International Journal of Contemporary Research in Multidisciplinary; 2023; 2(6): 16-23

Research Article

A Comprehensive Literature Survey for Crowd Scene Analysis techniques

Author (s): **Faisel G. Mohammed[1], Abbas F. Nori[2], Noor N. Thamer[3]**

[1] *Department of Remote Sensing & GIS, Baghdad, Collage of Science College, University of Baghdad, Iraq*

[2]*Department of Physics, Faculty of Science, University of Kufa, Najaf, Iraq*

[3]*Ministry of Education, General Directorate of Education, Baghdad, Al-Rusafa II / Education Department of the outskirts of eastern Baghdad, Iraq*

Corresponding Author: * Abbas F. Nori

Abstract

Understanding how people behave in crowded places is an important endeavor with several uses, like controlling the spread of COVID-19 or other diseases that spread through contact. An in-depth study of crowd scene analysis methods, including both crowd counting and crowd activity detection, is included in this survey article. This article fills the gap by exhaustively examining the spectrum up to contemporary deep learning techniques, whereas current studies frequently focus primarily on certain aspects or traditional approaches. The paper proposes the innovative idea of Crowd Divergence (CD) evaluation as a matrix for evaluating crowd scene analysis approaches, which was motivated by information theory. Contrary to conventional measurements, CD quantifies the agreement between expected and observed crowd count distributions. This paper makes three key contributions: an examination of readily available crowd scene datasets, the use of CD for thorough technique evaluation, and a thorough examination of crowd scene methodologies. The investigation starts with conventional computer vision methods, closely examining density estimates, detection, and regression strategies. Convolutional neural networks (CNNs) become effective tools as deep learning progresses, as seen by new models like ADCrowdNet and PDANet, which make use of attention mechanisms and structured feature representation. To evaluate algorithmic effectiveness, a variety of benchmark datasets, including ShanghaiTech, UCF CC 50, and UCSD, are carefully examined. Computer vision's exciting and challenging topic of "crowd scene analysis" has numerous applications, from crowd control to security surveillance. This survey article offers a comprehensive viewpoint on crowd scene analysis, bringing several approaches under a single heading and presenting the CD measure to guarantee reliable assessment. This article provides a complete resource for researchers and practitioners alike through an elaborate investigation of methods, datasets, and cutting-edge evaluation approaches, paving the way for improved crowd scene analysis techniques across a variety of fields.

Manuscript Information

- **ISSN No:** 2583-7397
- **Received:** 27-09-2023
- **Accepted:** 13-11-2023
- **Published:** 17-11-2023
- **IJCRM:**2(6);2023:16-23
- **©2023, All rights reserved**
- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Manuscript

Faisel G. Mohammed, Abbas F. Nori, Noor N. Thamer. **A Comprehensive Literature Survey for Crowd Scene Analysis techniques**. International Journal of Contemporary Research in Multidisciplinary. 2023; 2(6):16-23.

Keyword: Crowd behavior analysis, Crowd scene methodologies, Crowd Divergence (CD) evaluation, Deep learning techniques, Benchmark datasets, Crowd control and security

1. Introduction:

The study of crowd scene analysis involves examining the behavior of people in groups in the same physical space [1]. It usually entails counting the number of people, in regions tracking their movements and identifying their behaviors. This type of analysis has applications. One such application is controlling the spread of COVID 19 by guaranteeing separation in public spaces such as malls and parks [2]. Additionally, it helps to maintain security on Muslim pilgrimages, carnivals, New Year's Eve celebrations, and sporting events. [3, 4]. Surveillance camera systems may detect behaviors within groups of individuals using automatic crowd scene analysis [5]. Additionally analyzing crowd scenes in places like train stations, supermarkets and shopping malls can provide insights, into crowd movement patterns. Identify design flaws. These studies contribute to improving safety considerations [6, 7].

As was previously shown, it is crucial to analyze crowd scenes, consequently, a number of survey articles have been proposed. The survey articles that are now available either mandate the application of conventional computer vision techniques for the examination of crowd situations or focus on a particular facet of crowd analysis, as in crowd counts. [8,9]. This overview article seeks to offer a thorough examination from the development of tools for crowd scene analysis to the latest advancements in deep learning [10,11]. Crowd activity recognition and crowd counting are the two main elements of crowd analysis, which are both included in this survey.

