{\text{out}}, \tag{4}$$

$$\text{Inference Time} = \frac{1}{M} \sum_{i=1}^{M} \text{Elapsed\_Time}(i), \tag{5}$$

$$\text{FPS} = \frac{1000}{\text{Inference Time}}, \tag{6}$$

where L denotes the number of layers in the model, W_i represents the weight parameters and B_i indicates the bias parameters for the i-th layer. C_{in} specifies the number of input channels, while H_{out} and W_{out} refer to the height and width of the output feature map, respectively. The convolution kernel dimensions are given by K_h (height) and K_w (width). C_{out} stands for the number of output channels. Additionally, Elapsed_Time(i) measures the inference time for each iteration i in milliseconds, and M represents the total number of iterations.

3. Lightweight Networks

With a growing emphasis on lightweight model architectures, researchers have developed compact neural networks such as MobileNet [47], ShuffleNet [48], and GhostNet [49]. These models maintain accuracy while reducing computational and storage demands [71]. Consequently, they are well-suited for deployment in resource-constrained environments, such as mobile devices and embedded systems. While lightweight networks have achieved notable optimizations in parameter and computation efficiency, they still face significant challenges in complex crowd scenes, such as identifying subtle variations within dense crowds and intricate backgrounds [44, 72, 46]. To address these issues, researchers have responded by designing networks specifically for crowd counting and analysis to strengthen performance and robustness in complex scenarios [73, 74, 75]. Further advancements in balancing lightweight design with improved adaptability and expressiveness remain a focus of this field. The following sections present a categorized overview of lightweight models and summarize recent advancements and innovative developments.

3.1. Lightweight Architecture-based Networks

Lightweight architecture-based networks reduce parameter counts and computational demands to create efficient neural models suitable for resource-limited environments [76, 69, 77]. These networks implement structural optimizations through strategies such as minimizing network depth and using depthwise separable and grouped convolutions. We divide these networks into two categories based on problem-oriented approaches. The first category includes methods that address the scale variation issue, such as MCNN [61], PCCNet [43], and some similar approaches [26, 54, 50, 67, 78, 79]. These methods focus on crowd distribution across various scales, especially in high-density scenes, and use techniques such as multi-scale feature extraction to manage

the differences between distant and near targets. The second category focuses on background noise, such as MCNet [80], TinyCount [81], and so on [51, 82, 83]. These methods aim at reducing the impact of background noise on crowd counting, particularly in environments with occlusion and complex backgrounds.

To solve the problem of scale variation in images, Zhang *et al.* [61] proposed the Multi-column Convolutional Neural Network (MCNN). This method employs three parallel branches with different receptive fields to extract features at various scales. It removes the fully connected layers and keeps only convolutional and pooling layers, with a final 1×1 convolutional layer used to produce the density map. This multi-column structure reduces computational costs, achieves model lightweight, and improves real-time performance, which makes it suitable for diverse crowd densities and multi-perspective scenarios. However, MCNN [61] primarily relies on local features and lacks global feature capture, which limits its accuracy in high-density scenes. Based on MCNN [61], Ma *et al.* [84] presented a cascaded small-filter approach to achieve finer multi-scale feature extraction while further reducing computational demands. This method designs a lightweight three-stage network that includes multi-scale feature extraction, density map estimation, and refinement to balance accuracy and efficiency. However, this network continues to struggle with capturing global features in high-density and complex scenes, which constrains its precision and generalization. Shi *et*

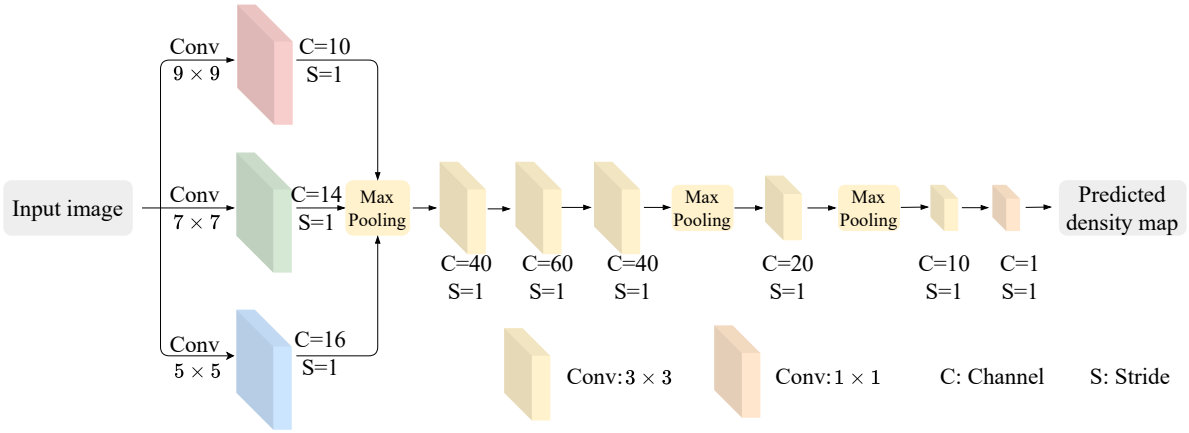

Figure 1: The architecture of the CCNN [50].

al. [50] developed the Compact Convolutional Neural Network (CCNN) to optimize computational efficiency further. As illustrated in Figure 1, CCNN [50] includes three parallel convolutional layers at the front, each with different kernel sizes to capture multi-scale local features. It merges the generated feature maps into a single unified feature map. This approach provides efficient real-time performance while maintaining a low computational cost.

However, CCNN [50] also has limitations due to the lack of a background noise suppression mechanism, which impacts its accuracy in complex scenes. Based on CCNN [50], Thai *et al.* [51] proposed the Dilated Compact Convolutional Neural

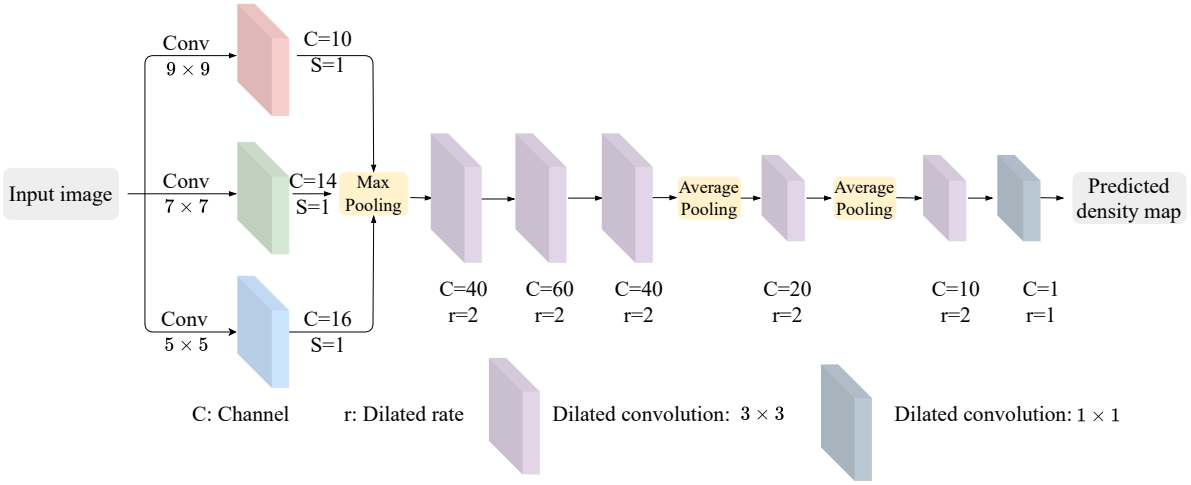

Figure 2: The framework of the DCCNN [51].

Network (DCCNN) to enhance background noise suppression in crowd counting. As depicted in Figure 2, DCCNN [51] introduces a dilated convolution in the second layer, with a dilation rate of 2. This adjustment expands the receptive field without increasing computational costs and strengthens noise suppression. Additionally, DCCNN [51] replaces max pooling with average pooling in the second and third layers. This modification helps reduce the loss of critical counting details. Batch normalization is also applied across all convolutional layers to improve model stability and accelerate convergence. Despite these advancements, DCCNN [51] still has limitations in capturing global features. It also performs less adaptively in sparse scenes, which affects its accuracy. To address the background noise issue, some studies combine multi-scale

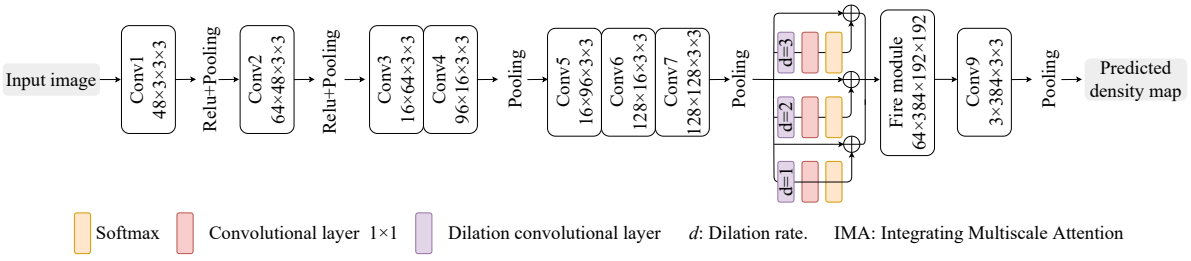

Figure 3: The framework of the MCNet [80].

features with attention mechanisms. Guo *et al.* [80] proposed MCNet, as shown in Figure 3. MCNet [80] improves model robustness in complex environments by merging multi-scale feature extraction with spatial attention mechanisms. First, it encodes the original image into high-level texture features through a series of convolutional layers, ReLU activations, and pooling operations. Then, these features are fed into the multi-scale attention layer. It uses multi-scale dilated convolutions and spatial attention to dynamically adjust the importance of various regions. This enables the network to focus on essential areas, improve accuracy in high-density regions, and effectively mitigate

background noise interference.

