Evaluation Framework for Crowd Counting Analysis Using Deep Learning Techniques

Fatima Jawad Kadhim and Ayad Hameed Mousa
College of Computer Science and Information Technology, University of Kerbala, 56001 Kerbala, Iraq
m05171132@s.uokerbala.edu.iq, *ayad.h@uokerbala.edu.iq*

Keywords: Crowd Counting, Convolutional Neural Network, Deep Learning, Evaluation Framework, Systematic Literature Review.

Abstract: Crowd counting refers to the process of counting the number of people in a specific area. It is a technique with broad applications in urban planning, healthcare, emergency management, security, and military strategies. However, this technique faces challenges such as visual distortions, perspective variation, and heterogeneous distribution of individuals, which increase the difficulty of calculations, especially in densely populated areas. Recent advances in convolutional neural networks (CNNs) and the creation of large datasets have contributed to significant advances in crowd counting methods in recent years. While deep learning has greatly advanced the field, a comprehensive analysis of the methodologies, challenges, and limitations of recent studies is lacking. This paper addresses this gap by: (1) conducting a Systematic Literature Review (SLR) of crowd-counting research; (2) proposing a novel evaluation framework (EFC2ADL) to classify and assess studies based on key criteria like datasets, loss functions, and metrics; and (3) validating the framework's relevance and comprehensiveness through expert review before employing it to evaluate 50 recent papers. The proposed Framework provides a structured basis for understanding trends and guiding future research in deep learning-based crowd counting.

1 INTRODUCTION

Accurate crowd counting contributes to a wide range of applications such as video surveillance[1] , public safety management [2], [3], and human behavior analysis [4], [5]. The algorithms can also be expanded to other fields such as cell microscopy [6], [7], vehicle counting [8], and environmental survey [9]. Conventional methods for crowd counting and density estimation have struggled due to occlusions and high clutter, especially when crowd density is very high, and substantial variations in head/body sizes and the uneven distribution of people.

The first traditional methods for crowd counting were detection-based counting methods. Detection methods are classified into two types: the model-based method, which segments individuals in the scene and detects each person individually, and then counts them using a model or human shape. However, the second method, based on trajectories, independently detects each movement in the scene by aggregating points of interest on the tracked individuals over time and then counts them. To improve the performance of the detection method, body detectors based on convolutional neural networks have been proposed. These detectors are more accurate in detecting objects because they extract features through the convolutional neural network, unlike simpler systems that rely on manually designed features. Examples of these methods include RCNN [10], Faster-RCNN [11].

Despite the performance of detection methods, they face difficulties in dealing with dense scenes. Therefore, researchers proposed a new method based on regression. These methods build a map from low-level features to crowd density. First, global features are extracted from the image. The next step after feature extraction is the training phase of the regression model to indicate a specific number of standard features. Examples of techniques for the regression method include linear regression [12], piecewise linear regression [13], and regression using a Gaussian mixture [14]. Recently, convolutional neural networks have demonstrated their efficiency in various fields, including computer vision and image processing, which has led researchers to be interested in employing them in crowd counting network architectures. The first to apply the convolutional

neural network approach was Min et al. [15]in crowd counting. This approach worked to determine the density level of crowds only without counting individuals in the crowd scene. These networks have high performance in challenging scenarios such as variations in the size of people's heads, differences in scenes and perspectives, and irregular density in the scene. In this work, we primarily focus on methods based on convolutional neural networks.

In this paper, a framework for evaluating studies related to crowd counting was proposed. The components of this framework were derived through an extensive and systematic investigation and analysis of the previous literature. The framework was then validated through an expert review method. It was then used to evaluate a carefully selected set of existing studies relevant to the research topic. The remainder of this paper will highlight the proposed framework, its components, and evaluation method, then use it to evaluate current research within the scope of the research. It will then be followed by results, conclusions, and future recommendations.

The main contributions of this paper are three aspects:

1) A systematic literature review (SLR) was conducted. This SLR aims to perform a critical assessment of crowd-counting problems.
2) We designed an assessment framework to evaluate relevant studies, identify their challenges, limitations, and obstacles, and derive future research directions that could enhance the performance of crowd counting techniques using deep learning.
3) We validated the proposed framework by using an expert review methodology, whereby the model was evaluated by twelve academic experts specialized in artificial intelligence and crowd management/counting.

The following paper is organized as follows: In section two, we explain how to determine and select relevant research and exclude unrelated studies, then we evaluate their methodological rigor. Finally, a systematic review of the literature was conducted. Section three presents our own evaluative framework designed to assess relevant studies. In section four, we validate the proposed framework. In section five, we discuss some important applications in the field of crowd counting.

2 METHOD

This paper's primary objective is to evaluate the effectiveness, obstacles, and constraints of crowd-counting studies and their methodologies. A systematic literature review (SLR) was conducted. This SLR aims to perform a critical assessment of crowd-counting problems.

2.1 Identification and Selection of Studies

The majority of relevant studies included in this selection utilized crowd detection as their primary data source for addressing the crowd counting problem. Crowd counting involves estimating the number of individuals within a defined region. This research investigates the methodologies employed by prior scholars in exploring machine learning techniques for crowd counting. This systematic literature review was conducted to investigate the application of deep learning methods in crowd counting research. The study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to systematically identify, evaluate, and synthesize relevant scholarly articles in the field.

The article selection process was guided by predefined inclusion and exclusion criteria. The authors conducted a systematic electronic literature search across prominent research databases, including Scopus, PubMed, IEEE Xplore, and ScienceDirect. The search strategy employed the following key terms: "Crowd Counting," "Crowd Counting Detection," "Crowd Counting Analysis," and "Crowd Counting Detection Using Deep Learning."." Crowd Counting", "Density Estimation in Crowd Counting", Deep Learning in Crowd Counting, Convolutional Neural Network in Crowd Counting, Density Maps in Crowd Counting" Following evaluation, 50 articles were designated and considered as potential candidates for inclusion in the study.

2.2 Inclusion and Exclusion Process

After an initial screening of articles identified through the search terms specified in Table 1, a total of 750 publications were retrieved. Following the removal of duplicates, the number of articles was reduced to 694. In the subsequent screening phase, two independent reviewers assessed the articles, resulting in the identification of 370 relevant publications. These eligible articles were then distributed among three reviewers for further evaluation to determine their suitability for inclusion in the next stage of analysis. The inclusion criteria were defined as follows:

1) The articles must be published works.

2) The proposed methodology must be based on deep learning techniques.

