IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12729

Deep Learning Techniques for Crowd Analysis

Prathmesh Jadhav[1], Pratika Murgod[2], Abhishek Nazare[3]

Department of MCA, KLS Gogte Institute of Technology/VTU, India[1,2,3]

Abstract: Crowd analysis plays a crucial role in various domains, including security, transportation, and social behavior understanding. Deep learning techniques have emerged as a powerful tool for handling the complexities and challenges associated with crowd analysis tasks. This survey report delves into the recent advancements in deep learning techniques for crowd analysis, highlighting their applications, strengths, and limitations. We explore various approaches used in crowd counting, crowd behavior understanding, and crowd anomaly detection. Additionally, we discuss the datasets commonly employed for evaluating these techniques. By shedding light on the current state-of-the-art, this survey aims to provide insights into the future prospects of deep learning in crowd analysis.

Keywords: Crowd Analysis, Deep Learning, Crowd Detection, Challenges, Convolutional Neural Network.

I.INTRODUCTION

1.1 Background

Crowd analysis has become an indispensable field of research with the proliferation of large gatherings in various scenarios, ranging from public spaces to major events and critical infrastructure sites. Understanding crowd dynamics, behavior patterns, and density estimation has significant implications for public safety, security, and urban planning. However, analyzing crowds manually is a challenging and time-consuming task, prompting the need for automated and intelligent systems.

In recent years, deep learning has revolutionized the landscape of computer vision and pattern recognition tasks. Leveraging the power of artificial neural networks, deep learning models have demonstrated exceptional performance in image and video analysis, leading to ground breaking advancements in crowd analysis. This integration of deep learning techniques with crowd analysis has unlocked novel opportunities to tackle the complexities posed by diverse and dynamic crowds.

Existing methods based on face detection struggle in highly dense circumstances, where conventional detectors fail due to intense occupation and perspective issues from various angles. Annotating dense crowds for training data is also arduous, and obtaining representative datasets is challenging. Innovative approaches are needed to handle non-stationary tracking cameras and address the difficulties in accurately estimating crowd density and detecting suspicious activity amidst the highly concentrated nature of such events.

1.2 Motivation

This paper presents a comprehensive exploration of the applications, advancements, and challenges in the domain of deep learning in crowd analysis. We delve into the fundamental components of crowd analysis, encompassing crowd counting, density estimation, and behavior understanding. Subsequently, we investigate how deep learning architectures, CNN and RNN, have been adapted and extended to address these crowd analysis tasks with greater accuracy and efficiency.

As we progress, we uncover the significant impact of deep learning in various crowd analysis applications. These applications include anomaly detection to identify unusual or suspicious behaviors, crowd event prediction to anticipate potential incidents or crowd movements, and social behavior analysis to understand collective actions and emotions within crowds. The incorporation of deep learning methods in these domains has shown remarkable promise in improving crowd management and enhancing public safety measures.

Despite the tremendous strides made in utilizing deep learning for crowd analysis, numerous challenges persist. The availability of comprehensive and diverse annotated datasets remains a critical bottleneck for training deep learning models effectively.

In this context, the primary objective of this paper is to shed light on the advances and potential limitations of deep learning in crowd analysis. By analysing the current state-of-the-art techniques, identifying promising applications, and addressing challenges, we aim to provide a deeper understanding of the possibilities and directions for future research in this field.

1.3 Challenges and gaps

The use of Fully Convolutional Neural Networks (FCNN) has gained prominence in crowd analysis and monitoring,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12729

particularly in densely populated areas like pilgrim gatherings. Deep learning technologies have become crucial for video analysis and crowd density detection. However, monitoring pilgrim crowds with densities of 7 to 8 people per square meter poses significant challenges.

1.4 Contributions

This article focuses on reviewing the latest technology used for analyzing crowd videos in video surveillance systems. The new approach for crowd analysis relies on a type of artificial intelligence called Fully Convolutional Neural Network (FCNN). The related works in this field are divided into two main categories: network-based and image-based approaches. The article discusses different strategies using Convolutional Neural Networks (CNN) to explain their strengths and weaknesses in each category. Detailed analyses are provided for various approaches, comparing their performance using metrics like Mean Absolute Error (MAE) on different datasets such as UCF, World Expo (WE), Shanghai Tech Part A (STA), and Shanghai Tech Part B (STB).