Crowd scene analysis is when we look at how many people are in a certain area. This is important for things like surveillance, planning cities, and keeping the public safe. This survey explores different ways of counting crowds, from traditional techniques to the more advanced method called deep learning. This text tells us about different ways to figure out how many people live in different places. It talks about the good and bad things about these methods and what they can teach us [12,13].

The survey starts by analyzing traditional ways of seeing and understanding things using computers. Three categories can be used to group these techniques: techniques for estimating density, methods that detect, and methods that predict. Density estimation methods help create density maps, which make it easier to see where the crowd is concentrated and reduce mistakes in counting. Detection-based techniques work by finding certain body parts like heads or shoulders. Regression-based methods, on the other hand, use regression models to convert basic features into counts. The advantages and disadvantages of each method are carefully investigated. The survey introduces a new method called deep learning. It focuses on a type of technology called convolutional neural networks. This change has made crowd analysis better by allowing us to count people faster and more accurately, even in complicated situations. This text talks about the importance of models like ADCrowdNet, PDANet, and DSSINet. These models show us how hierarchical structures, attention mechanisms, and generative adversarial networks (GANs) can be useful in counting crowds. Datasets are very important in influencing progress in research. UCSD, WorldExpo'10, UCF_CC_50, ShanghaiTech Part A and B, and other well-known datasets are all examined in this study. It shows different crowd sizes, types of scenes, and difficulties. These datasets show that there is a need for research methods that can handle the complex and varied situations we encounter in real-life crowds.

2. Crowd Counting

Crowd counting refers to determining how many people reside in a specific area. The subsections go through various approaches to estimating the population density of a given geographic area. To be thorough, we first discuss conventional computer vision techniques before reviewing deep learning-based techniques. A block diagram for crowd counting with deep learning has several blocks that each represent a part of the system. Graphical tools are used to graphically illustrate this structure in real-world circumstances.

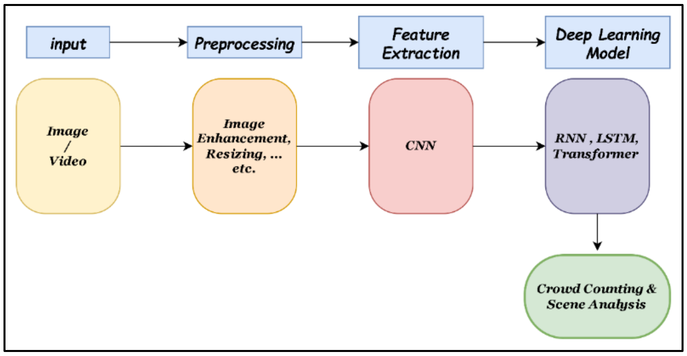


Figure 1: General Block Diagram of Crowd Counting

The above diagram starts by using an input (an image or video), then does feature extraction and pre-processing, which includes scaling and normalizing the picture. To identify different characteristics in the picture, the Convolutional Neural Network (CNN) is employed. The collected characteristics are then examined for crowd counting using a Deep Learning Model (RNN/LSTM/Transformer). The crowd count and potential extra scene analysis findings are represented in the output. This procedure makes sure that data analysis is reliable and effective.

2.1. Conventional Techniques for Computer Vision

Through the linear mapping of local patch features to their respective objects, the author constructed density maps. In situations when there are several things, this problem formulation minimizes the difficulty of dissecting each object for counting purposes and the potential for counting mistakes. To count the objects in terms of the number in the image, this method integrates local batches throughout [14].

The loss function was relied upon to create the density map, where it (maximizes the regularized risk quadratic cost function) [15]. Cutting-plane optimization was used to complete the solution [16]. The work in was improved by Pham et al. in by learning nonlinear mapping. They voted on the density of several target items using random forest regression [17, 18]. They also achieved performance in real time, and an alternative to mapping dense features and creating a density map computed the embedding of subspaces created by image patches.

An approach for density estimation that is scale- and resolution-invariant was suggested by Sirmacek et al. In order to ascertain probability density functions (pdfs) [19] of various places in successive frames, this technique uses Gaussian symmetric kernel functions [20]. The number of people per spot is then estimated using the value of the generated PDFS. The three primary categories of the conventional crowd-counting approach are listed in the Table 1.