3.2. Lightweight Module-based Networks

To enhance the feature extraction and adaptability of lightweight crowd counting models, specialized modules such as lightweight attention and multi-scale feature fusion are added to the lightweight models. Compared to lightweight architecture-based networks, lightweight module-based networks focus on optimizing model characteristics through modular design. By introducing modular components, the model can flexibly select or adjust the appropriate modules based on task requirements. It improves the performance of the model in complex environments.

We classify the lightweight module-based networks into two categories. The first category focuses on addressing scale variation by using multi-scale feature fusion and adaptive modules. The second category addresses background noise. It uses specific modules to suppress background interference. For instance, lightweight attention mechanisms enable the model to focus on important regions while minimizing irrelevant background noise. Meanwhile, it maintains high accuracy even in complex environments.

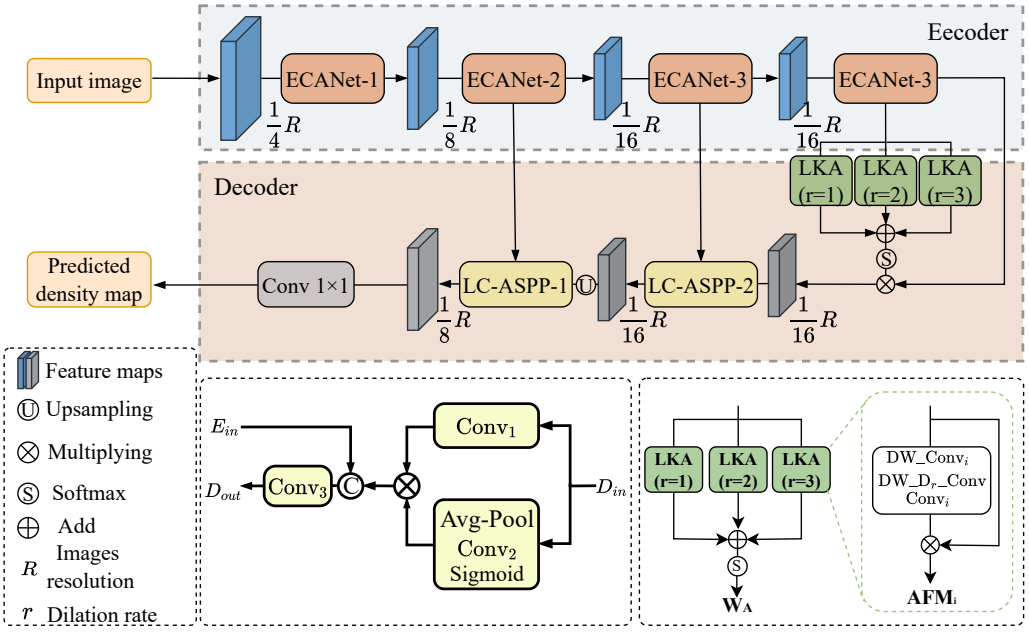

Figure 4: The architecture of the LEDCrowdNet [52].

To handle the challenge of scale variation, Yi *et al.* [52] proposed LEDCrowdNet, a lightweight network for crowd counting that optimizes both computational efficiency and accuracy. As shown in Figure 4, the LEDCrowdNet [52] used MobileViT [85] as the encoder to extract multi-scale crowd features. The decoder integrates enhanced AMLKA and LC-ASPP modules to generate high-quality density maps. The AMLKA module captures multi-scale features through convolutions with three dilation rates to address information loss associated with traditional large kernels. The LC-ASPP module

uses adaptive average pooling and sparse convolutions to improve feature localization accuracy and reduce computational cost. LEDCrowdNet [52] achieves a balance between efficiency and accuracy in sparse scenarios, but its feature extraction is less stable in dense environments.

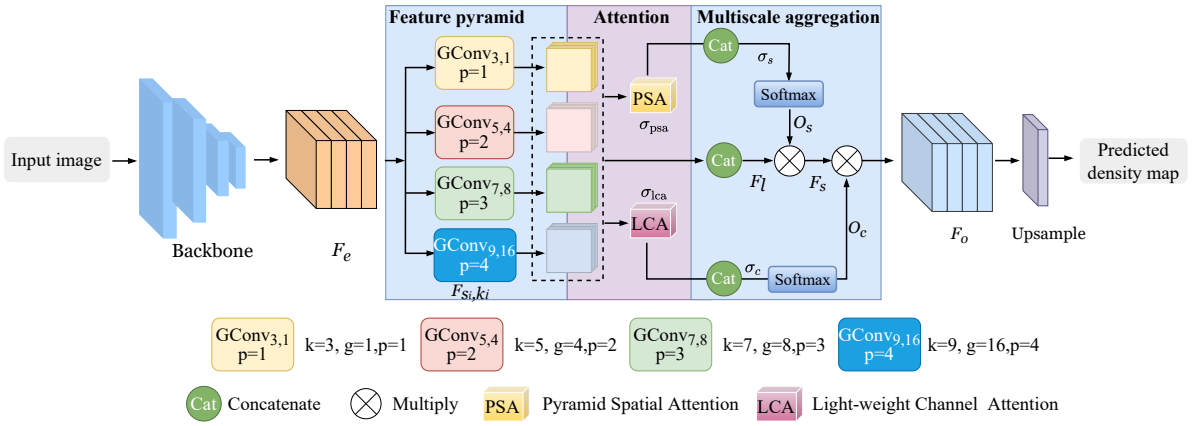

Figure 5: The framework of the FPANet [46].

Zhai *et al.* [46] developed FPANet, a lightweight network designed to address the challenges of scale variation in dense crowd scenarios. Based on shallow features from ResNet-50 [86], FPANet [46] includes a feature pyramid module, an attention module, and a multi-scale aggregation module (Figure 5). The feature pyramid module employs

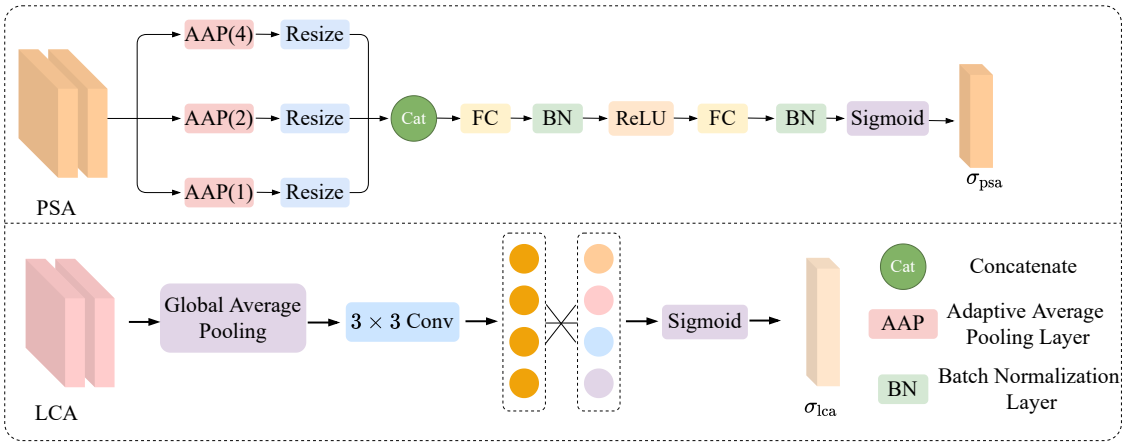

Figure 6: The architecture of the PSA unit and LCA unit [46].

multi-scale convolutional kernels to extract features in a pyramid structure, capturing crowd features across different scales. As illustrated in Figure 6, the attention module comprises a pyramid spatial attention (PSA) unit and a lightweight channel attention (LCA) unit. PSA highlights spatial features in head regions, reducing background interference, while LCA enhances connections between key channels for improved feature focus. The multi-scale aggregation module integrates spatial and channel attention information to enrich feature representation and increase adaptability to complex

crowd scenes. Although the two modules in FPANet [46] significantly enhance model performance, they also introduce additional computational overhead, which limits their applicability on devices with limited resources.