II.STUDIES ON CROWD ANALYSIS

2.1 Dataset

CrowdHuman serves as a benchmark dataset designed to enhance the evaluation of object detectors in crowded scenarios. The dataset is extensive, featuring rich annotations and exhibiting a high degree of diversity. It comprises 15,000, 4,370, and 5,000 images for training, validation, and testing, respectively. The dataset includes a total of 470,000 human instances from the train and validation subsets, with each image containing an average of 23 individuals, presenting various occlusion challenges. Notably, each human instance is meticulously annotated with bounding boxes for the head, visible region, and full body. This dataset aims to establish a robust foundation for advancing research in human detection tasks.

This article reviews various research works related to crowd analysis, covering different approaches and techniques. These include global regression, deep learning methods, scene labeling data-driven approaches, detection-based methods, CNN-based methods, optical flow detection, object tracking, 2D Convolutional Neural Network, 3D Convolutional Neural Network, crowd anomaly detection, abnormal event detection for deep models, feature learning based on the PCANet, and representation of neural event patterns with deep GMM. The article explores the strengths and limitations of each method and highlights their contributions to the field of crowd analysis.

2.2 Crowd analysis by global regression

Crowd analysis through global regression involves monitoring pedestrian crowds by sensing or clustering trajectories. However, occlusions among people often limit the effectiveness of these techniques. Specific methods have been introduced for global count predictions, using low-level-trained regression. These methods are particularly useful in crowded situations and are computationally efficient.

One approach proposed by Lempitsky is crowd detection analysis based on regressing the pixel-level object density map. Fiaschi later utilized a random forest to reduce object density and improve training efficiency. Regression-based methods offer the advantage of estimating the number of objects in a video region, allowing the creation of interactive object counter systems. These systems can visualize regions and efficiently provide relevant feedback to users.

2.3 Scene labelling data-driven approaches

Scene labelling data-driven approaches are recommended for large-scale crowd applications, as they offer advantages in non-parameter format. These methods are efficient because they do not require extensive preparation. In data-driven

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12729

methods, information from training images is transferred to test images by finding the most suitable training photographs that match the test picture.

Liu proposed a non-parameter image parsing technique that searches for dense areas of deformation between images. Using data-driven scene labelling methods, similar scenes and audience patches are extracted from training scenes for an unknown location in the test image. This allows for effective scene labelling and understanding the crowd distribution in the given context.

2.4 Detection-based methods

Detection-based methods are used to identify and count pedestrians in crowd tracking applications. Some authors have proposed extracting specific features from appearance-based crowd data to perform crowd counting. However, these methods have limitations in recognizing large crowds effectively. To address this issue, researchers have employed partial methods to detect specific parts of crowd bodies, such as the head or shoulder, in order to count individuals accurately. This approach helps improve the accuracy of crowd counting, especially in dense and crowded scenarios.

2.5 Optical flow detection

Optical flow detection is a method that estimates the movement of crowd objects by analyzing the motion of pixels between consecutive frames in a video. It uses Vector-based approaches to track the movement of objects as they move through the frames. Optical flow can be effective in locating moving crowd objects even when the subjects are turning or in motion.

One advantage of optical flow is that it provides a dynamic and complex approximation of crowd movement. However, it also has limitations. For example, the space-time filtering approach, which uses multiple adjacent frames to track objects based on their movement over time, may not work well in real-time implementations for stationary objects or when the crowd is not moving at all. In such cases, alternative methods may be more suitable for detecting and analyzing crowd behavior.

2.6 Deep learning

Deep learning has been widely applied in various monitoring systems, including individual re-identification and pedestrian recognition. Deep models have gained popularity due to their ability to extract powerful patterns automatically. Researchers, like Sermanet , have found that deep learning features outperform handcrafted features in many applications. In some cases, gathering a large dataset of individuals for training deep models can be challenging. To overcome this, approaches like the CNN-based algorithm used by [37] combine engineered features like HOG (Histogram of Oriented Gradients) based on head detections, Fourier measurements, and interest point counting. However, these methods may suffer from accuracy decline under challenging conditions like weather changes, occlusions, and distortions in the scene. Researchers like Zhang use deep networks to estimate the count of people guided by image maps, which can be complex to create. Wang train a deep model for crowd estimation that measures crowd density and distribution. The use of deep learning has shown promising results in various crowd analysis tasks, and it continues to be an active area of research in the field of computer vision and crowd analysis.

