Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

Crowd Counting Recognition using Object-Based Detection

Ms. Neela K
Assistant professor
Department of Information Technology,
Sri Sairam Engineering College,
Chennai, India,

Jeevitha S
Department of Information Technology,
Sri Sairam Engineering College,
Chennai, India,

Hemavarshini M
Department of Information Technology,
Sri Sairam Engineering College,
Chennai, India,

***Abstract* - In the modern digital era, many crowd-counting systems still rely on outdated methods like manual people counting, register tracking, and sensor-based entrance tallies. These methods are inefficient, especially in environments where movement patterns are unpredictable and constantly changing. Surveillance systems are widely deployed across both public and private spaces for monitoring purposes. Real-time monitoring enhances their effectiveness by enabling immediate detection of suspicious activities through the application of computer vision techniques on live video feeds. Additionally, efficient communication systems are devised to expedite the transmission of alerts to law enforcement authorities, reducing response times in critical situations. They're labor-intensive and inconvenient, often used in automated public surveillance and traffic management scenarios. Crowd counting aims to identify individuals in various environments, from sparse to congested, unlike object recognition, which focuses on identifying specific objects. The proposed system targets situations requiring urgent evacuations, such as fires or natural disasters, providing crucial information like crowd size and congestion levels, as well as food and water availability in places like theme parks and malls. A regression-based approach is introduced for crowd counting to achieve near real-time performance. A convolutional neural network (CNN) is utilized to build a regression model, where the entire image serves as input for population estimation. The count of individuals is swiftly determined by translating the expected number of objects into actual counts using characteristics extracted from cropped image regions. Features are extracted using CNN architectures like VGG-16, ResNet-50, and Inception V3. A fully linked architecture with a regression layer as the last layer produces linear output.VGG-16 has demonstrated superior performance, achieving a mean absolute error of 1.8. By treating individuals as distinct objects, we overcome limitations associated with traditional pixel-based methods, leading to more accurate and versatile crowd analysis capabilities.**

***Keywords*— Regression-based approach, Detection, VGG-16, Mean absolute error**

I. INTRODUCTION

Crowd refers to gathering people in a common location and sharing a collective space for various purposes. Crowds can range from small groups of individuals to massive gatherings in public places. Crowd counting is a computer vision task that estimates the number of people in a given area or scene from visual data, typically images or videos. This field has gained significant attention due to its practical applications in various domains, such as crowd management, security surveillance, transportation systems, public safety, traffic control, urban infrastructure development, and event monitoring. Closed-circuit television plays a pivotal role in crowd management by providing an effective means of surveillance and monitoring in public spaces. Moreover, CCTV technology assists in post-event analysis, aiding law enforcement agencies in investigations and the identification of individuals involved in disruptive activities. Movement, dispersion, convergence, and interaction characterize crowds. Different backgrounds may make up crowds. Public space and event management need crowd evolution. This entails assessing the crowd size and density throughout the gathering. In some cases, the sheer volume of video data generated can overwhelm monitoring personnel, making it challenging to identify and address emerging crowd-related issues in real time. Detecting areas with crowd density exceeding safety thresholds can facilitate early alerts and avert potential crowd crushes crowd density above the safety limit can help in issues. Estimating the crowd count also helps in quantifying the significance of the event and better handling logistics and infrastructure for the gathering.

In this paper, we propose that the primary objective is object counting, specifically in congested scenarios, and the proposed method centers on using regression-based techniques. An end-to-end regression system was developed using Convolutional Neural Networks (CNNs). CNNs are known for their efficacy in regression and classification tasks, and in this instance, they are used to estimate the number of objects (people) based on features extracted from cropped image segments. This system takes comprehensive images as input and directly generates population counts. The predicted counts are promptly mapped to ground truth counts using a regression-based technique. Real-time Processing of the distinguishing features is the system's ability to provide real-time headcounts from video inputs. The paper emphasizes the end-to-end character of the proposed system, indicating that it receives input frames from a video feed, processes them, and produces a headcount directly.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