3) The articles must focus on crowd counting approaches utilizing deep learning, published between 2019 and 2025 in the English language.

Only studies meeting these criteria were advanced to the subsequent phase of the review process.

Table 1: Inclusion and exclusion process.

Inclusion	Exclusion
Only Articles that published in the English language.	All articles unrelated to crowd counting problems and solutions.
Only Articles published between 2019 and 2025.	Irrelevant and unrelated to crowd counting detection
Articles that investigated a type of crowd counting problems and solutions.	Articles that did not meet any of the inclusion criteria.

Following the screening stage, the selected articles were evaluated by five experts to assess their eligibility based on the criteria specified in Table 1. The experts collaboratively reviewed and discussed their assessments until a consensus was reached. Ultimately, 50 articles were selected for detailed analysis.

2.3 Evaluation of Methodological Rigor

This systematic review employed the Critical Appraisal Skills Program (CASP) checklist to conduct a comprehensive evaluation of the methodological quality of the selected articles. Key aspects and limitations were examined through data extraction, including sources, keywords, temporal scope, and geographical coverage. Additionally, the quality of the data—particularly its relevance to crowd-counting challenges—was assessed, along with study design considerations such as the appropriateness of the applied methodologies. The analysis further focused on study outcomes, including the clarity of research objectives and findings, to identify strengths and weaknesses across the reviewed literature.

The inclusion and exclusion criteria, as outlined in Table 1, guided the selection of articles. The final selection was based on factors such as title relevance, research objectives, outcomes, datasets utilized, feature extraction techniques, feature fusion approaches, and deep learning methodologies.

2.4 SLR Finding

Initially, a comprehensive search yielded 750 articles. After removing duplicates, 694 records remained. Subsequent screening of titles and abstracts led to the exclusion of irrelevant studies, leaving 370 full-text articles for further evaluation. These articles were rigorously assessed against the predefined eligibility criteria detailed in Table 1. To ensure methodological rigor, only prospective studies were included for in-depth analysis. The selected studies were then thoroughly examined to derive comprehensive findings and analytical conclusions. Figure 1 illustrated the distribution of the selected articles.

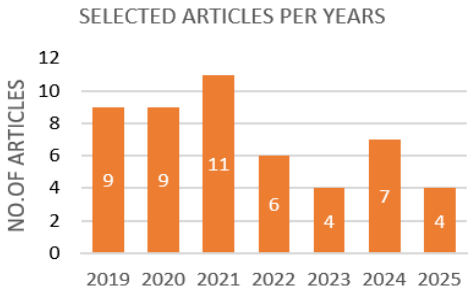


Figure 1: The distribution of selected articles.

The review included 50 articles published in English between 2019 and 2025. The systematic literature review (SLR) methodology followed a structured process, as illustrated in Figure 2, with protocols aligned to established guidelines.

3 THE PROPOSED FRAMEWORK

After conducting a thorough and systematic review of the relevant studies, the we concluded that there is a lack of established methods, frameworks, or approaches for evaluating these studies. This gap motivated the authors to propose an evaluative framework designed to assess relevant studies, identify their challenges, limitations, and constraints, and derive future research directions that could enhance the performance of crowd-counting techniques using deep learning. The analysis revealed that the majority of studies share key similarities, particularly in terms of dataset quantities, crowd counting techniques, proposed models, dataset categories, evaluation measures, and the types of loss functions employed. Figure 3 visualize the proposed framework and its components. In the ensuing paragraph, a brief description was given for every constituent of the suggested framework.

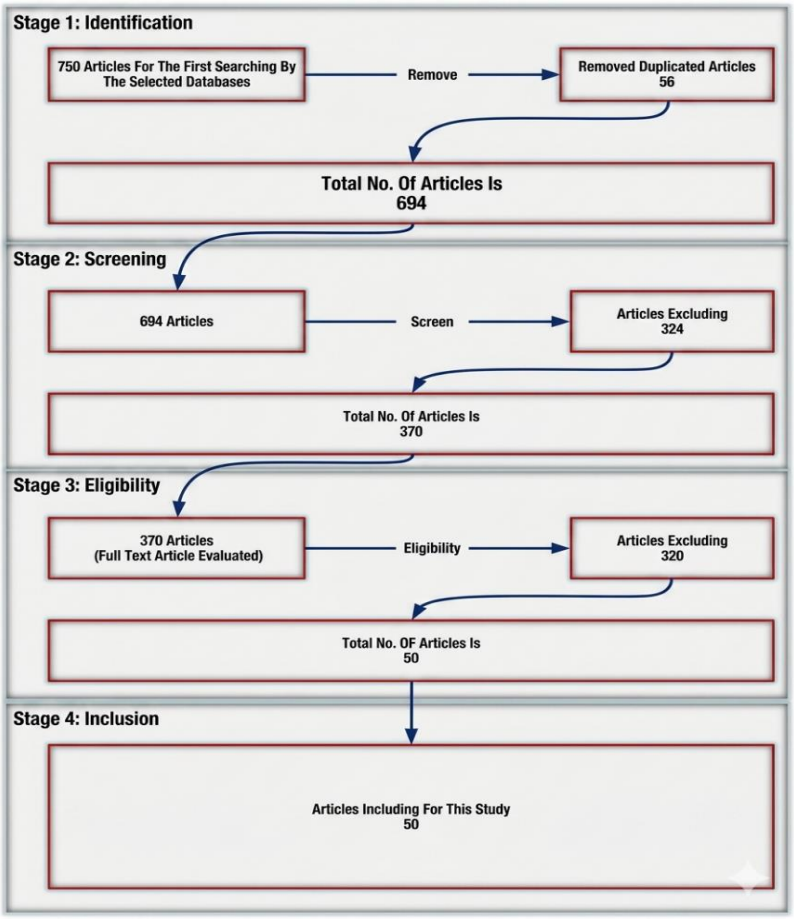


Figure 2: SLR protocol.

3.1 Loss Function Type

Loss functions play a central role in deep learning, determining how models learn and perform across a wide range of tasks. They measure the gap between predicted outputs and the actual ground truth, guiding the optimization process to reduce errors. Among the most common and widely used loss functions is for crowd counting: Mean Squared Error (MSE) [16].

3.2 The Proposed Artifact Name

This field includes the selected scientific name for each research paper according to the use of a specific type of crowd counting method and algorithm used to estimate density maps and crowd counts.