2.6.1 2D convolutional neural network

The 2D Convolutional Neural Network (CNN) is a type of neural network that uses weight sharing to work with real-world data. ConvNets are multi-layered deep neural networks that use kernels, also known as sensitive zone fields, with the same parameters for all potential entry points. This weight sharing is crucial as it allows each node in a previous layer to be filled with a small kernel window. The weights in the CNN are distributed across computing devices, reducing the number of free variables

and improving overall efficiency.In the 2D CNN, feature maps (FM) are used to distinguish different elements across layers. These layers are called convolutional layers.

The CNN uses a traditional gradient-descent propagation method to learn from data and derive spatial functions. It is particularly effective when applied to video datasets, where it can analyze and process spatial features to make predictions or extract valuable information.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12729

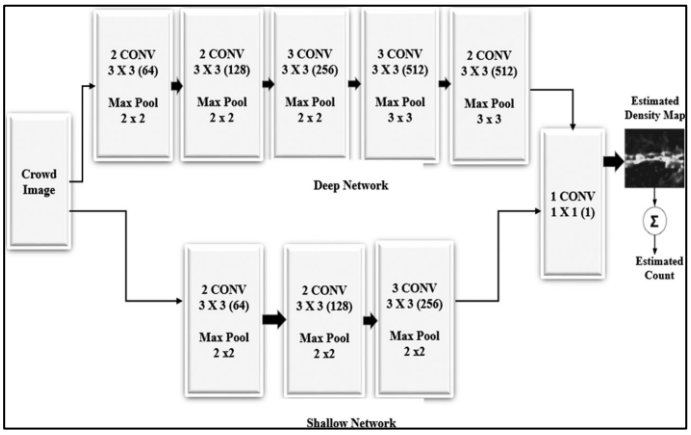

Fig. 1 Overview of the crowd counting

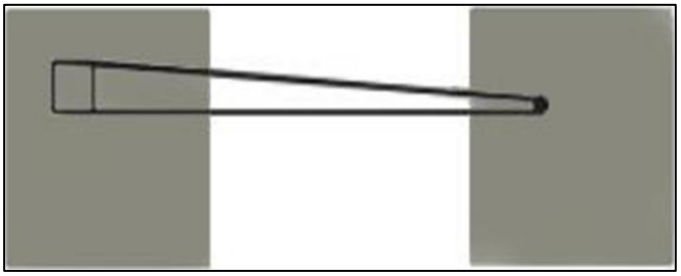

Fig. 2 2D convolution

2.6.2 3D convolutional neural network

The 3D Convolutional Neural Network (CNN) is used for anomaly detection in crowds. Before discussing anomaly detection, let's understand the concept of convolution. Convolution involves two functions, m and n, which produce a third function. This operation represents an altered version of one of the original functions, showing the difference between the two functions equivalent to the quantity of one of their original functions. This is denoted as m n and is represented with a star or an asterisk.

The 3D convolution is similar to 2D convolution but adds an additional dimension, which is the time dimension. It processes overlapping 3D cubes in the input video to extract motion information. The size of the convolution kernel in the time dimension is typically 3.

In a 3D CNN, the shared weights are used to extract features from the frame cube, and these weights may be applied again to the entire cube. The CNN's design philosophy aims to produce multiple feature maps by extracting various kinds of features from the same low-level feature maps. This is achieved by applying multiple 3D convolution operations to the previous layer at the same location with different cores.Multiple 3D convolutions can be applied to consecutive frames to extract various features, leading to multiple feature maps. Each connected set has unique weights, resulting in multiple feature maps on the right side, even though they may appear to have the same color due to the Shared weights. These multiple feature maps aid in Analyzing and understanding complex patterns and anomalies in the crowd video data.

2.7 Crowd anomaly detection

Crowd anomaly detection involves identifying irregular events in standard videos using various algorithms. Anomalies are classified into two types: (1) trajectory-dependent techniques, which focus on detecting abnormal trajectories that are rare compared to regular daily trajectories, and (2) local algorithms based on lines, where anomalies are seen as chains integrating dramatic case patterns.

In trajectory-dependent methods, abnormal behaviors are identified based on trajectory interpretation, speed, and acceleration. Clustering techniques are used to group trajectories, and anomalies are detected as clusters with few members or trajectories far from these cluster centers.