Liu *et al.* [55] introduced Lw-Count, a lightweight crowd counting network designed to address scale variation challenges while maintaining high accuracy and operational efficiency. As demonstrated in Figure 7 and Figure 8, this network uses a streamlined HRNet [87] as its baseline and introduces two key modules for enhanced performance: the Efficient Lightweight Convolution Module (ELCM) and the Scale Regression Module (SRM). ELCM utilizes a refined Ghost Block structure to extract features with minimal

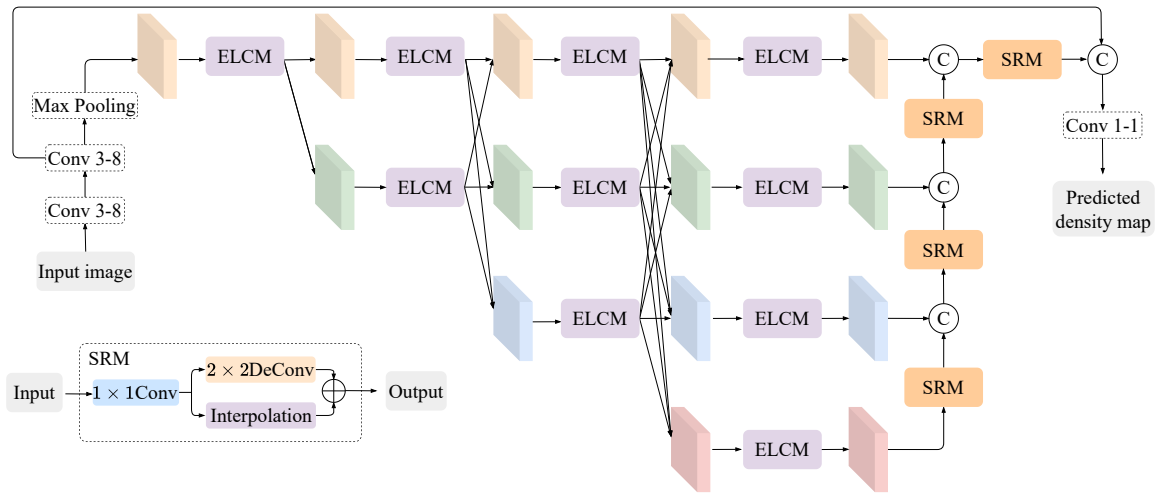

Figure 7: The architecture of the Lw-Count [55].

parameters, incorporating spatial group normalization to address counting challenges associated with uneven crowd distributions. The SRM module in the decoding stage aggregates multi-scale features layer by layer. This approach minimizes interpolation errors and prevents artifacts from transposed convolutions to enhance the quality of the density map. Additionally, Lw-Count [55] employs a region-normalized cross-correlation

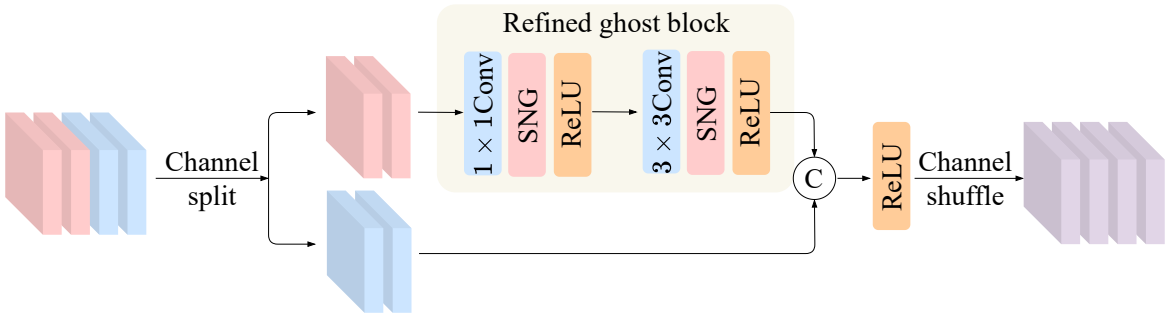

Figure 8: The structure of the ELCM [55].

loss to strengthen spatial consistency, which further refines counting accuracy. However,

the simplified HRNet [87] limits the ability of the network to capture fine-grained details across multiple scales in high-density and complex scenes. The redundant features generated by ghost block [49] are less effective in capturing boundary information, which reduces accuracy in high-density crowd regions.

Although multi-scale modules enhance the capacity of the model to handle scale variations, background noise remains a bottleneck under complex scenes. To address this, Guo *et al.* [44] proposed the Ghost Attention Pyramid Network (GAPNet) to suppress background interference. GAPNet [44] uses GhostNet [49] as the backbone to

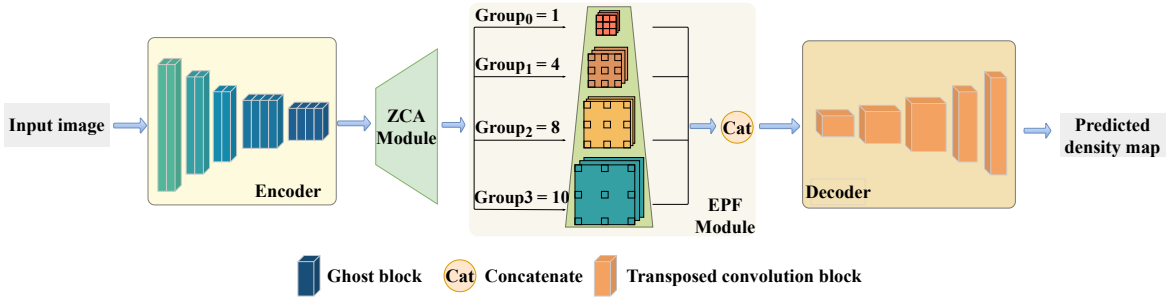

Figure 9: The framework of the GAPNet [44].

extract low-level features. As shown in Figure 9, the model takes a crowd image as input and generates the corresponding predicted density map as output. To identify discriminative crowd regions efficiently, GAPNet [44] introduces a zero-parameter channel attention (ZCA) (Figure 10) module that adjusts channel weights through linear transformations and activation functions. The ZCA module first applies global average pooling to obtain statistical features for each channel. It then introduces an energy-based weighting scheme to emphasize channels with significant differences from the target. This enhances the attention of the model to informative features. This module contains no learnable parameters. All attention scores are derived through predefined linear operations and normalization, which keeps the computation lightweight. Additionally, GAPNet [44] incorporates an Efficient Pyramid Fusion (EPF) module with a four-branch design to enhance background suppression. By combining group convolutions with dilated convolutions, the EPF module effectively filters out irrelevant background features while maintaining a low parameters. In the decoder stage, multiple transposed convolution layers generate the final density map. Although GAPNet [44] achieves reduced computational costs and improved efficiency, its streamlined design limits feature extraction depth, impacting accuracy in high-density settings.

Other lightweight module-based networks such as Chen *et al.* [88], Yi *et al.* [89], LMSNet [90], LigMSANet [91], MDCount [92], PDDNet [93], MLRNet [94], MobileCount [45], JMFEEL-Net [95], LSANet [96], and Yu *et al.* [97] address scale variation, whereas NeXtCrowd [98], ConNet [99], PSCC + DCL [53], LDNet [100], and LigMANet [101] focus on suppressing background interference.

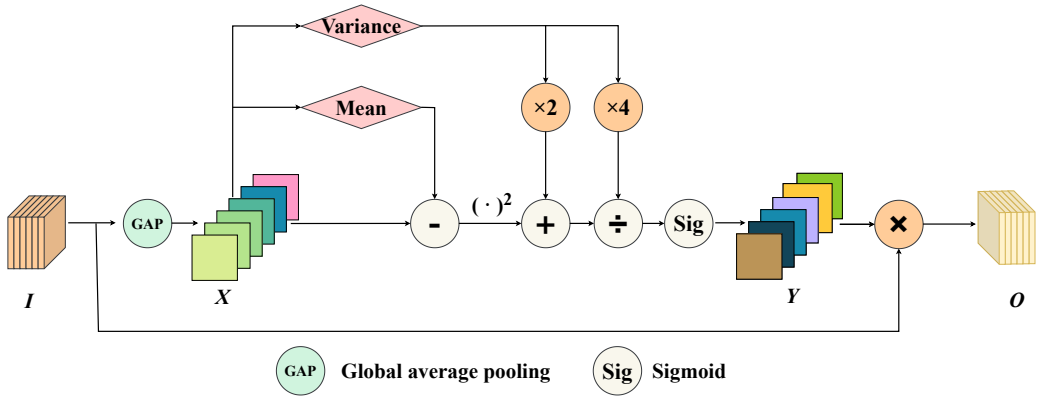

Figure 10: The structure of the ZCA module [44].

3.3. Knowledge Distillation-based Networks

Knowledge distillation enables lightweight networks by transferring expertise from a large teacher model to a smaller, yet more efficient student model [102, 103]. This strategy reduces parameters and computational load while preserving model accuracy. During distillation, the student model learns from both the ground truth and the guidance of the teacher. These additional cues help the model capture more detailed patterns and enhance its adaptability and accuracy in crowd counting tasks. By combining lightweight design with high accuracy, knowledge distillation is widely adopted for crowd counting applications in resource-limited settings.

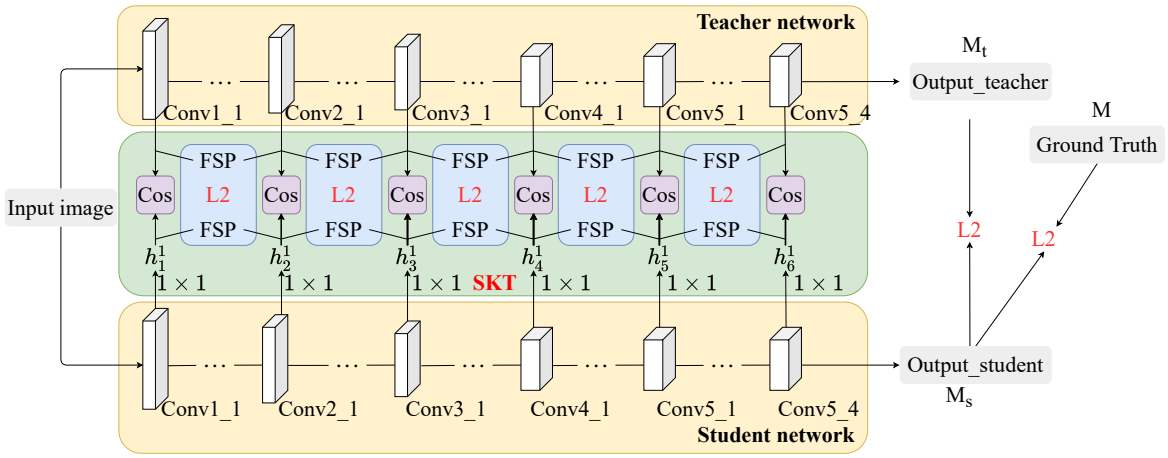

Figure 11: The structure of the SKT [58].