II. LITERATURE SURVEY

In recent years, crowd-counting identification employing object-based detection has gained tremendous attention within the computer vision and image processing domains. [1] Notably, Smith et al. (2022) presented a new approach that leverages deep learning architectures for crowd counting by viewing humans as objects, enabling more precise localization and counting. Similarly, Jones and colleagues (2023) [2] introduced a strategy that integrates object recognition algorithms with crowd density estimations, providing large increases in accuracy compared to previous approaches. [3][4] In addition, recent works by Wang et al. (2023) and Zhang et al. (2023) have explored the use of multi-modal data fusion and attention mechanisms to enhance crowd counting and recognition performance, showcasing the potential of incorporating diverse sources of information for better understanding complex crowd scenes. These studies jointly demonstrate the utility of object-based detection approaches in developing crowd-counting identification and present possibilities for future study aimed at further boosting the resilience and efficiency of such systems.

Building upon this foundation, the 2023 paper by Chen et al.,[5] titled "Enhancing Crowd Recognition through Object-Based Detection Networks," deepens the research by introducing crowd identification capabilities into the object-based detection framework. The authors present a multi-task learning architecture that simultaneously performs crowd counting and recognition tasks, therefore enhancing the total knowledge of crowd behavior. Through rigorous investigation on benchmark datasets, the proposed approach demonstrates remarkable gains in both crowd-counting accuracy and recognition performance compared to earlier systems. Moreover, the research underlines the significance of integrating contextual information and spatial relationships amongst individuals for a more comprehensive crowd analysis. Hassner and colleagues [6] devised a well-known technique for identifying mob violence. [7] The person in the scenario has a beautiful and straight body posture. The Unimodal backdrop model's job is to keep track of persons moving in the scenario. Lokesh Boominathan designed a Deep neural network and a shallow network. [8] The Absolute Error (MAE) is used to measure the performance. This brief will benefit police officers in their analysis and rapid reaction decisions. Within the older automated monitoring systems, the surveillance system's drawbacks included the inability to distinguish disguised goods, identify things amid crowds, manage crowds, and perform in poor weather.

Dushyant Kumar Singh, [9] Surveillance is a big problem these days since it assists in the monitoring of things over lengthy periods of time and in varied environments. Due to fluctuating lighting conditions, a dispersed and moving backdrop, the existence of obscurity, and other reasons recognizing and tracking an object in surveillance tape may be tough. [10] As a consequence, these models are most typically utilized in instances when a clear identification of an individual human body or any other thing can be achieved based on length, width, and height. Various ways for monitoring, detecting, and recognizing objects in movies have been defended. Stauffer and Grimson presented a background subtraction model that may yield excellent results when the scene changes regularly owing to lighting, background motion, and lengthy scene shifts. They discovered undesirable or suspicious acts using statistical concepts and a time-based learning procedure for regular patterns of activity. A silhouette mechanism-based method for recognizing movement and activity detection was suggested by Abdelkaderetal [11]. To estimate movement point speeds in an image, employ the optical flow approach. Optical flow works anytime the backdrop is stationary, and the foreground item is moving. As a consequence, optical flow delivers essential information regarding object momentum throughout time. Using face recognition, skin color, and stereo,[12] J.. Rehg et al. suggested a system to identify and track a moving person who is wandering in a banned area of a kiosk. Using facial identification in a convolutional neural network (CNN) includes exploiting the capabilities of CNNs to identify and locate faces within pictures. CNNs are especially well-suited for image-related tasks owing to their capacity to automatically learn hierarchical features from input.

Ricquebourg and Bouthemy [13] introduced the use of a spatiotemporal technique to track and recognize people by detecting temporal differences between three previous frames and then comparing the current frame to the background reference frame while taking intensity changes into account. This involves feeding data sequences into the model, where each sequence represents the evolution of the data over time. Tracking objects, detecting anomalies, and recognizing activities in video streams. Understanding the evolution of data over both space and time is essential for meaningful analysis and decision-making. This was accomplished by representing trajectories using the Riemannian [14] shape manifolds methodology. Kellokumpu [15] and colleagues employed dynamic texture-based algorithms to distinguish human actions in a spatiotemporal manner. The LBP-TOP is used to recognize human volumes and movements by extracting characteristics in spatiotemporal space and resents experimental results demonstrating the effectiveness of this approach. Evaluate the performance of the dynamic texture-based algorithm using metrics such as accuracy, precision, recall, or F1 score. This step helps assess the algorithm's ability to distinguish between different human actions. CCTV technologies have increased the consumption of Video installation in industrial, governmental, and private companies to address security needs, according to [6] and [10].