There are many researchers used Detection-Based Approaches to study crowded sence analysis like, Early methods, like those in [21], relied on detectors to locate heads or shoulders to numbering persons in a crowd. Typically, calculating using detection uses either parts-based or monolithic detection. Pedestrian detection methods for monolithic detection, such as optical flow [22], oriented gradient histogram (HOG), Typically, the detection is based on particle movement, shapelets, edgelets, and Haar wavelets. Nonlinear classifiers like the Support Vector Machine (SVM) are then fed the characteristics collected from the earlier detectors, although at a slow rate. Typically, linear SVM, via forests, or boosting are linear classifiers that provide a compromise between accuracy and speed [23]. After then, it moved the classifier around the whole image to identify prospects and exclude those who are less certain. The sliding results show the number of people in attendance.

When the partial occlusion problem [24] arises, the earlier approaches are unable to handle it; as a result, part-based detection is used. Instead of focusing on the entire body, Similar to the head and shoulders, specific body parts are the subject of part-based detection. It is believed that part-based detection is more trustworthy than monolithic detection. Humans were modelled using ellipsoids based on 3D shapes, and a stochastic approach was used to determine the arrangement of numbers and shapes that most clearly describes a divided foreground item. Later, Ge et al expanded the same concept using a Bernoulli form prototype and a Bayesian marked point process (MPP) [25]. The Markov chain Monte Carlo was utilized by Zhao et al. To benefit from temporal coherence in 3D human models between successive frames [26].

Even when part-based or detection-based counting yields findings that are satisfactory, they fall short in densely populated areas and where there is significant occlusion. Regression counting tries to address the prior issues. This approach typically consists of two key parts. The first part of the process is obtaining low-level features, including gradient, edge, texture, and foreground features. The second step entails applying a regression function, such as linear, piecewise, ridge, or Gaussian process regression, to transform the gathered features into counts, as in figure 1 depicts this method's whole workflow [27].

A multi-feature by York et al, the technique was proposed for accurate crowd counting. They combined many features, such as head locations, SIFT interest points, Fourier interest points, uneven and nonhomogeneous texture, into one overall feature descriptor. Then, a multi-scale Markov Random Field (MRF) was utilized using this global descriptor to estimate counts,. Also, the authors included a brand-new dataset (UCF-CC-50). Regression-based techniques usually yield respectable outcomes, but because they rely on a global count, they lack spatial information. [28,29]

2.2. The Deep Learning Methodologies

Similar to neural networks (NNs), convolutional neural networks (CNNs) include learnable weights and biases, as well as neurons and receptive fields. The output of every responsive field's convolution operation is supplied to a function of nonlinearity when it receives a batch of inputs. (ReLU, Sigmoid, etc.). Since CNN assumes that the input image is RGB, rich information is acquired by the hidden layers, improving the overall network's performance (classifier and hidden layers together). This structure provides advantages in terms of accuracy and speed because there are multiple elements to identify in the photographs of the crowd scene. Networks that receive an input image and immediately produce the desired output are known as end-to-end networks [30]. Some of the articles that examined crowd scene analysis with deep learning approaches are listed below: -

1. The study offers a review of the literature on crowd analysis using deep learning methods in the context of intelligent video surveillance. The study focuses on numerous crowd analysis topics, such as crowd analysis, object recognition, action recognition, and violence detection. The study analyzes several deep learning techniques regarding their models and algorithms because most of the publications it reviews are based on deep learning methods. The report also touches on the technological implementation of various crowd video analysis techniques and emphasizes the demand for real-time processing in this field [31].

2. For crowd scene analysis, the research suggests a Compressed Sensing Output Encoding (CSOE) strategy that enhances localisation performance in densely populated situations without significant scale variation. Multiple Dilated Convolution Branches (MDCB), which provide various receptive field sizes, are introduced in this research to help localization accuracy when object sizes vary significantly in a picture. The research also introduces an Adaptive Receptive Field Weighting (ARFW) module that emphasizes informative channels with the right receptive field size to address the scale variation problem. Experiments show that the suggested technique, especially in densely populated settings, reaches cutting-edge performance on four popular datasets. The study emphasizes the significance of addressing target size variation in crowd analysis and contends that crowd localization can be effectively achieved by modelling it as regression in encoding signal space for crowd analysis [32].