Liu *et al.* [58] proposed a lightweight crowd counting network termed the Structured Knowledge Transfer (SKT) framework, as shown in Figure 11. It is designed to effectively transfer knowledge from a trained large teacher model to a smaller student model. The SKT framework [58] employs CSRNet [24] as the teacher network and a simplified CSRNet [24] as the student network. To support effective knowledge transfer, SKT [58] includes two main modules: the Intra-Layer Pattern Transfer (Intra-PT) and

the Inter-Layer Relation Transfer (Inter-RT). Intra-PT allows for the stepwise transfer of intra-layer patterns from the teacher to guide the student's local feature learning, while Inter-RT captures relationships across layers in the teacher model, helping the student understand feature evolution. These modules equip the student model for efficient learning under lightweight design principles. However, due to its reliance on fixed feature patterns, SKT [58] has limited capacity to capture global features in dense and complex scenes. Moreover, the generalization ability of the student model depends on the performance of the teacher model.

Building on this, Liu *et al.* [104] introduced the ReviewKD method, which is an enhancement of the SKT framework. This approach strengthens feature learning in the

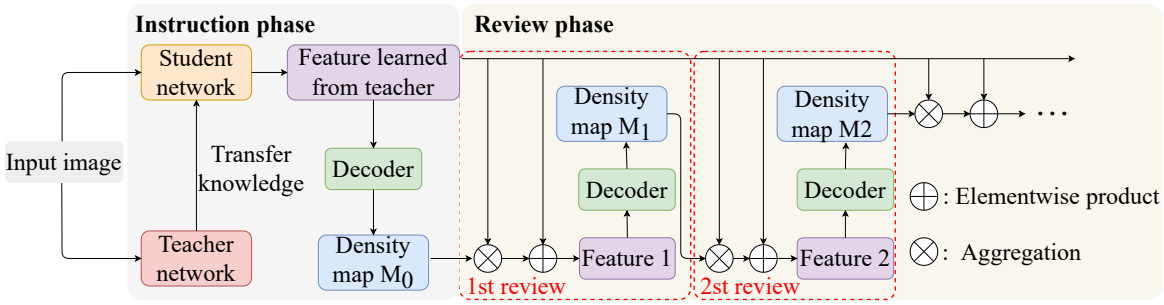

Figure 12: The architecture of the ReviewKD [104].

student model through two distillation stages: Instruction and Review, as demonstrated in Figure 12. In the Instruction stage, the teacher network gradually transfers its feature patterns to the student, guiding its learning process. In the Review stage, a density map serves as an attention weight to help the student focus on key areas and progressively improve feature extraction ability. This dual-stage approach narrows the performance disparity between teacher and student networks effectively. While ReviewKD [104] significantly improves the counting accuracy of the student network, it heavily relies on high-level features from the teacher model. This reliance limits robustness in complex environments. Furthermore, the review mechanism focuses on optimizing local features, neglecting the need for dynamic scene adaptation.

To address these limitations, Huang *et al.* [60] proposed a lightweight crowd counting network based on knowledge distillation, named Improved Knowledge Distillation (IKD), which aims to enhance counting performance in compact models. The framework of IKD [60] is illustrated in Figure 13. EffCC-lite2 [105] serves as the teacher network to provide knowledge for a student network. The student network, as a variant of EffCC-lite with reduced parameters, is suitable for deployment on resource-limited devices. IKD incorporates two main modules: self-transformed hints and outlier-tolerant loss, which address challenges related to information loss and outliers. In the IKD framework [60], the self-transformed hints module ensures feature dimensions remain consistent between teacher and student networks. This consistency eliminates the transformer dependency typically found in traditional knowledge distillation, thus reducing information loss. The outlier-tolerant loss module further improves the model

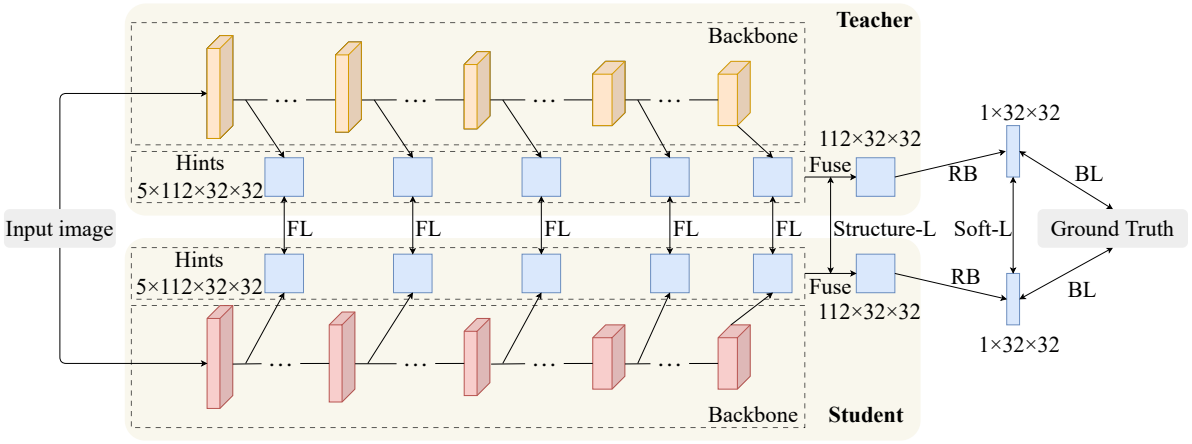

Figure 13: The framework of the IKD [60].

by limiting the influence of abnormal data during distillation. This enhancement increases the model's robustness and counting accuracy in high-density scenes. Although IKD [60] significantly improves the student network's counting accuracy, its adaptability and generalization in dynamic scenes remain limited.

Other knowledge distillation-based networks include ShuffleCount [56], DKD [57], Repmobilenet [59], D2PT [106], Gu [107], MJPNet-S* [108], EdgeCount [109], Duan *et al.* [110] and several others.

4. Analysis of Experimental Results

To comprehensively evaluate the performance of various crowd counting methods, this study presents a comparative analysis of representative models based on accuracy and efficiency, as shown in Table 2 and Table 3. The evaluation is conducted on four datasets, including ShanghaiTech Part A [61], ShanghaiTech Part B [61], UCF_CC_50 [62], UCF-QNRF [63], and NWPU [64], which are widely recognized benchmarks in the field of crowd counting.

As shown in Table 2, heavyweight models generally outperform the lightweight models in precision. APGCC [32] performs best across multiple datasets. It reaches 48.8 in MAE on ShanghaiTech Part A. On UCF-QNRF and NWPU datasets, where scenes are dense and complex, APGCC [32] records MAEs of 80.1 and 71.7, respectively. UEPNet [115] also performs well. It achieves an MAE of 6.9 and an RMSE of 10.6 on the ShanghaiTech Part B dataset, which demonstrates superior performance over most lightweight models. Other methods, such as STNet [116] and DLPTNet [35], achieve good performance in multiple metrics and provide further evidence for the superiority of large-scale architectures. On the other hand, although the performances of the lightweight models in accuracy are inferior to the heavyweight models, lightweight models demonstrate advantages in Parameters, which makes them suitable for edge devices and limited-resource settings. For instance, TinyCount [81] achieves 78.2 in

Table 2: Comparison of accuracy across different crowd counting methods. The upper part of the table represents heavyweight networks, and the lower parts represent lightweight networks. The "LA-based" represents Lightweight Architecture-based Networks, the "LM-based" represents Lightweight Module-based Networks, and the "KD-based" represents Knowledge Distillation-based Networks.