3.3 Dataset Utilized & No. of Datasets

Datasets play a significant and important role in training and evaluating crowd counting models. These datasets may contain various types of data, such as humans, cars, and pets. They include a wide range of factors related to the samples: different crowd scenarios, varying crowd densities, a wide range of head factors, and so on, to simulate the distribution of data in real-world applications. In terms of labeling accuracy as well, these datasets produce more accurate and realistic density maps. Among the most famous and commonly used datasets. A- Shanghai-Tech, UCF CC 50**,** UCSD**,** UCFQRNF.

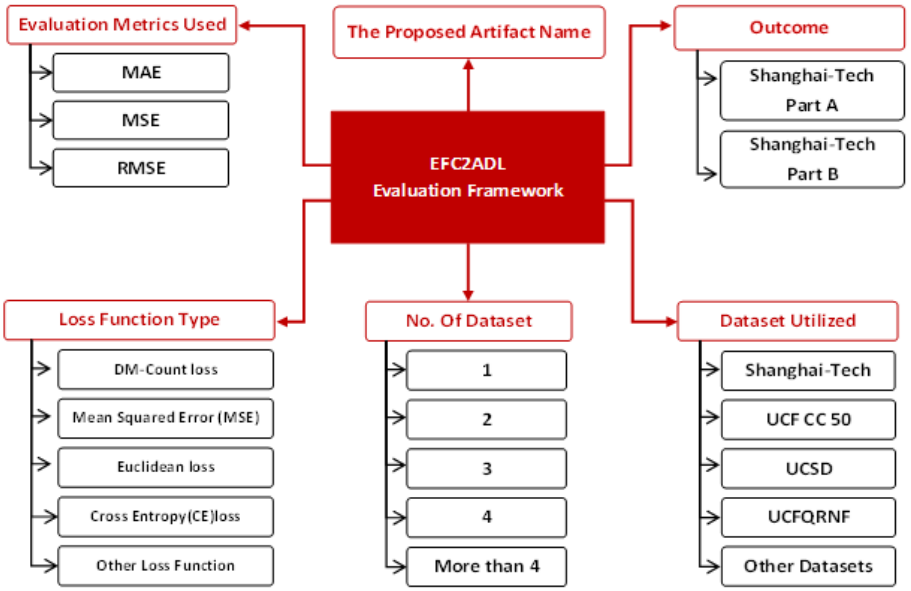


Figure 3: The proposed EFC2ADL framework.

3.4 Evaluation Metrics

These evaluation metrics can be used to compare the performance of different proposed methods for crowd counting on the various datasets mentioned earlier. Among the most common and widely used evaluation metrics for evaluating crowd counting models are:

1) The Mean Absolute Error (MAE): MAE calculates the absolute value of the average differences between actual numbers in the image and the expected numbers:

$$\text{Mean Absolute Error (MAE)} = 1/N \sum_{i=1}^{N} |yi - yi'| \quad (1)$$

2) The Root Mean Square Error (RMSE): RMSE calculates the root mean square of the differences between the actual values in the image and the predicted values:

$$\text{Root Mean Square Error (RMSE)} = \sqrt{1/N \sum_{i=1}^{N} (yi - yi')^2} \quad (2)$$

3) The Mean Squared Error (MSE): MSE calculates the absolute value of the average of the squares of the differences between the actual values in the picture and the expected values:

$$\text{Mean Squared Error (MSE)} = 1/N \sum_{i=1}^{N} (yi - yi')^2 \quad (3)$$

- N: is the number of test samples.
- yi: is the ground truth result corresponding to sample i.
- yi': is the estimated result corresponding to sample i[17].

Practical example (Table 2). To demonstrate how the proposed framework can be applied, the study "An object counting network based on hierarchical context and feature fusion (HFNet,2019)" was evaluated using EFC2ADL criteria [18].

1) Proposed Artifact Name. The researcher based the naming of the model HFNet on the context-based counting network, which relies on a hierarchical structure and feature integration.
2) Loss Function criterion. The researcher used the Euclidean loss to measure the distance between the estimated density map and the actual map, and this type of loss is the most common in the field of crowd counting.
3) Dataset Utilized & No. of Datasets criterion. This study uses the three datasets, Shanghai-Tech, UCF_CC_50, TRANCOS datasets; the first is the most famous dataset used for crowd counting.

4) Evaluation Metrics criterion. This study reported its results using only the evaluation metrics MAE and MSE.
5) Outcome.

Table 2. Practical application of the proposed EFC2ADL framework: evaluation of HFNet (2019) using hierarchical criteria, including model design, loss function, datasets, evaluation metrics, and reported performance results.

ShanghaiTech A		ShanghaiTech B		UCF_CC_50	
MAE	MSE	MAE	MSE	MAE	MSE
81.1	123.4	16.8	31.3	270.6	387.3

4 THE PROPOSED FRAMEWORK VALIDATION

To validate the proposed framework, we employed an expert review methodology, where the model was evaluated by twelve academic experts specializing in artificial intelligence and crowd management/counting in Table 3. Two validation criteria were adopted: comprehensibility and relevance.

Table 3: Academic experts.

	Field of Expertise	Experience (Year)	Work Location
1	Academician	12	University of Kerbala
2	Developers	14	Software development company
Total		12	

The experts were provided with a detailed explanation of the proposed framework, along with a validated evaluation tool. Their responses were collected, systematically analyzed, and the results are presented in Figures 4 and 5, which illustrate the final validation outcomes in terms of relevance and comprehensibility

Figure 4 indicates that most experts expressed agreement or strong agreement regarding the interrelatedness of the proposed framework's components, reflecting its relevance a ssessment.

Figure 5 indicates that most experts expressed agreement or strong agreement regarding the interrelatedness of the proposed framework's components, reflecting its comprehensibility assessment.