3. A literature review of profound learning-based approaches to analyzing crowded environments, with a focus on crowd measurement and identification of crowd behavior, is presented in the publication: "Deep Learning-Based Crowd Scene Analysis Survey". The study examines crowd scene datasets as well, which are crucial for training and assessing

crowd scene analysis techniques. The research also suggests an appraisal score for methods of crowd scene analysis. This statistic calculates the contrast between crowd counts that are calculated and those that are observed in crowd scene videos [33].

4. In high-population cities where big crowds in public spaces pose issues for safety and transit, the study focuses on machine and deep learning for crowd analytics. Existing crowd analytics techniques are problem-specific and difficult to adapt to other videos. More training examples drawn from a variety of videos are needed. The goal of the research is to examine various scene crowd analytics using conventional and deep learning models while weighing the benefits and drawbacks of each strategy. Large datasets are needed for training and evaluating deep learning models and techniques. Manual annotations and a growing variety of videos and images are used to collect datasets. The paper offers a variety of deep learning models and training methods for feature modeling in crowd analytics [34].

5. In 2018, C. Santhini1 and V. Gomathi concentrated on crowd scene analysis, a critical step in comprehending crowded scenes. Convolutional neural networks (CNNs) and deep learning models are suggested by the authors as tools for assessing crowd scenes. They want to develop maps of crowd density and estimate how many individuals are in a crowd. Estimating the crowd size in very dense groups is important for video surveillance and anomaly warning, according to the report. The authors note the difficulties with the current approaches, including the scarcity of training samples, serious obstructions, disorderly scenes, and perspective changes. They contend that CNNs, deep learning networks, perform better in determining crowd size and density. The usage of Lucas Kanade optical flow is also mentioned in the paper for locating displacement vectors between adjacent frames and sequencing 3D volume video slices. The authors created an attribute set with 94 attributes using a neural crowd dataset made up of 100 films from 800 crowd scenarios [35].

6. In 2019, Wang Zhiyu, Yang Jiaxin, and Yang Jiaye proposed a deep learning-based scene analysis system and method. A cloud AI platform and a data-collecting subsystem make up the system. While the cloud AI platform has modules for face recognition, facial expression analysis, speech recognition, voice analysis, and complete analysis, the data acquisition subsystem collects photos and audio. The system makes use of deep learning technology to deliver precise recognition results for facial expressions, voice semantics, and intonations. The technique enables simultaneous speech and facial identification, guaranteeing speed and accuracy of findings. This method considerably enhances scene analysis technology [36].

7. P. Mahesha et al. discuss applying deep learning models to criminal justice scene investigation in 2021, particularly in terms of producing phrases that describe the crime scene based on photographs. Since there isn't a recognized dataset accessible specifically for crime scene photos, the suggested method involves training the (MSCOCO), are models on a large dataset. The models get 9 separate segments of the crime scene photos and process each segment individually to produce 9 phrases that describe the scene of the crime. The study mentions three deep learning models—the Inceptionv3-LSTM network, the VGG-16-LSTM network, and the ResNet-50-LSTM network—that are utilized to generate sentences. The BLEU scores of these models are 0.1771, 0.11, and 0.1784, in that order. With a 14.8% vote difference, the Inceptionv3-LSTM model was favored by customers over the ResNet-50-LSTM model. [37].

8. In 2021, R. Abinaya et al. concentrated on the application using convolutional neural networks (CNNs) for deep learning classifier for categorizing ambient event sounds based on extracted MFCC features. The experiment's findings demonstrate that, with a high classification accuracy of 90.65%, the proposed MFCC-CNN outperforms other existing methods. The cornerstone for classifying acoustic data is feature extraction, according to the literature review section. Based on temporal resolution, it divides feature extraction into three subcategories: frame-level features, segment-level features, and texture windows. Local characteristics are described by frame-level features, which are generated from short analysis frames/windows with sample sizes ranging from 10ms to 100ms. Examples include spectral, cepstral, and temporal features including roll-off, flatness, ZCR, time energy, LPC, LPCC, and MFCC. features at the segment level Compared to frame-level features, these features have larger analysis windows and capture the sound's textural qualities [38].