	Methods	Year	Part A		Part B		UCF_CC_50		UCF-QNRF		NWPU		Params(M)
			MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
Heavyweight	CSRNet [24]	2018	68.2	115.0	10.6	16.0	266.1	397.5	135.4	207.4	121.3	387.8	16.26
	CAN [111]	2019	62.3	100.0	7.8	12.2	212.2	243.7	107.0	183.0	106.3	386.5	18.10
	BL [112]	2019	61.5	103.2	7.5	12.6	229.3	308.2	87.7	158.1	105.4	454.2	21.50
	SFCN [113]	2021	64.8	107.5	7.6	13.0	214.2	318.2	102.0	171.4	105.7	424.1	38.60
	UOT [114]	2021	58.1	95.9	6.5	10.2	-	-	83.3	142.3	87.8	387.5	21.50
	UEPNet [115]	2021	54.6	91.2	6.4	10.9	165.2	275.9	81.1	131.7	-	-	26.12
	STNet [116]	2022	52.9	83.6	6.3	10.3	162.0	230.4	87.9	166.4	-	-	15.56
	SRRNet [11]	2023	60.8	103.0	7.4	13.6	172.9	256.3	89.5	162.9	-	-	66.14
	PET [117]	2023	49.3	78.8	6.2	9.7	-	-	79.5	144.3	74.4	328.5	20.90
	RAQNet [118]	2024	59.0	101.2	9.0	15.4	177.1	247.6	106.5	186.1	-	-	42.77
	DLPTNet [35]	2024	58.4	95.0	9.3	15.6	-	-	121.0	225.8	103.3	421.9	110.90
	SDANet [25]	2024	54.9	90.4	7.1	12.0	104.1	154.4	107.3	195.5	-	-	68.50
	APGCC [32]	2024	48.8	76.7	5.6	8.7	154.8	205.5	80.1	136.6	71.7	284.4	18.68
LA-based	MCNN [61]	2016	110.2	173.2	26.4	41.3	377.6	509.1	277.0	426.0	232.5	714.6	0.13
	SANet [26]	2018	75.3	122.2	10.5	17.9	258.4	334.9	152.6	547.0	190.6	491.4	0.91
	TDF-CNN [83]	2018	97.5	145.1	20.7	32.8	354.7	491.4	-	-	-	-	0.13
	ACSCP [79]	2018	75.7	102.7	17.2	27.4	291.0	404.6	-	-	-	-	5.10
	PCCNet [43]	2019	73.5	124.0	11.0	19.0	240.0	315.5	148.7	247.3	-	-	0.55
	LCNet [84]	2019	93.3	149.0	15.3	25.2	326.7	430.6	-	-	-	-	0.86
	CCNN [50]	2020	88.1	141.7	14.9	22.1	-	-	-	-	-	-	0.07
	DCCNN [51]	2020	84.1	133.5	12.2	21.9	-	-	-	-	-	-	0.07
	Li *et al.* [67]	2024	63.8	110.1	7.1	12.1	239.6	332.9	90.4	217.8	-	-	0.07
	TinyCount [81]	2024	78.2	120.8	10.8	18.4	-	-	134.7	223.3	-	-	0.06
	CPAS [78]	2024	64.5	102.9	7.1	10.9	115.0	139.6	94.2	164.4	-	-	2.20
LM-based	Yu *et al.* [97]	2019	78.5	126.4	12.8	22.1	299.1	391.8	-	-	-	-	0.13
	MobileCount [45]	2020	89.4	146.0	9.0	15.4	284.8	392.8	131.1	222.6	-	-	3.40
	PSCC+DCL [53]	2020	65.0	108.0	8.1	13.3	-	-	108.0	182.0	-	-	8.96
	MDCount [92]	2021	84.2	130.7	11.8	19.2	103.1	158.1	111.3	203.0	-	-	5.33
	Lw-Count [55]	2022	69.7	100.5	10.1	12.4	239.3	307.6	149.7	238.4	90.2	311.8	0.07
	MLRNet [94]	2022	82.6	130.2	10.6	15.9	265.8	389.1	127.4	222.4	-	-	1.26
	LSANet [96]	2022	66.1	110.2	8.6	13.9	-	-	112.3	186.9	-	-	0.20
	LigMSANet [91]	2022	76.6	121.4	10.9	17.5	231.5	339.7	-	-	-	-	0.63
	LEDCrowdNet [52]	2023	74.6	118.6	8.9	14.1	195.6	282.2	122.6	199.4	-	-	2.06
	FPANet [46]	2023	70.9	120.6	8.8	15.5	159.5	218.4	108.9	197.6	97.1	372.8	7.80
	GAPNet [5]	2023	67.1	110.4	9.8	15.2	202.8	246.9	118.5	217.2	174.1	514.7	2.85
	Yi *et al.* [89]	2023	85.9	139.9	9.2	15.1	105.7	120.3	112.8	201.6	-	-	4.58
	PDDNet [93]	2023	72.6	112.2	10.3	17.0	-	-	130.2	246.6	91.5	381.0	1.10
	LMSNet [90]	2024	62.9	108.4	8.2	13.5	223.5	281.0	110.7	178.7	-	-	0.73
KD-based	1/4SAN+SKT [58]	2020	78.0	126.6	11.9	19.8	-	-	157.5	257.7	-	-	0.06
	ReviewKD-VSKT [104]	2022	61.5	101.3	7.2	12.0	192.0	277.6	88.2	149.0	-	-	2.20
	DKD [57]	2023	64.4	103.0	7.4	12.7	210.3	283.8	91.7	150.1	-	-	1.35
	Repmobilenet [59]	2024	84.2	127.5	8.6	13.7	-	-	122.5	216.2	-	-	3.41
	EdgeCount [109]	2024	69.0	118.6	8.1	13.7	-	-	111.4	189.2	-	-	1.32

MAE and 120.8 in RMSE on ShanghaiTech Part A with only 0.06 M parameters. It reflects a compact structure with acceptable accuracy. However, in dense and complex scenes, such as those in UCF-QNRF and NWPU, most lightweight methods report MAE values above 100. This accuracy gap highlights the need for improved feature representation and generalization in lightweight designs.

The efficiency of some state-of-the-art (SOTA) methods is verified on an RTX 3080Ti, and the results are shown in Table 3. We mainly focus on methods with open-source code that were tested with a consistent evaluation standard and unified input resolution of 576×768. The results show that lightweight models provide clear advantages in inference speed and real-time performance. TinyCount [81] requires only 3.37 ms per inference and achieves 296.65 FPS. EffCC-Lite0.25 [60] reaches 731.49 FPS and demonstrates strong potential for use in real-time applications. Even 1/4CSRNet+SKT [58], with a slightly larger size, attains 309.21 FPS and balances compression with speed. In contrast, most heavyweight models suffer from higher latency despite better accuracy. SRRNet [11] and SASNet [119] record inference times of 29.45 ms and 45.42 ms, with FPS below 35 FPS. These delays reduce their suitability for real-time applications.

Overall, Tables 2 and Tables 3 highlight a core challenge in current crowd counting research: heavyweight models lead in accuracy but pose deployment challenges due to high computational cost, whereas lightweight models offer superior speed and parameter efficiency, but they struggle to reach the accuracy of heavyweight networks. Future work should aim to reduce model complexity and improve inference speed without sacrificing accuracy, to achieve scalable and deployable solutions in real-world applications.

Table 3: Comparison of efficiency across different methods.

	Methods	Params (M)↓	FLOPs (G)↓	Time (ms)↓	FPS↑
Heavyweight	SRRNet [11]	66.14	162.09	29.45	33.96
	RAQNet [118]	42.77	250.86	36.83	27.15
	SASNet [119]	38.90	393.16	45.42	22.02
Lightweight	GAPNet [5]	2.85	3.29	4.27	234.40
	$\frac{1}{4}$ CSRNet+SKT [58]	1.12	12.92	3.23	309.21
	EffCC-Lite0.25 [60]	0.23	1.01	1.37	731.49
	TinyCount [81]	0.06	1.37	3.37	296.65

5. Challenging and Research Directions

5.1. Challenging

In the design and application of lightweight networks, researchers encounter several key challenges that significantly impact both model performance and practical utility.

The balance between accuracy and lightweight design. Lightweight networks offer significant advantages in reducing computational overhead, but they often come at the cost of accuracy, especially in high-density, detail-sensitive crowd counting tasks. As shown in Table 2, SRRNet [11] performs well across multiple datasets. It achieves good accuracy with MAE values of 60.8 for Part A and 7.4 for Part B. However, the parameters are 66.14 M, which results in a high computational cost. In contrast, methods like TinyCount [81] have a much smaller parameter size of only 0.06 M. While this results in a slight drop in accuracy (Part A MAE of 78.2), it improves inference speed (296.65 in FPS as shown in Table 3). This trade-off highlights the advantages of lightweight methods in real-time and resource-constrained environments, particularly for edge computing and embedded devices. Therefore, one of the core challenges for lightweight crowd counting networks remains achieving efficient and robust performance while minimizing accuracy loss in practical applications.

Scene complexity and model robustness. In practical crowd counting applications, environmental factors, *e.g.,* scene variations, occlusions, background noise, and varying crowd densities degrade model performance. While lightweight networks show advantages in computational efficiency, they often exhibit poor performance in robustness in complex environments, especially in high-density crowds and dynamic backgrounds. The QNRF dataset demonstrates the effect of scene diversity and crowd density on model performance. This dataset covers diverse crowd arrangements and environments. It requires the model to be highly adaptable to these variations. As shown in Table 2, although RAQNet [118] performs well in MAE (106.5) and RMSE (186.1) on the QNRF dataset, the inference times (36.83 ms, as shown in Table 3) is much longer due to the large parameters. In contrast, the lightweight network GAPNet [5] offers faster inference times (4.27 ms) but struggles with robustness in complex environments, such as occlusion and high-density crowds. Therefore, future research should focus on improving the robustness of lightweight networks across diverse scenarios, especially in handling environmental changes like complex backgrounds and lighting variations.

Lack of generalization ability. Generalization remains a significant challenge in the research of lightweight crowd counting models. Existing lightweight models typically rely on large labeled datasets for training and perform well on standard benchmarks. However, in practical applications, their generalization ability is limited due to differences in data distribution and environmental changes (such as location, weather, and density) in real-world scenarios. These factors lead to unstable performance in complex and dynamic environments. Additionally, unsupervised methods based on large-scale pre-trained models, such as Contrastive Language-Image Pretraining (CLIP) [120], can perform class-agnostic counting but their large parameters and lack of spatial awareness make them less effective in real-world applications. For example, Chen *et al.* [6] demonstrated that while the CLIP model can count objects based on textual instructions, it lacks sensitivity to object locations and focuses more on global content rather than precise positioning. Furthermore, CLIP [120] typically freezes its pre-trained encoders and ignores misalignments between modalities, which reduces its

effectiveness in counting tasks. Future work should focus on enhancing the generalization of lightweight crowd counting models to ensure high accuracy and stability across various environments.

5.2. Research Directions

To further enhance the performance and practicality of lightweight networks, future research can explore the following directions.

Optimized lightweight network compression technology. At present, lightweight models often reduce computational cost and memory usage through pruning and quantization. However, these techniques usually lead to a loss in accuracy. Future studies can focus on structure-aware pruning methods, such as channel and block pruning, as well as mixed-precision quantization strategies. For example, HAWQ-V3 [121] applies second-order information to implement mixed-precision quantization. This method significantly lowers model complexity while maintaining accuracy. Lw-Count [55] focuses on crowd counting and redesigns the network with lightweight components. It achieves high compression rates in parameters and FLOPs with little precision loss. Further improvements could come from considering spatial density patterns and regional differences in weights. This enables tailored compression strategies for better performance in dense crowd settings.