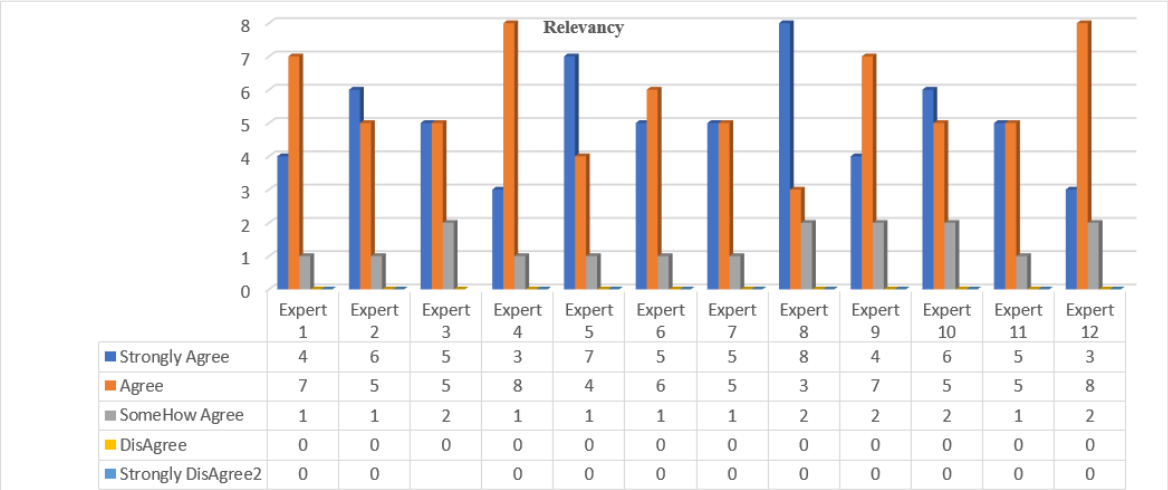
5 USING EFC2ADL PROPOSED MODEL FOR EVALUATION CROWD COUNTING STUDIES

In the context of this study, after proposing the proposed framework, which contains six dimensions, and after verifying its validity using the expert review method in terms of understanding and relevance between the dimensions, in this section, the proposed framework will be used to evaluate 50 recent studies relevant to the study topic.in Appendix A (Table A1) shows the results obtained from the evaluation process using the proposed framework

Table A1 shows a comprehensive comparison of 50 research papers in the field of crowd counting. Most of these studies utilized feature fusion techniques for multi-level or multi-scale deep convolutional neural networks. Meanwhile, some other studies indicated an interest in using attention mechanisms.

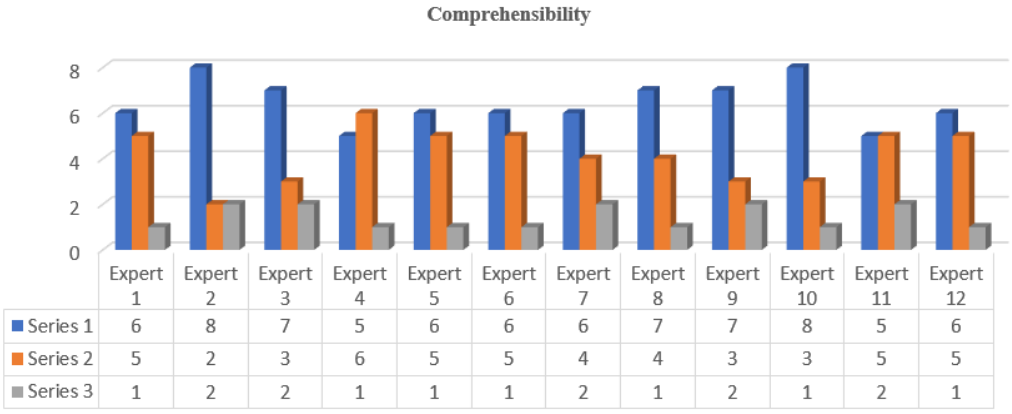
Despite the results achieved by these studies, there are still existing obstacles, including: a lack of generalization across different scenes, and insufficient focus on the use of post-processing techniques. These studies represent a real gap and highlight the need for more research.

Based on this review, it can be concluded that future trends will focus on generalizing models to different scenes, simplifying models and making them less complex, in addition to creating more diverse datasets for different scenes and densities, and using post-processing techniques. These trends reflect researchers' efforts to address shortcomings and achieve more accurate and efficient results in the future.



	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Expert 12
Strongly Agree	4	6	5	3	7	5	5	8	4	6	5	3
Agree	7	5	5	8	4	6	5	3	7	5	5	8
SomeHow Agree	1	1	2	1	1	1	1	2	2	2	1	2
DisAgree	0	0	0	0	0	0	0	0	0	0	0	0
Strongly DisAgree2	0	0		0	0	0	0	0	0	0	0	0

Figure 4: The validation result (relevancy).



	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Expert 12
Series 1	6	8	7	5	6	6	6	7	7	8	5	6
Series 2	5	2	3	6	5	5	4	4	3	3	5	5
Series 3	1	2	2	1	1	1	2	1	2	1	2	1

Figure 5: The validation result (comprehensibility**).**

6 RESULTS AND DISCUSSION

Figures 6 - 9 present the main results obtained from applying the framework to 50 selected studies, including the datasets used, the frequency of loss functions, and the evaluation metrics reported in the literature.

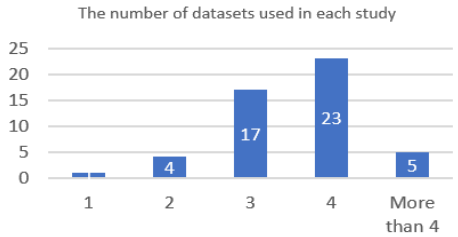


Figure 6: The number of datasets used in each study

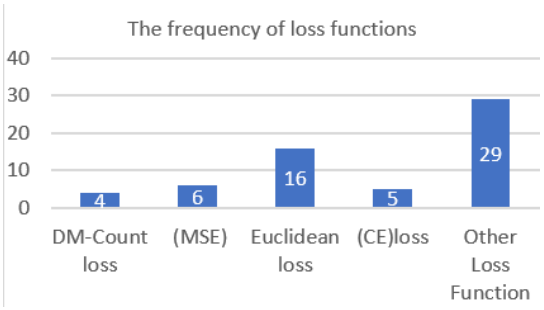


Figure 7: The frequency of loss functions**.**

The results indicate that using the proposed framework in the evaluation process of the fifty relevant studies is meaningful and achieves the following:

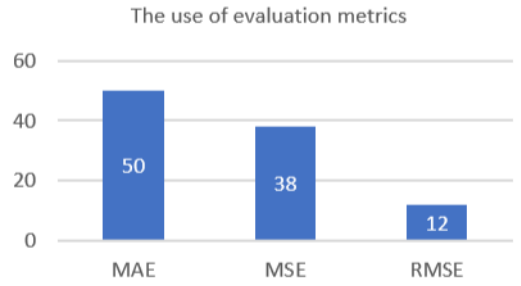


Figure 8: The use of evaluation metrics.