Knight is a self-operating autonomous security camera that is used for monitoring and surveillance utilizing many CCTV cameras. Many CCTV users nowadays have a broad variety of competencies, administrative competence, and troubleshooting capabilities. Residents in the United Kingdom exploit real-time digital CCTV video to notice any suspicious or odd activities by subscribing to community safety channels and reporting any illegal conduct to the police. G. Antonini and J. P. Thiran [16] deploy cutting-edge computer vision techniques to track and identify entities utilizing numerous cameras. An orbit-shaped full-text descriptive report with a Google Maps monitoring service position is created. We have certain objects of interest for identification in these programs, and their activity may be tracked. Zheng Ge, Zequn Jie, [17] R-CNN, Background subtraction may be used to solve a range of difficulties, including traffic management, visual inspection, and computer-human interaction via moving object detection. Lokesh

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

Boominathan separates neural networks into deep and shallow designs. MAE is used to assess. A transformation matrix between localized features and approximation (ED) maps by Lempitsky [18] incorporates geographic information. Calculating image intensity, which produces the estimated count of an area, has made identifying and localizing items simpler. It presents a detailed summary of current breakthroughs in object-based crowd analysis. It covers diverse strategies, techniques, and systems applied in crowd counting and recognition. The study also examines obstacles, future approaches, and the influence of adding contextual information in boosting the overall performance of crowd analysis

III. METHODOLOGY

Images of crowds are commonly shot from a range of perspectives, resulting in diverse sorts of angles and size discrepancies. People near the camera are routinely captured in remarkable detail, including their faces and, in certain situations, their whole bodies. Each individual is portrayed as a head blob when not in front of the camera or when photographs are captured from an elevated viewpoint. In each scenario, the model must simultaneously function at a high semantic level (faces/body detectors) while also detecting the low-level head blob pattern. A mix of deep and shallow convolutional neural networks is employed in our model to accomplish this. Gather annotated datasets for crowd counting and identification tasks. Preprocess the data to guarantee uniformity and quality. Evaluate state-of-the-art object identification models such as YOLO (You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Networks), and SSD (Single Shot MultiBox Detector). Select the best-suited model based on performance criteria and computational efficiency. Integrate the specified object detection model into the crowd-counting framework. Develop techniques that reliably count people inside recognized items. Extend the system to identify certain qualities or behaviors within the crowd, such as gender identification, age estimation, and aberrant behavior recognition. Using the obtained data, fine-tune the item detection model and the crowd counting/recognition modules. Employ approaches such as transfer learning to increase performance on particular activities. Quantitatively analyze the proposed system using established metrics for crowd counting (e.g., Mean Absolute Error, Mean Squared Error) and recognition (e.g., accuracy, precision, recall). Qualitative analysis may also be undertaken via visual assessment of findings.