Dynamic adaptation to different environments. Changes in data distribution, such as differences between urban and rural areas, daytime and nighttime, or occlusion and clear views, often challenge the generalization of crowd counting models. To improve performance under these challenging environments, future research should explore domain adaptation and transfer learning methods for lightweight models. Methods like adversarial training and feature alignment can help bridge the gap between source and target domains. Nguyen *et al.* [122] proposed a self-training method that combines source domain labels with target domain unlabeled samples. It applies adversarial training and entropy map minimization to improve generalization in cross-domain settings. Additionally, Meta-learning methods like dynamic β-MAML [123] allow for rapid model adaptation to new domains. Moreover, Generative Adversarial Networks (GANs) can generate domain-invariant features or synthetic data and further improve model adaptability. For example, ASNet [124] employed adversarial learning with dual discriminators to minimize domain gaps and improve counting accuracy in complex environments. In future work, it is suggested to explore the practicality of these methods and address the domain adaptation challenges faced by lightweight crowd counting models.

Combination of self-supervised and unsupervised learning. In crowd counting tasks, lightweight models often face challenges such as high annotation costs and limited generalization due to the scarcity of labeled data. Future research should explore the combination of self-supervised and unsupervised learning to improve model performance and cross-domain generalization with limited labeled data. Semi-supervised learning

methods combine small amounts of labeled data with large unlabeled datasets to enhance model learning. For example, Lin *et al.* [125] proposed a method using pixel-level density regression and alternating consistency self-supervision to improve both accuracy and generalization. Unsupervised methods, such as the approach by Liu *et al.* [126], generate pseudo-labels to further enhance model capabilities. Their transfer learning approach also improves adaptability and cross-domain performance using unlabeled data. Additionally, combining self-supervised and unsupervised strategies, such as using Generative Adversarial Networks (GANs) to generate pseudo-labels, could further improve feature learning from unlabeled data. However, balancing higher cross-domain adaptability with model efficiency and stability in complex environments remains a key challenge for future research.

Hardware optimization and energy efficiency improvement. As the demand for lightweight crowd counting networks increases on edge devices and embedded systems, efficient inference and low power consumption become critical. It is crucial in resource-constrained applications like intelligent surveillance, autonomous driving, and drones. Future research should focus on optimizing the performance of lightweight networks on specific hardware platforms through hardware/software co-design. For instance, the Quantized Deconvolution Generative Adversarial Network (QDCGAN) model can be deployed on hardware platforms like FPGAs to achieve efficient inference while balancing throughput and resource usage. Alhussain *et al.* [127] proposed a hardware/software co-design method that implements QDCGAN on FPGA. It utilizes scalable dataflow architectures to improve inference speed while reducing resource consumption. This approach offers high parallelism and is effective for computationally intensive applications such as GANs, especially in edge computing scenarios. Future research should further explore the application of these technologies to enhance lightweight crowd counting models. Specifically, more future work is expected to optimize the balance between low power consumption and high performance across various hardware platforms.

6. Conclusion

With the rapid advancement of deep learning, crowd counting technology has made significant strides across various fields, particularly in accuracy and robustness. However, as model complexity increases, the demand for computational resources and training data has become a bottleneck in practical applications. Thus, optimizing computational efficiency while maintaining high accuracy has emerged as a critical challenge in the field of crowd counting. To address this issue, lightweight crowd counting methods have increasingly become a focal point of research. This paper presents a review of recent progress in lightweight crowd counting methods. We classify these methods into three main categories: lightweight architecture-based networks, lightweight module-based networks, and knowledge distillation-based networks. For each category, we detail the design principles and introduce key representative methods.

Another key contribution of this work is the analysis of current SOTA lightweight crowd counting networks, which clarifies the trade-off between accuracy and computational efficiency. Heavyweight models offer higher accuracy but come with a significant computational burden. Lightweight models offer significant advantages in inference speed and parameter efficiency. These advantages make them well-suited for real-time applications with limited resources. However, accuracy remains a challenge for these models. Although current optimization techniques have improved the precision of lightweight models to some extent, further researches are expected to enhance accuracy without compromising efficiency. Finally, we summarize current challenges and suggest potential directions for future research.

7. Acknowledgements

This work has been funded by the Natural Science Foundation of Shandong Province (No. ZR2022ME156), Shandong Province Undergraduate Teaching Reform Project (No. Z2024184) and Research Project on Artificial Intelligence Education in Shandong Province (SDDJ202501097).

8. Reference

[1] Wenzhe Zhai, Qilei Li, Ying Zhou, Xuesong Li, Jinfeng Pan, Guofeng Zou, and Mingliang Gao. Da 2 net: a dual attention-aware network for robust crowd counting. *Multimedia Systems*, 29(5):3027–3040, 2023.

[2] Muhammad Asif Khan, Hamid Menouar, and Ridha Hamila. Revisiting crowd counting: State-of-the-art, trends, and future perspectives. *Image and Vision Computing*, 129:104597, 2023.

[3] Xiangyu Guo, Mingliang Gao, Wenzhe Zhai, Qilei Li, Jinfeng Pan, and Guofeng Zou. Multiscale aggregation network via smooth inverse map for crowd counting. *Multimedia Tools and Applications*, 83(22):61511–61525, 2024.

[4] Lijia Deng, Qinghua Zhou, Shuihua Wang, Juan Manuel Górriz, and Yudong Zhang. Deep learning in crowd counting: A survey. *CAAI Transactions on Intelligence Technology*, 2023.

[5] Xiangyu Guo, Mingliang Gao, Jinfeng Pan, Jianrun Shang, Alireza Souri, Qilei Li, Alessandro Bruno, et al. Crowd counting via attention and multi-feature fused network. *Human-Centric Computing And Information Sciences*, 13, 2023.

[6] Jinyong Chen, Qilei Li, Mingliang Gao, Wenzhe Zhai, Gwanggil Jeon, and David Camacho. Towards zero-shot object counting via deep spatial prior cross-modality fusion. *Information Fusion*, page 102537, 2024.

[7] Wenzhe Zhai, Xianglei Xing, Mingliang Gao, and Qilei Li. Zero-shot object counting with vision-language prior guidance network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[8] Xiangyu Guo, Jinyong Chen, Guisheng Zhang, Guofeng Zou, Qilei Li, and Mingliang Gao. Cell counting via attentive recognition network. *International Journal of Computational Science and Engineering*, 27(1):1–8, 2024.

[9] Akshita Patwal, Manoj Diwakar, Vikas Tripathi, and Prabhishek Singh. Crowd counting analysis using deep learning: A critical review. *Procedia Computer Science*, 218:2448–2458, 2023.

[10] Jinyong Chen, Mingliang Gao, Xiangyu Guo, Wenzhe Zhai, Qilei Li, and Gwanggil Jeon. Object counting in remote sensing via selective spatial-frequency pyramid network. *Software: Practice and Experience*, 54(9):1754–1773, 2024.

[11] Xiangyu Guo, Mingliang Gao, Wenzhe Zhai, Qilei Li, and Gwanggil Jeon. Scale region recognition network for object counting in intelligent transportation system. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[12] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57:137–154, 2004.

[13] Sheng-Fuu Lin, Jaw-Yeh Chen, and Hung-Xin Chao. Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 31(6):645–654, 2001.

[14] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *2008 19th international conference on pattern recognition*, pages 1–4. IEEE, 2008.

[15] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2009.

[16] Ibrahim Saygin Topkaya, Hakan Erdogan, and Fatih Porikli. Counting people by clustering person detector outputs. In *2014 11th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pages 313–318. IEEE, 2014.

[17] Bo Wu and Ram Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International journal of computer vision*, 75:247–266, 2007.

[18] Ke Chen, Chen Change Loy, Shaogang Gong, and Tony Xiang. Feature mining for localised crowd counting. In *Bmvc*, volume 1, page 3, 2012.

[19] Anthony C Davies, Jia Hong Yin, and Sergio A Velastin. Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, 7(1):37–47, 1995.

[20] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[21] Xiangyu Guo, Marco Anisetti, Mingliang Gao, and Gwanggil Jeon. Object counting in remote sensing via triple attention and scale-aware network. *Remote Sensing*, 14(24):6363, 2022.

[22] Wenzhe Zhai, Mingliang Gao, Xiangyu Guo, Qilei Li, and Gwanggil Jeon. Scale-context perceptive network for crowd counting and localization in smart city system. *IEEE Internet of Things Journal*, 10(21):18930–18940, 2023.

[23] Xiangyu Guo, Mingliang Gao, Wenzhe Zhai, Qilei Li, Kyu Hyung Kim, and Gwanggil Jeon. Dense attention fusion network for object counting in iot system. *Mobile Networks and Applications*, 28(1):359–368, 2023.

[24] Yuhong Li, Xiaofan Zhang, and Deming Chen. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1091–1100, 2018.

[25] Jianyong Wang, Xiangyu Guo, Qilei Li, Ahmed M Abdelmoniem, and Mingliang Gao. Sdanet: scale-deformation awareness network for crowd counting. *Journal of Electronic Imaging*, 33(4):043002–043002, 2024.

[26] 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.

[27] Xiangyu Guo, Mingliang Gao, Wenzhe Zhai, Jianrun Shang, and Qilei Li. Spatial-frequency attention network for crowd counting. *Big data*, 10(5):453–465, 2022.

[28] Ching-Hao Wang, Kang-Yang Huang, Yi Yao, Jun-Cheng Chen, Hong-Han Shuai, and Wen-Huang Cheng. Lightweight deep learning: An overview. *IEEE consumer electronics magazine*, 2022.