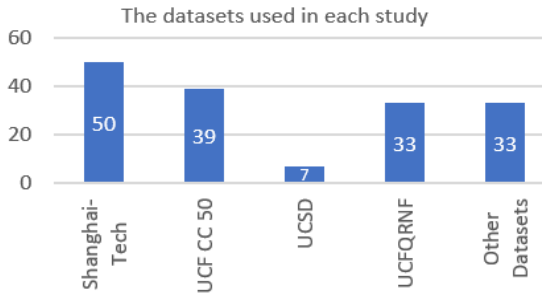


Figure 9: The datasets used in each study.

The vast majority used four academically well-known datasets for deep learning-based crowd counting.

1) All studies share the Shanghai-Tech dataset, indicating that any subsequent study should include this dataset in their proposed models.
2) Authors differed in their use of the loss function. Some used one or more loss functions alongside the mentioned loss functions, with the Euclidean loss being the most commonly used. This indicates the flexibility of researchers in using the appropriate loss function.
3) In terms of performance evaluation, the majority of researchers used the MAE and MSE, while a few used the RMSE to evaluate the performance of their proposed deep learning-based crowd counting models.
4) In terms of outcomes, this study focused on the results obtained using the Shanghai-Tech dataset, given its widespread use in most studies. The dataset is divided into two parts, A and B.

Crowd counting is considered an important and inspiring topic for researchers recently. Researchers have focused on developing crowd counting models using deep learning through convolutional neural networks and transformer-based methods, which have demonstrated their power and accuracy in estimating density maps and crowd counts, surpassing traditional crowd counting methods based on detection and regression approaches. The task of crowd counting faces common problems and challenges, namely scale variation, occlusion, and perspective distortion.

1) Scale variation. Variations in size pose a significant challenge in crowd counting and density estimation. These variations include differences in both the overall crowd size and the size of individual heads.
2) Occlusion. Occlusion is a common issue present in almost all crowd images, and it becomes increasingly severe as crowd density rises. In densely populated scenes, heavy occlusion makes the counting process highly challenging. To address this, most recent studies leverage the powerful feature extraction and learning capabilities of convolutional neural networks (CNNs) to mitigate the effects of occlusion and improve counting accuracy.
3) Uneven distribution. In many situations, people are unevenly distributed throughout the scene, leading to significant variations in crowd density.
4) Perspectives variation. Changes in camera position and viewing angle directly cause variations in scale within the image, as well as issues such as occlusion and uneven crowd distribution [19].

The researchers used techniques such as multi-scale and multi-level networks to address the problem of scale variations in terms of density and head size, in addition to attention mechanisms and convolutional neural networks to handle issues of occlusion and uneven distribution.

Researchers used techniques such as multi-scale and multi-level networks to address the problem of scale variation in terms of density and head size, in addition to attention mechanisms and convolutional neural networks to tackle issues of occlusion and uneven distribution. Previous work focused on explaining and clarifying the above techniques. Through our framework, we highlight the importance and impact of frame parameters on model accuracy and development. The selection and number of datasets contribute to generalizing the model across a larger number of images and diverse scenes, and choosing the appropriate loss function to feed the network helps increase counting accuracy and density estimation.

Although the proposed EFC2ADL framework provides a structured and comprehensive approach for assessing deep learning-based crowd counting

studies, several limitations must be acknowledged. First, the validation process was primarily qualitative, relying on expert judgment rather than large-scale quantitative experiments. Second, despite following the PRISMA protocol, there may be potential selection bias in the reviewed studies. Third, some assessment dimensions, such as model interpretability or reproducibility, involve subjective evaluations that may vary among reviewers. Additionally, the framework is domain-specific for crowd counting and may require adaptation before applying it to other computer vision tasks. Finally, as transformer-based models and foundation models continue to evolve, regular updates to the framework will be necessary to maintain its applicability and relevance.

7 CONCLUSIONS

In this research, a summary and evaluation of 50 research papers in the field of crowd counting were presented. After selecting certain studies and excluding others based on several factors, we assessed methodological rigor and conducted systematic literature reviews. In this study, a framework was proposed to evaluate these studies in terms of the datasets used and their sizes, the loss function, evaluation metrics, and evaluation results for each study based on the core datasets used in crowd counting research. The evaluation results were discussed in Section Six, highlighting the importance of this framework and its role in crowd counting research, as well as its relation to the main issues and challenges in crowd counting such as size variation, occlusion, uneven distribution, and perspective changes, the main techniques used to address these challenges, and their relation to the standards of this framework. Finally, we mention the limitations of this framework in terms of the verification process and bias in the selection of cross-sectional studies despite following the PRISMA protocol. Some evaluation dimensions, such as model interpretability or reproducibility, include subjective assessments that may vary between reviewers. In addition, this framework is restricted to the level of crowd counting only and may require further expansion to cover computer vision areas and new techniques used in crowd counting.

In line with the above, the proposed framework reveals that the authors are still striving to achieve higher accuracy in crowd counting, and that achieving higher accuracy remains a challenge, motivating researchers to further investigate this area.

REFERENCES

[1] F. Xiong, X. Shi, and D.-Y. Yeung, “Spatiotemporal modeling for crowd counting in videos,” in Proc. IEEE Int. Conf. Computer Vision, 2017, pp. 5151–5159.

[2] J. E. Almeida, R. Rosseti, and A. L. Coelho, “Crowd simulation modeling applied to emergency and evacuation simulations using multi-agent systems,” 2013. [Online]. Available: http://arxiv.org/abs/1303.4692

[3] A. Abdelghany, K. Abdelghany, H. Mahmassani, and W. Alhalabi, “Modeling framework for optimal evacuation of large-scale crowded pedestrian facilities,” Eur. J. Oper. Res., vol. 237, no. 3, pp. 1105–1118, 2014, doi: 10.1016/j.ejor.2014.02.054.

[4] S. Saxena, “Crowd behavior recognition for video surveillance,” in Int. Conf. Adv. Concepts for Intelligent Vision Systems, Berlin, 2008, pp. 970–981.

[5] T. Ko, “A survey on behavior analysis in video surveillance for homeland security applications,” in 2008 37th IEEE Applied Imagery Pattern Recognition Workshop, 2008, pp. 1–8.

[6] Y. Wang and Y. Zou, “Fast visual object counting via example-based density estimation,” in 2016 IEEE Int. Conf. Image Processing (ICIP), 2016, pp. 3653–3657.

[7] V. Lempitsky and A. Zisserman, “Learning to count objects in images,” in Advances in Neural Information Processing Systems, 2010, p. 23.

[8] D. Onoro-Rubio and R. J. López-Sastre, “Towards perspective-free object counting with deep learning,” in European Conf. Computer Vision, Cham, 2016, pp. 615–629, doi: 10.1007/978-3-319-46478-7.