9. Using deep learning methods, Xiaogang Wang and Chen Change Loyin 2017, proposed scene-independent crowd analysis. It draws attention to the shortcomings of scene-specific previous efforts and suggests the use of general deep models that can be applied to a variety of crowd settings without the need for retraining. Numerous crowd analysis topics are covered in the study, includes an estimate of crowd density, crowd attribute recognition, and crowd counting. It introduces some freshly created large-scale datasets and highlights how crucial large-scale training sets are for advancing deep learning. The difficulties of annotating crowd datasets and boosting scene diversity are also covered in the paper. It offers various deep neural network topologies and instruction methods for acquiring feature representations for crowd analysis [39].

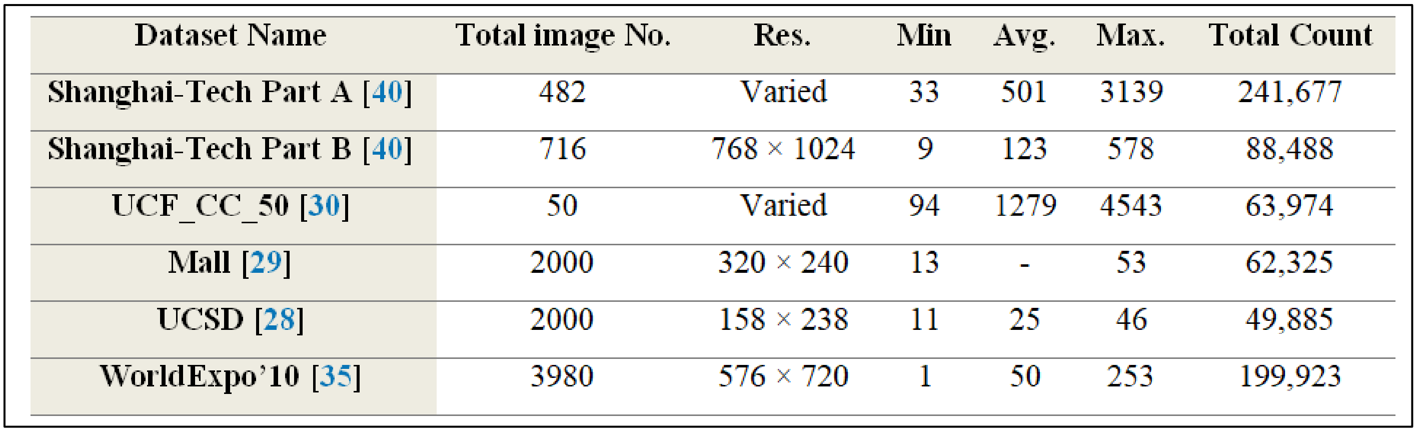
Table 1: Crowded Scene Analysis by Deep Learning Comparison Table