Another approach involves considering traceability at the particle and function factor level. For example, some researchers derive messy invariant properties from symbolic pathways using an approach to particle dynamics. Others analyze the direction of interest and reflect the crowd's dynamism using potential measurements.These methods contribute to crowd anomaly detection by detecting unusual and abnormal behavior patterns amidst regular crowd dynamics, helping in the identification of potential security threats and unusual events.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12729

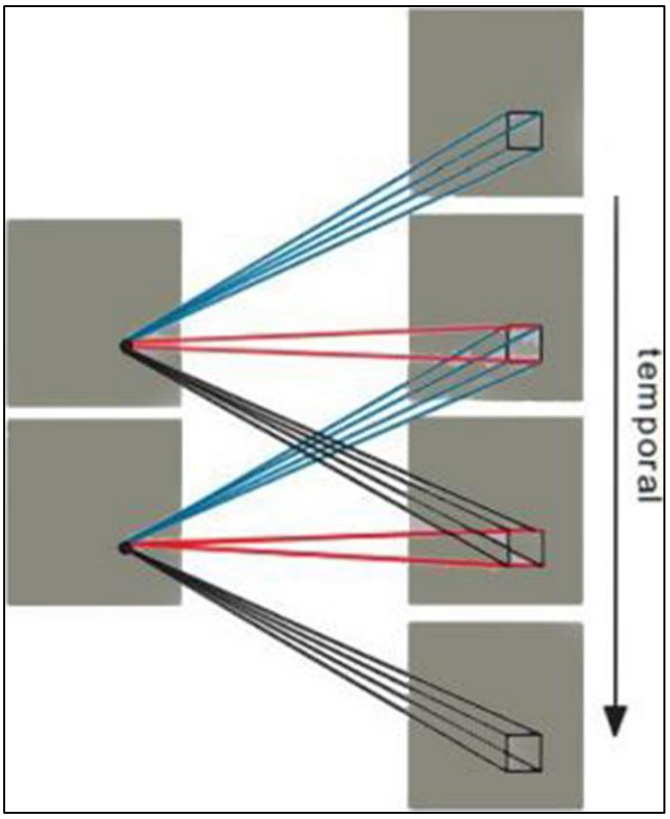

Fig. 3 3D convolution

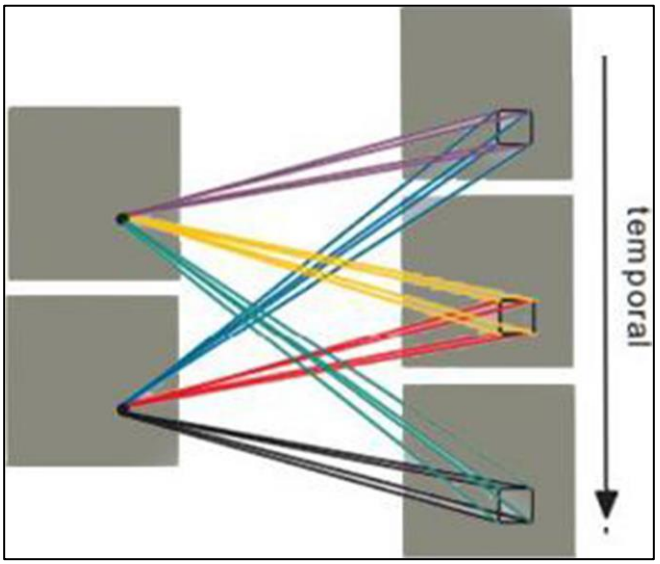

Fig. 4 Feature extraction from numerous consecutive frames

2.8 Feature learning based on the PCANet

The PCANet (Principal Component Analysis Network) is used for feature learning in anomaly detection. In most current techniques, spatial and temporal characteristics like intensity, color, gradient, and optical flow are manually selected. However, this paper focuses on calculating the 3D gradient for video events and explores its power and effectiveness in detecting unusual occurrences.

The approach combines both appearance (look) and motion clues for the 3D gradient prototype. A deep neural network is utilized to abstract high-level features primarily based on these 3D gradients. Deep learning has shown significant performance improvement in computer vision applications due to its ability to extract meaningful and discriminatory

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023
DOI: 10.17148/IJARCCE.2023.12729

features adaptively through non-linear multi-layer transformations.
In the anomaly detection domain, there are no marked unusual activities available for training. Therefore, this paper adopts an approach of learning from video opportunities using PCANet features, which are simple and effective, making them suitable for unsupervised anomaly detection. This method allows the system to learn and identify abnormal patterns in videos without the need for labelled abnormal samples in the training dataset.