[9] G. French, M. Fisher, M. Mackiewicz, and C. Needle, “Convolutional neural networks for counting fish in fisheries surveillance video,” in Proc. British Machine Vision Conf. Workshop, BMVA Press, 2015, doi: 10.5244/c.29.mvab.7.

[10] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, 2014, pp. 580–587.

[11] S. Ren, K. He, R. B. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[12] N. Paragios and V. Ramesh, “A MRF-based approach for real-time subway monitoring,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2001), IEEE, 2001, pp. 1034–1040.

[13] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, “Privacy preserving crowd monitoring: Counting people without people models or tracking,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Anchorage, Alaska, 2008, pp. 1–7.

[14] Y. Tian, L. Sigal, H. Badino, F. D. la Torre, and Y. Liu, “Latent Gaussian mixture regression for human pose estimation,” in Asian Conf. Computer Vision, Springer, Berlin, Heidelberg, 2010, pp. 679–690.

[15] F. Min, X. Pei, X. Li, Q. Liu, and Y. Huang, “Fast crowd density estimation with convolutional neural networks,” Eng. Appl. Artif. Intell., vol. 43, pp. 81–88, 2015, doi: 10.1016/j.engappai.2015.04.006.

[16] O. Elharrouss et al., "Loss functions in deep learning: A comprehensive review," 2025. [Online]. Available: http://arxiv.org/abs/2504.04242

[17] R. Gouiaa, M. A. Akhloufi, and M. Shahbazi, "Advances in convolution neural networks based crowd counting and density estimation," Big Data Cogn. Comput., vol. 5, no. 4, p. 50, 2021.

[18] S. Zhang, H. Li, W. Kong, L. Wang, and X. Niu, "An object counting network based on hierarchical context and feature fusion," J. Vis. Commun. Image Represent., vol. 62, pp. 166–173, 2019, doi: 10.1016/j.jvcir.2019.05.003.

[19] B. Li, H. Huang, A. Zhang, P. Liu, and C. Liu, "Approaches on crowd counting and density estimation: A review," Pattern Anal. Appl., vol. 24, no. 3, pp. 853–874, 2021, doi: 10.1007/s10044-021-00959-z.

[20] M. A. Hossain, M. Hosseinzadeh, O. Chanda, and Y. Wang, "Crowd counting using scale-aware attention networks," Mar. 2019. [Online]. Available: http://arxiv.org/abs/1903.02025

[21] N. Liu, C. Zou, Y. Long, Q. Niu, L. Pan, and H. Wu, "ADCrowdNet: An attention-injective deformable convolutional network for crowd understanding," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2019, pp. 3225–3234.

[22] Y. Liu, M. Shi, Q. Zhao, and X. Wang, "Point in, box out: Beyond counting persons in crowds," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2019, pp. 6469–6478.

[23] L. Zhu, Z. Zhao, C. Lu, Y. Lin, Y. Peng, and T. Yao, "Dual path multi-scale fusion networks with attention for crowd counting," Feb. 2019. [Online]. Available: http://arxiv.org/abs/1902.01115

[24] W. Liu, M. Salzmann, and P. Fua, "Context-aware crowd counting," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2019, pp. 5099–5108. [Online]. Available: https://sites.google.com/view/weizheliu/home/

[25] Z. Ma, X. Wei, X. Hong, and Y. Gong, "Bayesian loss for crowd count estimation with point supervision," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2019, pp. 6142–6151.

[26] F. Dai, H. Liu, Y. Ma, X. Zhang, and Q. Zhao, "Dense scale network for crowd counting," in ICMR 2021 - Proc. Int. Conf. Multimedia Retrieval, ACM, 2021, pp. 64–72, doi: 10.1145/3460426.3463628.

[27] V. M. Patel and V. A. Sindagi, "Multi-level bottom-top and top-bottom feature fusion for crowd counting," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2019, pp. 1002–1012.

[28] Z. Cheng, J. Li, Q. Dai, X. Wu, J. He, and A. G. Hauptmann, "Improving the learning of multi-column convolutional neural network for crowd counting," in Proc. 27th ACM Int. Conf. Multimedia, 2019, pp. 1897–1906.

[29] P. Thanasutives, K. Fukui, M. Numao, and B. Kijsirikul, "Encoder-decoder based convolutional neural networks with multi-scale-aware modules for crowd counting," in 2020 25th Int. Conf. Pattern Recognit. (ICPR), IEEE, Jan. 2021, pp. 2382–2389, doi: 10.1109/ICPR48806.2021.9413286.

[30] L. Liu, J. Chen, H. Wu, T. Chen, G. Li, and L. Lin, "Efficient crowd counting via structured knowledge transfer," in Proc. 28th ACM Int. Conf. Multimedia, 2020, pp. 2645–2654. [Online]. Available: http://arxiv.org/abs/2003.10120

[31] J. J. Cheng, Z. Chen, X. Zhang, Y. Li, and X. Jing, "Exploit the potential of multi-column architecture for crowd counting," 2020, pp. 1–9.

[32] X. Ding, F. He, Z. Lin, Y. Wang, H. Guo, and Y. Huang, "Crowd density estimation using fusion of multi-layer features," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 8, pp. 4776–4787, 2020, doi: 10.1109/TITS.2020.2983475.

[33] Y. Wang, W. Zhang, Y. Liu, and J. Zhu, "Multi-density map fusion network for crowd counting," Neurocomputing, vol. 397, pp. 31–38, 2020, doi: 10.1016/j.neucom.2020.02.010.

[34] L. Dong, H. Zhang, Y. Ji, and Y. Ding, "Crowd counting by using multi-level density-based spatial information: A multi-scale CNN framework," Inf. Sci. (N. Y.), vol. 528, pp. 79–91, 2020, doi: 10.1016/j.ins.2020.04.001.

[35] Z. Huo, B. I. N. Lu, A. Mi, F. E. N. Luo, and Y. Qiao, "Learning multi-level features to improve crowd counting," IEEE Access, vol. 8, pp. 211391–211400, 2020, doi: 10.1109/ACCESS.2020.3039998.

[36] M. Zhu, X. Wang, J. Tang, N. Wang, and L. Qu, "Attentive multi-stage convolutional neural network for crowd counting," Pattern Recognit. Lett., vol. 135, pp. 279–285, 2020, doi: 10.1016/j.patrec.2020.05.009.