REF.	Conclusions	Results	Methods Used	Limitations	Contributions	Practical Implications
[31]	Techniques for analyzing intelligent surveillance video were reviewed - Crowd analysis is difficult because of how big and shifting the crowd is.	**N/A**	Deep Learning techniques - alternatives to deep learning (not specified)	Understanding the underlying concepts can be difficult with deep learning models. - Limited comprehension of video crowd behavior.	Crowd analysis using deep learning techniques - Using deep learning in surveillance videos	Provides methods for deep learning to analyze crowds. focuses on difficulties with real- time processing and abnormal event identification in congested settings
[32]	The CSOE technique enhances crowd localization in densely populated environments Accurate localization is improved with MDCB and ARFW modules.	Reaches cutting-edge performance across four popular datasets - Excellent outcomes in busy environments	Multiple Dilated Convolution Branches (CSOE) method for Compressed Sensing-based Output Encoding (MDCB)	**N/A**	Developed Multiple Dilated Convolution Branches (MDCB) and Adaptive Receptive Field Weighting (ARFW) module for Compressed Sensing based Output Encoding (CSOE) scheme.	Improves accuracy in crowd analysis with varied item sizes - Boosts localization performance in densely populated scenes
[34]	We evaluate deep learning-based approaches to crowd counting and crowd action detection. - There is a suggested assessment metric for crowd scene analysis techniques.	An assessment metric for crowd scene analysis methods is proposed in a survey of deep learning-based techniques.	Crowd tally Recognition of crowd behavior	Heavy occlusion, intricate activities, and altered posture- **N/A**	An assessment metric for crowd scene analysis algorithms is proposed in a survey of deep learning-based methods.	Automatic crowd management, counting, security, and tracking - Crowd scene analysis method evaluation metric.
[36]	system for analyzing scenes using deep learning. Recognition outcomes are quicker and more precise.	Face expressions have an impact on recognition Results of voice semantic and intonation recognition	Face recognition technique using deep learning. Deep learning voice recognition technology	**N/A**	system for analyzing scenes using deep learning. Recognition of voice semantics, intonations, and facial expressions	improved video surveillance crowd counts. Improved assessment of population density in crowded situations
[37]	**N/A**	Identification of rehearsing issues quickly and accurately Provides additional tools and data support for on- site commanding	Using deep learning to find errors. System for creating simulation data during practice	**N/A**	Method for detecting errors using deep learning. method for simulating practice and producing simulation data	Quick and accurate detection of rehearsing issues. provides additional tools and information to support on-site commanding
[38]	Analysis of crowd scenes using deep learning networks is useful. 94 attribute convolutional crowd dataset is utilized	Crowd scene analysis with a deep learning model using a convolutional crowd dataset with 94 attributes	Deep learning model using convolutional neural networks (CNN)	Absence of training samples, significant obstacles, chaotic surroundings, and perspective shift	Convolution neural networks are introduced for crowd scene interpretation, and crowd counting, and density map estimation are suggested	-improved video surveillance crowd counts. Improved assessment of population density in crowded situations
[39]	Three deep learning algorithms for creating sentences are suggested ResNet-50- LSTM model is favored over Inceptionv3-LSTM model	ResNet-50-LSTM model is chosen above Inceptionv3-LSTM model according to BLEU scores of 0.1771, 0.11, and 0.1784.	ResNet-50-LSTM network, Inceptionv3-LSTM network, and VGG-16-LSTM network	No known dataset for training on crime scene photos is available, and it is challenging to capture even the smallest aspects of crime scenes.	Deep learning models used for crime scene analysis segmenting crime scene photos and training models on large dataset	aids in the analysis of crime scene photographs produces complex sentences regarding crime scenes

3. Crowd Scene Datasets

Crowd scene algorithms can be trained and tested on a variety of datasets, as indicated in Table 2. The ShanghaiTec dataset is the most used, particularly in deep learning algorithms [40]. There are 1198 photos with annotations, including street view images and internet images. The 108 security cameras that were watching Shanghai WorldExpo 2010 produced the WorldExpo'10 dataset [35]. There are 1132 annotated video clips in this collection. There are 50 annotated crowd frames in the UCF dataset _CC 50 [27]. Due to the wide variation in crowd sizes and scenario types, this dataset is one of the most difficult to analyze. The crowd size typically ranges from 94 to 4543 people. 2000 annotated photos with a dimension of 158 by 238 pixels each make up the UCSD dataset [25]. The maximum number of persons is 46, and each object has a label in the middle that indicates the ground truth. There are varied densities in the mall [26]. There are also a variety of static and dynamic activity patterns.

Table 2: Datasets specifications



Dataset Name	Total image No.	Res.	Min	Avg.	Max.	Total Count
Shanghai-Tech Part A [40]	482	Varied	33	501	3139	241,677
Shanghai-Tech Part B [40]	716	768 × 1024	9	123	578	88,488
UCF_CC_50 [30]	50	Varied	94	1279	4543	63,974
Mall [29]	2000	320 × 240	13	-	53	62,325
UCSD [28]	2000	158 × 238	11	25	46	49,885
WorldExpo'10 [35]	3980	576 × 720	1	50	253	199,923

4. Discussion

Crowd scene analysis is a crucial field in public safety, surveillance, urban planning, and event management. It has been instrumental in maintaining social distancing during the COVID-19 pandemic, ensuring security during mass events, and enhancing public safety in crowded spaces. The survey explores How crowd scene analysis has evolved techniques, starting with traditional computer vision methods and transitioning to deep learning methodologies. Traditional methods, such as density estimation, detection-based approaches, and regression-based methods, have limitations in densely populated scenarios. However, the accuracy of activity recognition and crowd counting have been transformed since the advent of Convolutional Neural Networks (CNNs). The survey emphasizes the need for continued research and development of new methods to handle complex situations in real-life crowd scenes. As urbanization and large crowd events become more common, the demand for effective crowd analysis techniques will increase.