[29] Fanghui Chen, Shouliang Li, Jiale Han, Fengyuan Ren, and Zhen Yang. Review of lightweight deep convolutional neural networks. *Archives of Computational Methods in Engineering*,

31(4):1915–1937, 2024.

[30] Jun Jiang, XinYue Wang, Mingliang Gao, Jinfeng Pan, Chengyuan Zhao, and Jia Wang. Abnormal behavior detection using streak flow acceleration. *Applied Intelligence*, pages 1–18, 2022.

[31] Wenzhe Zhai, Mingliang Gao, Marco Anisetti, Qilei Li, Seunggil Jeon, and Jinfeng Pan. Group-split attention network for crowd counting. *Journal of Electronic Imaging*, 31(4):041214–041214, 2022.

[32] I Chen, Wei-Ting Chen, Yu-Wei Liu, Ming-Hsuan Yang, Sy-Yen Kuo, et al. Improving point-based crowd counting and localization based on auxiliary point guidance. *arXiv preprint arXiv:2405.10589*, 2024.

[33] Xiangyu Guo, Mingliang Gao, Guofeng Zou, Alessandro Bruno, Abdellah Chehri, and Gwanggil Jeon. Object counting via group and graph attention network. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[34] Shaokai Wu and Fengyu Yang. Boosting detection in crowd analysis via underutilized output features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15609–15618, 2023.

[35] Jinyong Chen, Mingliang Gao, Qilei Li, Xiangyu Guo, Jianyong Wang, Xuening Xing, et al. Privacy-aware crowd counting by decentralized learning with parallel transformers. *Internet of Things*, 26:101167, 2024.

[36] Mingliang Gao, Alireza Souri, Mayram Zaker, Wenzhe Zhai, Xiangyu Guo, and Qilei Li. A comprehensive analysis for crowd counting methodologies and algorithms in internet of things. *Cluster Computing*, 27(1):859–873, 2024.

[37] Dao-Hui Ge, Hong-Sheng Li, Liang Zhang, R Liu, P Shen, and Qi-Guang Miao. Survey of lightweight neural network. *J. Softw*, 31:2627–2653, 2020.

[38] Feifei He, Chang Liu, Mengdi Wang, Enshan Yang, and Xiaoqin Liu. Network lightweight method based on knowledge distillation is applied to rv reducer fault diagnosis. *Measurement Science and Technology*, 34(9):095110, 2023.

[39] Yingwei Sun, Xiyu Liu, Xiaodi Zhai, Kuizhi Sun, Mengmeng Zhao, Yankang Chang, and Yan Zhang. Automatic pixel-level detection of tire defects based on a lightweight transformer architecture. *Measurement Science and Technology*, 34(8):085405, 2023.

[40] Deqiang He, Chenyu Liu, Yanjun Chen, Zhenzhen Jin, Xianwang Li, and Sheng Shan. A rolling bearing fault diagnosis method using novel lightweight neural network. *Measurement Science and Technology*, 32(12):125102, 2021.

[41] Hou-I Liu, Marco Galindo, Hongxia Xie, Lai-Kuan Wong, Hong-Han Shuai, Yung-Hui Li, and Wen-Huang Cheng. Lightweight deep learning for resource-constrained environments: A survey. *ACM Computing Surveys*, 2024.

[42] Yan Zhou, Shaochang Chen, Yiming Wang, and Wenming Huan. Review of research on lightweight convolutional neural networks. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pages 1713–1720. IEEE, 2020.

[43] Junyu Gao, Qi Wang, and Xuelong Li. Pcc net: Perspective crowd counting via spatial convolutional network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(10):3486–3498, 2019.

[44] Xiangyu Guo, Kai Song, Mingliang Gao, Wenzhe Zhai, Qilei Li, and Gwanggil Jeon. Crowd counting in smart city via lightweight ghost attention pyramid network. *Future Generation Computer Systems*, 147:328–338, 2023.

[45] Peng Wang, Chenyu Gao, Yang Wang, Hui Li, and Ye Gao. Mobilecount: An efficient encoder-decoder framework for real-time crowd counting. *Neurocomputing*, 407:292–299, 2020.

[46] Wenzhe Zhai, Mingliang Gao, Qilei Li, Gwanggil Jeon, and Marco Anisetti. Fpanet: feature pyramid attention network for crowd counting. *Applied Intelligence*, 53(16):19199–19216, 2023.

[47] Andrew G Howard. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[48] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018.

[49] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1580–1589, 2020.

[50] Xiaowen Shi, Xin Li, Caili Wu, Shuchen Kong, Jing Yang, and Liang He. A real-time deep network for crowd counting. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2328–2332. IEEE, 2020.

[51] Thien Thai and Ngoc Quoc Ly. Lightweight solution to background noise in crowd counting. In *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 185–190. IEEE, 2020.

[52] Jun Yi, Fan Chen, Zhilong Shen, Yi Xiang, Shan Xiao, and Wei Zhou. An effective lightweight crowd counting method based on an encoder-decoder network for the internet of video things. *IEEE Internet of Things Journal*, 2023.

[53] Qi Wang, Wei Lin, Junyu Gao, and Xuelong Li. Density-aware curriculum learning for crowd counting. *IEEE Transactions on Cybernetics*, 52(6):4675–4687, 2020.

[54] Shuo Wang, Ziyuan Pu, Qianmu Li, Yaming Guo, and Meng Li. Edge computing-enabled crowd density estimation based on lightweight convolutional neural network. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–7. IEEE, 2021.

[55] Yanbo Liu, Guo Cao, Hao Shi, and Yingxiang Hu. Lw-count: An effective lightweight encoding-decoding crowd counting network. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6821–6834, 2022.

[56] Minyang Jiang, Jianzhe Lin, and Z Jane Wang. Shufflecount: Task-specific knowledge distillation for crowd counting. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 999–1003. IEEE, 2021.

[57] Rui Wang, Yixue Hao, Long Hu, Xianzhi Li, Min Chen, Yiming Miao, and Iztok Humar. Efficient crowd counting via dual knowledge distillation. *IEEE Transactions on Image Processing*, 2023.

[58] Lingbo Liu, Jiaqi Chen, Hefeng Wu, Tianshui Chen, Guanbin Li, and Liang Lin. Efficient crowd counting via structured knowledge transfer. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2645–2654, 2020.

[59] Chenxi Lin and Xiaojian Hu. Efficient crowd density estimation with edge intelligence via structural reparameterization and knowledge transfer. *Applied Soft Computing*, 154:111366, 2024.

[60] Zuo Huang and Richard O Sinnott. Improved knowledge distillation for crowd counting on iot devices. In *2023 IEEE International Conference on Edge Computing and Communications (EDGE)*, pages 207–214. IEEE, 2023.

[61] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 589–597, 2016.

[62] Haroon Idrees, Imran Saleemi, Cody Seibert, and Mubarak Shah. Multi-source multi-scale counting in extremely dense crowd images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2547–2554, 2013.

[63] 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. In *Proceedings of the European conference on computer vision (ECCV)*, pages 532–546, 2018.

[64] Qi Wang, Junyu Gao, Wei Lin, and Xuelong Li. Nwpu-crowd: A large-scale benchmark for crowd counting and localization. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):2141–2149, 2020.

[65] Cong Zhang, Hongsheng Li, Xiaogang Wang, and Xiaokang Yang. Cross-scene crowd counting

via deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 833–841, 2015.

[66] Vishwanath A Sindagi, Rajeev Yasarla, and Vishal M Patel. Jhu-crowd++: Large-scale crowd counting dataset and a benchmark method. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2594–2609, 2020.

[67] Yong-Chao Li, Rui-Sheng Jia, Ying-Xiang Hu, and Hong-Mei Sun. A lightweight dense crowd density estimation network for efficient compression models. *Expert Systems with Applications*, 238:122069, 2024.

[68] Jing-an Cheng, Qilei Li, Jinyong Chen, and Mingliang Gao. Efficient vehicular counting via privacy-aware aggregation network. *Measurement Science and Technology*, 36(2):026213, 2025.

[69] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.

[70] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[71] Ying Liu, Peng Xiao, Jie Fang, and Dengsheng Zhang. A survey on image classification of lightweight convolutional neural network. In *2023 19th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 1–10. IEEE, 2023.

[72] Rui Wang, Yixue Hao, Yiming Miao, Long Hu, and Min Chen. Rt3c: Real-time crowd counting in multi-scene video streams via cloud-edge-device collaboration. *IEEE Transactions on Services Computing*, 2024.

[73] Mengru Feng, Jiangjun Hu, Minghui Ou, and Dongchun Li. Crowd counting algorithm based on depthwise separable of dilated convolution. *Procedia Computer Science*, 208:319–324, 2022.

[74] Shaopeng Yang, Weiyu Guo, and Yuheng Ren. Crowdformer: An overlap patching vision transformer for top-down crowd counting. In *IJCAI*, volume 1, page 2, 2022.

[75] Shuo Wang, Ziyuan Pu, Qianmu Li, and Yinhai Wang. Estimating crowd density with edge intelligence based on lightweight convolutional neural networks. *Expert Systems with Applications*, 206:117823, 2022.

[76] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[77] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018.

[78] Qian Li, Chao Ma, Hao Chen, Xinyuan Chen, and Xiaokang Yang. Combinatorial progressive architecture search for crowd counting. *Displays*, 83:102686, 2024.

[79] Zan Shen, Yi Xu, Bingbing Ni, Minsi Wang, Jianguo Hu, and Xiaokang Yang. Crowd counting via adversarial cross-scale consistency pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5245–5254, 2018.