[37] Y. Wang, W. Zhang, Y. Liu, and J. Zhu, "Two-branch fusion network with attention map for crowd counting," Neurocomputing, vol. 411, pp. 1–8, 2020, doi: 10.1016/j.neucom.2020.06.034.

[38] C. Wang et al., "Uniformity in heterogeneity: Diving deep into count interval partition for crowd counting," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2021, pp. 3234–3242.

[39] Y. Tian, X. Chu, and H. Wang, "CCTrans: Simplifying and improving crowd counting with transformer," 2021. [Online]. Available: http://arxiv.org/abs/2109.14483

[40] U. Sajid, W. Ma, and G. Wang, "Multi-resolution fusion and multi-scale input priors based crowd counting," in 2020 25th Int. Conf. Pattern Recognit. (ICPR), 2020, pp. 5790–5797. [Online]. Available: http://arxiv.org/abs/2010.01664

[41] M. Tian, H. Guo, and C. Long, "Multi-level attentive convolutional neural network for crowd counting," 2021. [Online]. Available: http://arxiv.org/abs/2105.11422

[42] X. Zeng, Q. Guo, H. Duan, and Y. Wu, "Multi-level features extraction network with gating mechanism for crowd counting," IET Image Process., vol. 15, no. 14, pp. 3534–3542, 2021.

[43] G. Chen and P. Guo, "Enhanced information fusion network for crowd counting," Jan. 2021. [Online]. Available: http://arxiv.org/abs/2101.04279

[44] B. Zhang, N. Wang, Z. Zhao, A. Abraham, and H. Liu, "Crowd counting based on attention-guided multi-scale fusion networks," Neurocomputing, vol. 451, pp. 12–24, 2021, doi: 10.1016/j.neucom.2021.04.045.

[45] F. Zhu, H. Yan, X. Chen, T. Li, and Z. Zhang, "A multi-scale and multi-level feature aggregation network," Neurocomputing, vol. 423, pp. 46–56, 2020, doi: 10.1016/j.neucom.2020.09.059.

[46] Y. Xia, Y. He, S. Peng, Q. Yang, and B. Yin, "CFFNet: Coordinated feature fusion network for crowd counting," Image Vis. Comput., vol. 112, p. 104242, 2021, doi: 10.1016/j.imavis.2021.104242.

[47] S. D. Khan, Y. Salih, B. Zafar, and A. Noorwali, "A deep-fusion network for crowd counting in high-density crowded scenes," Int. J. Comput. Intell. Syst., vol. 14, no. 1, p. 168, Dec. 2021, doi: 10.1007/s44196-021-00016-x.

[48] Y. Ma, "Inception-based crowd counting – being fast while remaining accurate," 2022. [Online]. Available: http://arxiv.org/abs/2210.09796

[49] Y. Ma, V. Sanchez, and T. Guha, "Fusioncount: Efficient crowd counting via multiscale feature fusion," in 2022 IEEE Int. Conf. Image Processing (ICIP), IEEE, 2022, pp. 3256–3260.

[50] M. Wang, H. Cai, X. Han, J. Zhou, and M. Gong, "STNet: Scale tree network with multi-level auxiliator for crowd counting," IEEE Trans. Multimedia, vol. 25, pp. 2074–2084, 2022.

[51] H. Lin, Z. Ma, R. Ji, Y. Wang, and X. Hong, "Boosting crowd counting via multifaceted attention," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022, pp. 19628–19637.

[52] W. Shu, J. Wan, K. C. Tan, S. Kwong, and A. B. Chan, "Crowd counting in the frequency domain," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022, pp. 19618–19627.

[53] J. Gao et al., "Deep rank-consistent pyramid model for enhanced crowd counting," IEEE Trans. Neural Netw. Learn. Syst., Nov. 2023, pp. 1–13, doi: 10.1109/TNNLS.2023.3336774.

[54] Z. Du, M. Shi, J. Deng, and S. Zafeiriou, "Redesigning multi-scale neural network for crowd counting," IEEE Trans. Image Process., vol. 32, pp. 3664–3678, 2023, doi: 10.1109/TIP.2023.3289290.

[55] Z. Miao, Y. Zhang, Y. Peng, H. Peng, and B. Yin, "DTCC: Multi-level dilated convolution with transformer for weakly-supervised crowd counting," Comput. Vis. Media (Beijing), vol. 9, no. 4, pp. 859–873, 2023.

[56] J. Zhang, L. Ye, J. Wu, D. Sun, and C. Wu, "A fusion-based dense crowd counting method for multi-imaging systems," Int. J. Intell. Syst., vol. 2023, no. 1, p. 6677622, 2023.

[57] X. Guo et al., "Crowd counting via attention and multi-feature fused network," Human-centric Comput. Inf. Sci., vol. 13, no. Nov., 2023.

[58] Y. Chaudhuri, A. Kumar, O. C. Phukan, and A. B. Buduru, "A lightweight feature fusion architecture for resource-constrained crowd counting," 2024. [Online]. Available: http://arxiv.org/abs/2401.05968.

[59] Y. Yin and B. Yin, "Cross-level feature relocation: Mitigating information loss in cross-layer feature fusion for crowd counting," in Proc. 16th Asian Conf. Mach. Learn., 2024.

[60] L. Chen et al., "The effectiveness of a simplified model structure for crowd counting," 2024. [Online]. Available: http://arxiv.org/abs/2404.07847.

[61] H. Ma, L. Zhang, and S. Shi, "VMambaCC: A visual state space model for crowd counting," 2024.

[62] Y. Zhang, W. Song, M. Shao, and X. Liu, "MRSNet: Multi-resolution scale feature fusion-based universal density counting network," Sensors, vol. 24, no. 18, p. 5974, 2024.

[63] H.-Y. Ma, L. Zhang, and X.-Y. Wei, "FGENet: Fine-grained extraction network for congested crowd counting," in Int. Conf. Multimedia Modeling, 2024, pp. 43–56. [Online]. Available: http://arxiv.org/abs/2401.01208

[64] J. Yue, J. Cheng, W. Wu, and X. Tang, "FGEFNet: Fine-grained extraction and flow network for crowd counting," 2024, doi: 10.21203/rs.3.rs-4607436/v1.

[65] S. Goel and D. Koundal, "A MaskFormer EfficientNet instance segmentation approach for crowd counting," Sci. Rep., vol. 15, no. 1, p. 13275, 2025.