5. Conclusion

Crowd scene analysis is a crucial field with applications in public safety, surveillance, urban planning, and event management. Traditional computer vision methods, such as density maps and regression-based methods, struggle in densely populated areas. Deep learning approaches, particularly convolutional neural networks, have revolutionized crowd scene analysis, showing exceptional performance in crowd counting and activity recognition. Crowd scene datasets, such as ShanghaiTech, WorldExpo'10, UCF_CC_50, and UCSD, are essential for training and evaluating crowd analysis algorithms.

6. Future Directions

This survey provides an overview of crowd scene analysis techniques, there are a number of interesting avenues for further study. The first is to improve the robustness of crowd analysis methods, particularly in handling complex scenarios with severe occlusion, variable lighting conditions, and diverse crowd behaviors. The second is to optimize deep learning models for real-time processing, ensuring accuracy without sacrificing accuracy. The third is to explore multimodal analysis, combining information from various sensors to provide a holistic understanding of crowd scenes. The fourth is to address privacy and ethical concerns, exploring techniques for anonymizing data and ensuring the responsible use of crowd analysis tools. The fifth is to explore human-AI collaboration in crowd management and security, integrating AI insights with human judgment.

References

1. Musse, S.R.; Thalmann, D. A model of human crowd behavior: Group inter-relationship and collision detection analysis. In Computer Animation and Simulation'97; Springer: Berlin/Heidelberg, Germany, 1997; pp. 39–51. [Springer],[Google Scholar]
2. Watkins, J. Preventing a Covid-19 Pandemic. 2020. Available online: https://www.bmj.com/content/368/bmj.m810.full (accessed on 8 May 2012).
3. Jarvis, N.; Blank, C. The importance of tourism motivations among sport event volunteers at the 2007 world artistic gymnastics championships, stuttgart, germany. J. Sport Tour. 2011, 16, 129–147. [Taylor & Francis], [Academia]
4. Drury J, DaMatta R. Carnivals, rogues, and heroes: An interpretation of the Brazilian dilemma. University of Notre Dame Press; 2020. [Google Scholar]