III.FUTURE DIRECTIONS

1. Potential research directions in crowd analysis using deep learning:

In the future, crowd analysis using deep learning could explore more advanced architectures and models tailored specifically for crowd-related tasks. This may involve developing novel neural network architectures to handle large-scale crowd data and overcome challenges like occlusion, perspective changes, and varying crowd densities.

2. Multi-modal crowd analysis techniques:
Integrating multiple sources of data, such as video, audio, and other sensors, can lead to more comprehensive and accurate crowd analysis. Future research can explore multi-modal deep learning techniques that combine information from various sources to enhance crowd understanding. This may include audio-based crowd emotion recognition, fusion of visual and thermal imagery for crowd tracking, or incorporating social media data for crowd sentiment analysis.

3. Explainable deep learning for crowd analysis:
As deep learning models become more complex, explaining their decisions becomes essential, especially in critical applications like crowd analysis. Future research can focus on developing explainable deep learning techniques to provide insights into why certain predictions or detections are made. This could improve the transparency and trustworthiness of crowd analysis systems.

4. Integration with other AI technologies for enhanced crowd understanding:
Integrating deep learning with other AI technologies, such as natural language processing and knowledge graphs, can lead to more advanced crowd understanding. By incorporating textual information and domain knowledge, AI systems can gain a deeper understanding of crowd behavior and intent, enabling more context-aware analysis.

5. Real-time crowd analysis systems:

Efforts can be directed towards developing real-time crowd analysis systems that can process and analyse crowd data in real-time, enabling prompt responses to potential crowd-related events. This would require optimizing deep learning models for speed and developing efficient algorithms for real-time data processing.

IV.CONCLUSION

This paper has reviewed different approaches, techniques and frameworks used for crowd Analysis, In conclusion, crowd analysis and monitoring have become critical areas of research, with real-world applications in crowd management, security, and event planning. The use of advanced technologies such as deep learning has significantly improved the accuracy and efficiency of crowd analysis. Fully Convolutional Neural Networks (FCNN) and Convolutional Neural Networks (CNN) have emerged as powerful tools for crowd density estimation, anomaly detection, and individual counting in densely populated areas.
In future research, exploring multi-modal crowd analysis techniques that integrate data from various sources, such as video, audio, and other sensors, can lead to more comprehensive insights. Additionally, explainable deep
learning can enhance the transparency and trustworthiness of crowd analysis systems, ensuring their reliability in critical applications.

REFERENCES

[1] Video analytics using deep learning for crowd analysis: a review by Md Roman Bhuiyan, Junaidi Abdullah, Noramiza Hashim and Fahmid Al Farid.
[2] Lempitsky V and Zisserman A (2010) “Learning To Count Objects in Images,” pp. 1–9.
[3] Brostow GJ, Cipolla R (2006) ‘Unsupervised bayesian detection of independent motion in crowds’, Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 1, pp.
[4] Wu B, Nevatia R (2007) Detection and tracking of multiple, partially occluded humans by Bayesian Combination of edgelet based part detectors. Int J Comput Vis 75(2):247–266.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023
DOI: 10.17148/IJARCCE.2023.12729

[5] Chan AB, Vasconcelos N (2009) 'Bayesian poisson regression for crowd counting'. Proceed IEEE Int Conf Comput Vision, (Iccv). 545–551.
[6] Fiaschi T, Giannoni E, Taddei ML, Chiarugi P (2012) Globular adiponectin activates motility and Regenerative traits of muscle satellite cells. PLoS One 7(5):e34782.
[7] Chen K et al (2012) 'Feature mining for localized crowd counting', BMVC 2012 - electronic proceedings of the British machine vision conference 2012.
[8] Ryan D et al (2009) 'Crowd counting using multiple local features', DICTA 2009 - digital image computing: techniques and applications, pp. 81–88.
[9] Dargan S, … Kumar G (2020) A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng. Springer Netherlands 27(4):1071–1092.
[10]Forsyth D (2014) Object detection with discriminatively trained part-based models'. Computer 47(2):6–7.
[11] Kang D, Chan A (2019) 'Crowd counting by adaptively fusing predictions from an image pyramid', British machine vision conference 2018. BMVC 2018:1–12
[12] Bendali-Braham M et al (2021) 'Recent trends in crowd analysis: a review', machine learning with applications. Elsevier ltd., 4(October 2020), p. 100023.