[80] Qiang Guo, Rubo Zhang, and Di Zhao. Mcnet: A crowd denstity estimation network based on integrating multiscale attention module. *arXiv preprint arXiv:2403.20173*, 2024.

[81] Hyeonbeen Lee and Jangho Lee. Tinycount: an efficient crowd counting network for intelligent surveillance. *Journal of Real-Time Image Processing*, 21(4):153, 2024.

[82] Vishwanath A Sindagi and Vishal M Patel. Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2017.

[83] Deepak Babu Sam and R Venkatesh Babu. Top-down feedback for crowd counting convolutional neural network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[84] Xiangyu Ma, Shan Du, and Yu Liu. A lightweight neural network for crowd analysis of images with congested scenes. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 979–983. IEEE, 2019.

[85] Sachin Mehta and Mohammad Rastegari. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021.

[86] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. https://doi.org/10.1109/cvpr.2016.90.

[87] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3349–3364, 2020.

[88] Yongqi Chen, Huailin Zhao, Ming Gao, and Mingfang Deng. A weakly supervised hybrid lightweight network for efficient crowd counting. *Electronics*, 13(4):723, 2024.

[89] Jun Yi, Zhilong Shen, Fan Chen, Yiheng Zhao, Shan Xiao, and Wei Zhou. A lightweight multiscale feature fusion network for remote sensing object counting. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[90] Mengyuan Xi and Hua Yan. Lightweight multi-scale network with attention for accurate and efficient crowd counting. *The Visual Computer*, 40(6):4553–4566, 2024.

[91] Guoquan Jiang, Rui Wu, Zhanqiang Huo, Cuijun Zhao, and Junwei Luo. Ligmsanet: Lightweight multi-scale adaptive convolutional neural network for dense crowd counting. *Expert Systems with Applications*, 197:116662, 2022.

[92] Xiaolong Meng and Zhengfei Ren. Mdcount: a lightweight encoder-decoder architecture for resource-saving crowd counting. In *Journal of Physics: Conference Series*, volume 2024, page 012031. IOP Publishing, 2021.

[93] Lanjun Liang, Huailin Zhao, Fangbo Zhou, Mingyang Ma, Feng Yao, and Xiaojun Ji. Pddnet: lightweight congested crowd counting via pyramid depth-wise dilated convolution. *Applied Intelligence*, 53(9):10472–10484, 2023.

[94] Peirong Ji, Xiaofeng Xia, Zhiwei Wu, Fusen Wang, Xinyue Liu, and Jun Sang. Mlrnet: Towards real-time crowd counting with mobile-based lightweight framework. In *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*, pages 1201–1208. IEEE, 2022.

[95] Mingtao Wang, Xin Zhou, and Yuanyuan Chen. Jmfeel-net: a joint multi-scale feature enhancement and lightweight transformer network for crowd counting. *Knowledge and Information Systems*, pages 1–21, 2024.

[96] Fushun Zhu, Hua Yan, Xinyue Chen, and Tong Li. Real-time crowd counting via lightweight scale-aware network. *Neurocomputing*, 472:54–67, 2022.

[97] Yang Yu, Jifeng Huang, Wen Du, and Naixue Xiong. Design and analysis of a lightweight context fusion cnn scheme for crowd counting. *Sensors*, 19(9):2013, 2019.

[98] Jun Hu and Hui Han. Nextcrowd: Lightweight and efficient network design for dense crowd counting. In *2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, pages 90–97. IEEE, 2023.

[99] Regina Marie A Masilang, Bianca Joy R Benedictos, Mikayla M Tejada, Giann Jericho Mari F Marasigan, and Arren Matthew C Antioquia. Connet: Designing a fast, efficient, and robust crowd counting model through composite compression. *International Journal of Pattern Recognition and Artificial Intelligence*, 37(07):2350017, 2023.

[100] Ye Tian, Chengzhen Duan, Ruilin Zhang, Zhiwei Wei, and Hongpeng Wang. Lightweight dual-task networks for crowd counting in aerial images. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1975–1979. IEEE,

2021.

[101] Lander Peter E Cua, Jacob Bryan B Gaba, Hylene Jules G Lee, Ian Angelo T Racoma, and Arren Matthew C Antioquia. Ligmanet: Towards designing a lightweight crowd counting model. In *2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pages 131–135. IEEE, 2023.

[102] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.

[103] Lin Wang and Kuk-Jin Yoon. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE transactions on pattern analysis and machine intelligence*, 44(6):3048–3068, 2021.

[104] Yunxin Liu, Qiaosi Yi, and Jinshan Zeng. Reducing capacity gap in knowledge distillation with review mechanism for crowd counting. *arXiv preprint arXiv:2206.05475*, 2022.

[105] Zuo Huang, Richard Sinnott, and Qiuhong Ke. Crowd counting using deep learning in edge devices. In *Proceedings of the 2021 IEEE/ACM 8th International Conference on Big Data Computing, Applications and Technologies*, pages 28–37, 2021.

[106] Fan LI, Enze YANG, Chao LI, Shuoyan LIU, and Haodong WANG. D2pt: Density to point transformer with knowledge distillation for crowd counting and localization. *IEICE Transactions on Information and Systems*, 2024.

[107] Yue Gu. Perspective-aware distillation-based crowd counting. In *Proceedings of the 2020 4th International Conference on Deep Learning Technologies*, pages 123–128, 2020.

[108] Wujie Zhou, Xun Yang, Xiena Dong, Meixin Fang, Weiqing Yan, and Ting Luo. Mjpnet-s*: Multistyle joint-perception network with knowledge distillation for drone rgb-thermal crowd density estimation in smart cities. *IEEE Internet of Things Journal*, 2024.

[109] Zhilong Shen, Guoquan Li, Ruiyang Xia, Hongying Meng, and Zhengwen Huang. A lightweight object counting network based on density map knowledge distillation. 2024.

[110] Zuodong Duan, Shunzhou Wang, Huijun Di, and Jiahao Deng. Distillation remote sensing object counting via multi-scale context feature aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2021.

[111] Weizhe Liu, Mathieu Salzmann, and Pascal Fua. Context-aware crowd counting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5099–5108, 2019.

[112] Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Bayesian loss for crowd count estimation with point supervision. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6142–6151, 2019.

[113] Qi Wang, Junyu Gao, Wei Lin, and Yuan Yuan. Pixel-wise crowd understanding via synthetic data. *International Journal of Computer Vision*, 129(1):225–245, 2021.

[114] Zhiheng Ma, Xing Wei, Xiaopeng Hong, Hui Lin, Yunfeng Qiu, and Yihong Gong. Learning to count via unbalanced optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2319–2327, 2021.

[115] Changan Wang, Qingyu Song, Boshen Zhang, Yabiao Wang, Ying Tai, Xuyi Hu, Chengjie Wang, Jilin Li, Jiayi Ma, and Yang Wu. Uniformity in heterogeneity: Diving deep into count interval partition for crowd counting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3234–3242, 2021.

[116] Mingjie Wang, Hao Cai, Xianfeng Han, Jun Zhou, and Minglun Gong. Stnet: Scale tree network with multi-level auxiliator for crowd counting. *IEEE Transactions on Multimedia*, 2022.

[117] Chengxin Liu, Hao Lu, Zhiguo Cao, and Tongliang Liu. Point-query quadtree for crowd counting, localization, and more. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1676–1685, 2023.

[118] Wenzhe Zhai, Xianglei Xing, and Gwanggil Jeon. Region-aware quantum network for crowd counting. *IEEE Transactions on Consumer Electronics*, 2024.

[119] Qingyu Song, Changan Wang, Yabiao Wang, Ying Tai, Chengjie Wang, Jilin Li, Jian Wu, and

Jiayi Ma. To choose or to fuse? scale selection for crowd counting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 2576–2583, 2021.

[120] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[121] Zhewei Yao, Zhen Dong, Zhangcheng Zheng, Amir Gholami, Jiali Yu, Eric Tan, Leyuan Wang, Qijing Huang, Yida Wang, Michael Mahoney, et al. Hawq-v3: Dyadic neural network quantization. In *International Conference on Machine Learning*, pages 11875–11886. PMLR, 2021.

[122] Pha Nguyen, Thanh-Dat Truong, Miaoqing Huang, Yi Liang, Ngan Le, and Khoa Luu. Self-supervised domain adaptation in crowd counting. In *2022 IEEE international conference on image processing (ICIP)*, pages 2786–2790. IEEE, 2022.

[123] Xiaoyu Hou, Jihui Xu, Jinming Wu, and Huaiyu Xu. Cross domain adaptation of crowd counting with model-agnostic meta-learning. *Applied Sciences*, 11(24):12037, 2021.

[124] Zhikang Zou, Xiaoye Qu, Pan Zhou, Shuangjie Xu, Xiaoqing Ye, Wenhao Wu, and Jin Ye. Coarse to fine: Domain adaptive crowd counting via adversarial scoring network. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 2185–2194, 2021.

[125] Hui Lin, Zhiheng Ma, Rongrong Ji, Yaowei Wang, Zhou Su, Xiaopeng Hong, and Deyu Meng. Semi-supervised counting via pixel-by-pixel density distribution modelling. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (01):1–14, 2025.

[126] Yuting Liu, Zheng Wang, Miaojing Shi, Shin'ichi Satoh, Qijun Zhao, and Hongyu Yang. Towards unsupervised crowd counting via regression-detection bi-knowledge transfer. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 129–137, 2020.

[127] Azzam Alhussain and Mingjie Lin. Hardware-efficient deconvolution-based gan for edge computing. In *2022 56th Annual Conference on Information Sciences and Systems (CISS)*, pages 172–176. IEEE, 2022.