[66] S. Jiang et al., "ProgRoCC: A progressive approach to rough crowd counting," 2025. [Online]. Available: http://arxiv.org/abs/2504.13405

[67] J. Yu and H. Hu, "Multiscale regional calibration network for crowd counting," Sci. Rep., vol. 15, no. 1, p. 2866, 2025.

[68] P. Liu, H. Li, S. Lei, N. Liu, B. Feng, and X. Wu, "RCCFormer: A robust crowd counting network based on transformer," Apr. 2025. [Online]. Available: http://arxiv.org/abs/2504.04935.

Appendix A

Table A1: Using EFC2ADL proposed model for evaluation crowd counting studies.

Author	Dataset No.					Dataset Type					The Proposed Artifact	Loss Function Type					Evaluation Metrics Used			Outcome					
																				Shanghai-Tech Part A			Shanghai-Tech Part B		
	1	2	3	4	More than 4	Shanghai-Tech	UCF CC 50	UCSD	UCFQRNF	Other Datasets		DM-Count loss	Mean Squared Error (MSE)	Euclidean loss	Cross Entropy (CE)loss	Other Loss Function	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
[20]			√			√	√			√	SAAN					√	√	√					16.86	28.41	
[21]				√		√	√	√		√	ADCrowdNet					√	√	√		70.9	115.2		7.7	12.9	
[22]				√		√	√				PSDDN					√	√	√		65.9	112.3		9.1	14.2	
[23]				√		√	√	√	√		SFANet					√	√	√		59.8	99.3		6.9	10.9	
[24]				√		√	√		√	√	ECAN					√	√	√		62.3	100.0		7.8	12.2	
[25]			√			√	√		√		BAYESIAN+					√	√	√		62.8	101.8		7.7	12.7	
[26]				√		√	√	√	√		DSNet			√		√	√		√	61.3		97.3	6.7		10.5
[27]			√			√	√		√		MBTTBF			√			√	√		60.2	94.1		8.0	15.5	
[18]				√		√	√			√	HFNet			√			√	√		81.1	123.4		16.8	31.3	
[28]				√		√	√	√		√	McML					√	√	√		63.8	110.5		10.1	13.9	
[29]					√	√	√		√	√	*M-SFANet *M-SegNet					√	√	√		57.55	94.48		6.32	10.06	
[30]			√			√			√	√	SKT					√	√		√	62.73		102.33	7.98		13.13
[31]				√		√	√		√	√	PSNet			√		√	√		√	55.5		90.1	6.8		10.7
[32]					√	√	√			√	CFFNet			√			√	√		69.8	114.7		10.2	14.9	
[33]				√		√	√	√		√	MDMF			√			√	√		64.1	105.6		7.7	12.9	
[34]				√		√	√	√	√		MM- Net					√	√	√		60.8	99.0		7.6	11.7	
[35]			√			√	√	√			FFANet		√				√	√		62.4	102.6		8.3	11.1	
[36]			√			√	√			√	AMCNN			√	√		√	√		76.1	110.7		15.3	27.4	
[37]				√		√	√		√	√	None			√		√	√	√		57.7	99.7		7.4	11.1	
[38]				√		√	√		√	√	UEPNet		√		√		√	√		54.64	91.15		6.38	10.88	
[39]				√		√	√		√	√	CCTrans					√	√	√		64.4	95.4		7.0	11.5	
[40]		√				√			√		None		√				√		√	67.1		81.0			
[41]			√			√	√		√		MLAttnCNN			√			√	√					7.5	11.6	
[42]				√		√	√		√	√	MFEN			√			√		√	58.0		95.8	6.6		11.1
[43]			√			√	√		√		IFM		√			√	√	√		61.1	111.7		7.6	12.1	
[44]					√	√	√		√	√	AMS-Net			√	√		√	√		63.8	108.5		7.3	11.8	
[45]				√		√	√		√	√	MFANet					√	√		√	58.5		98.4	7.2		11.6

Author	Dataset No.					Dataset Type					The Proposed Artifact	Loss Function Type					Evaluation Metrics Used			Outcome					
																				Shanghai-Tech Part A			Shanghai-Tech Part B		
	1	2	3	4	More than 4	Shanghai-Tech	UCF CC 50	UCSD	UCFQRNF	Other Datasets		DM-Count loss	Mean Squared Error (MSE)	Euclidean loss	Cross Entropy (CE)loss	Other Loss Function	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
[46]				√		√			√	√	CFFNet					√	√	√		53.4	93.5		6.6	10.6	
[47]			√			√	√		√		None			√			√	√		77.58	129.7		14.1	21.10	
[48]		√				√				√	ICC	√					√		√	76.97		130.16	8.46		15.20
[49]	√					√					FusionCount	√					√		√	62.2		101.2	6.9		11.8
[50]			√			√	√		√		STNet				√	√	√	√		52.85	83.64		6.25	10.30	
[51]				√		√			√	√	AMCNN		√			√	√	√		56.8	90.3				
[52]				√		√			√	√	None					√	√	√		57.5	94.3		6.9	11.0	
[53]				√		√	√		√	√	DREAM					√	√		√	62.6		102.0	7.9		13.4
[54]					√	√	√			√	HMoDE					√	√		√	54.4		87.4	6.2		9.8
[55]				√		√	√		√	√	DTCC-Dynamic					√	√	√		60.8	97.0		7.2	10.8	
[56]		√				√	√				AMNet			√			√	√		113.0	71.7		17.0	11.0	
[57]					√	√	√		√	√	AMFNet		√				√		√	66.8		107.6	7.7		12.2
[58]		√				√	√				ASFNet					√	√	√		59.32	88.67		8.2	11.02	
[59]				√		√	√			√	CFRM			√	√		√	√		51.1	85.2		6.4	10.9	
[60]			√			√	√			√	FFNet	√					√	√		48.3	80.5		6.1	9.0	
[61]				√		√	√		√	√	VMambaCC					√	√	√		51.87	81.3		7.48	12.47	
[62]				√		√			√	√	MRSNet					√	√	√		54.2	88.5		6.3	9.7	
[63]			√			√	√		√		FGENet					√	√	√		51.66	85		6.34	10.53	
[64]			√			√	√			√	FGEFNet			√			√	√		49.1	77.6		6.9	11.3	
[65]			√			√			√	√	MFEFNet					√	√		√	24.18		98.1	6.5		10.4
[66]			√			√			√	√	ProgRoCC			√			√	√		70.0	112.6				
[67]			√			√	√		√		MRCNet					√	√	√		55.8	96.9		6.7	10.4	
[68]			√			√			√	√	RCCFormer	√					√	√		48.3	72.1		6.6	10.4	