5. Winter, T. Landscape, memory and heritage: New year celebrations at angkor, cambodia. Curr. Issues Tour. 2004, 7, 330–345. [Taylor & Francis],[Research Gate]
6. Peters, F.E. The Hajj: The Muslim Pilgrimage to Mecca and the Holy Places; Princeton University Press: Princeton, NJ, USA, 1996. [Google Books]
7. Cui, X.; Liu, Q.; Gao, M.; Metaxas, D.N. Abnormal detection using interaction energy potentials. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20 June 2011; pp. 3161–3167.
8. Mehran R, Moore BE, Shah M. A streakline representation of flow in crowded scenes. InComputer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11 2010 (pp. 439-452). Springer Berlin Heidelberg. [Springer], [Google Scholar]
9. Ihaddadene N, Djeraba C. Motion Pattern Extraction and Event Detection for Automatic Visual Surveillance. EURASIP Journal on Image and Video Processing. 2011;2011(1):163682. [Google Scholar]
10. Chow, W.K.; Ng, C.M. Waiting time in emergency evacuation of crowded public transport terminals. Saf. Sci. 2008, 46, 844–857. [Elsevier]
11. Sindagi, V.A.; Patel, V.M. A survey of recent advances in cnn-based single image crowd counting and density estimation. Pattern Recognit. Lett. 2018, 107, 3–16. [Elsevier], [Google Scholar]
12. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [Nature]
13. Wang, Z.; Bovik, A.C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. IEEE Signal Process. Mag. 2009, 26, 98–117. [Google Scholar]
14. Goffin, J.L.; Vial, J.P. Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method. Optim. Methods Softw. 2002, 17, 805–867. [Research Gate]
15. Pham, V.Q.; Kozakaya, T.; Yamaguchi, O.; Okada, R. Count Forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 3–17 December 2015; pp. 3253–3261.[Google Scholar]
16. Liaw, A.; Wiener, M. Classification and regression by randomforest. News 2002, 2, 18–22.[Google Scholar]
17. Scaillet O. Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric statistics. 2004 Feb 1;16(1-2):217-26. [Google Scholar]
18. Cha, S.H. Comprehensive survey on distance/similarity measures between probability density functions. City 2007, 1,1.
19. Dollar P, Wojek C, Schiele B, Perona P. Pedestrian detection: An evaluation of the state of the art. IEEE transactions on pattern analysis and machine intelligence. 2011 Aug 4;34(4):743-61.[Google Scholar]
20. Li, M.; Zhang, Z.; Huang, K.; Tan, T. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL, USA, 8 December 2008; pp. 1–4.
21. Brox, T.; Bruhn, A.; Papenberg, N.; Weickert, J. High accuracy optical flow estimation based on a theory for warping. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2004; pp. 25–36.[Springer]
22. Viola, P.; Jones, M.J. Robust real-time face detection. Int. J. Comput. Vis. 2004, 57, 137–154. [Springer]
23. Kilambi P, Ribnick E, Joshi AJ, Masoud O, Papanikolopoulos N. Estimating pedestrian counts in groups. Computer Vision and Image Understanding. 2008 Apr 1;110(1):43-59. [Google Scholar], [Academia]
24. Whitt, W. Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2002.[Springer]
25. Bouwmans, T.; Silva, C.; Marghes, C.; Zitouni, M.S.; Bhaskar, H.; Frelicot, C. On the role and the importance of features for background modeling and foreground detection. Comput. Sci. Rev. 2018, 28, 26–91. [Elsevier]
26. Tuceryan M, Jain AK. Texture analysis. Handbook of pattern recognition and computer vision. 1993:235-76.[Google Scholar]
27. Mikolajczyk, K.; Zisserman, A.; Schmid, C. Shape rEcognition With Edge-Based Features. 2003. Available online: https://hal.inria.fr/inria-00548226/
28. Vu TH, Osokin A, Laptev I. Context-aware CNNs for person head detection. InProceedings of the IEEE International Conference on Computer Vision 2015 (pp. 2893-2901). [Google Scholar]
29. Li, S.Z. Markov Random Field Modeling in Computer Vision; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.[Google Books]
30. Lempitsky V, Zisserman A. Learning to count objects in images. Advances in neural information processing systems. 2010;23.[Google Scholar]
31. Sun Z, Wang Y, Tan T, Cui J. Improving iris recognition accuracy via cascaded classifiers. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2005 Jul 25;35(3):435-41. [Google Scholar]
32. Zhang C, Li H, Wang X, Yang X. Cross-scene crowd counting via deep convolutional neural networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 833-841). [Google Scholar]
33. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. Photo-realistic single image super-resolution using a generative adversarial network. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4681-4690). [Google Scholar]
34. Shen Z, Xu Y, Ni B, Wang M, Hu J, Yang X. Crowd counting via adversarial cross-scale consistency pursuit. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 5245-5254).

35. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481–2495. [Google Scholar]
36. Yao, Xue., Siming, Liu., Yonghui, Li., Xueming, Qian. (2020). Crowd Scene Analysis by Output Encoding. arXiv: Computer Vision and Pattern Recognition.[Google Scholar]
37. Elbishlawi S, Abdelpakey MH, Eltantawy A, Shehata MS, Mohamed MM. Deep learning-based crowd scene analysis survey. Journal of Imaging. 2020 Sep 11;6(9):95. [Crossref]
38. Siraj M. Machine and deep learning for crowd analytics. arXiv preprint arXiv:1909.04150. 2019 Aug 25.[Google Scholar]
39. Wang, Zhiyu., Yang, Jiaxin., Yang, Jiaye. (2019). Scene analysis system based on deep learning technology and method thereof.
40. Reddy MK, Hossain M, Rochan M, Wang Y. Few-shot scene adaptive crowd counting using meta-learning. InProceedings of the IEEE/CVF winter conference on applications of computer vision 2020 (pp. 2814-2823).[Google Scholar]