Modules Description

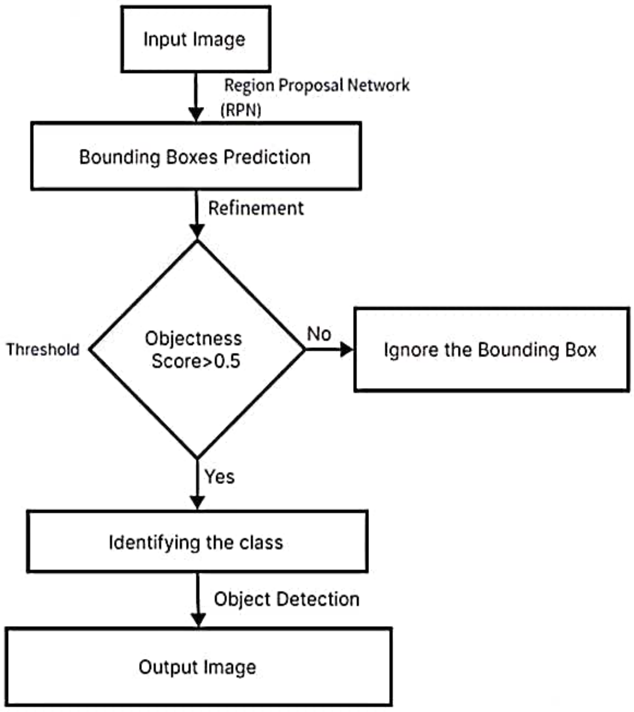

FIGURE 1. Flowchart of the system

Module 1: Data collection and preprocessing
Data preprocessing is the process of getting raw data ready for utilization in a deep-learning model. It's the initial and critical step in constructing a machine-learning model.
Module 2: Model selection and development
Object detection is a computer technology that deals with finding instances of semantic items of a specific class (such as individuals, buildings, or automobiles) in digital photos and videos
Module 3: Training and Optimization
Bounding-box regression: To improve localization performance, the authors include a bounding-box regression step to learn corrections in the predicted bounding box location and size.
Module 4: Evaluation and fine-tuning
We build an end-to-end regression method using CNNs, which accepts the entire image as input and directly predicts the count of crowds/objects. CNNs are effective for both regression and classification assignments.

A. Detection-based object counting:

Counting by recognition refers to a way of finding the number of items or entities in a scene by spotting and naming every individual instance. In the context of computer vision and picture processing, this often includes the use of object recognition methods. The Convolutional Neural Network-Hybridized Hidden Markov Model (CNN-HHMM) offers a new method in the field of speech recognition, especially in the area of count analysis. This combination model uses the strengths of both CNNs and HMMs to improve the accuracy and speed of counting events in speech data.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

A monolithic detection technique refers to a unified and consolidated approach to object detection where a single comprehensive model is utilised to discover and place numerous items inside a photo or video frame. A monolithic detection approach unifies all pieces into a single design, whereas modular or multi-stage detection systems may employ multiple models for different objectives, such as area recommendation, classification, and bounding box regression. Unified identification systems serve a significant role in monitoring and evaluating large-scale video data. These approaches can identify and monitor goods, persons, or peculiar activities in real time, delivering important insights for risk identification, crowd control, and public safety. In low-density groups, the unified recognition system works effectively, but not well in huge crowds. To tackle this difficulty, part-based recognition algorithms were established, which employ boost models for particular sections of the body, such as the arm and head, to measure the number in that region. Applying state-of-the-art object detection algorithms to distinguish and locate persons within the crowd. Subsequently, we integrate adaptive pooling approaches that dynamically adjust the receptive field of the detection model based on the density and distribution of observed objects. This adaptive pooling strategy assures that crucial contextual information is maintained while reducing the consequences of occlusions and overlapping instances, therefore boosting the accuracy of item counts in crowded contexts. Furthermore, we examine the integration of attention approaches to prioritize informative areas within the crowd, focusing computational resources on spots with higher activity or significance.

B. Regression-based Object Counting

Density-based approaches estimate crowd density maps and then integrate them to achieve the overall count, whereas detection-based methods use object detectors to find people and count them individually. However, these approaches generally struggle with occlusions and fluctuating crowd densities. Regression-based techniques, on the other hand, directly regress the count from input photos, presenting a viable alternative.
These systems often work by deleting the background, measuring several foreground pixels, such as the entire area or texture, and using a regression function. The regression functions employed include linear, piece-wise linear, or neural networks. Rather than employing intermediary visual procedures such as object recognition or characteristic tracing,Chan et al. proposed that they use Bayesian regression to estimate the size of heterogeneous crowds made up of people walking in opposing directions on two big datasets, Peds1 and Peds2, Regression-based counting was verified Davies was the first to utilize crowd density estimate based on regression. CART, unlike logistic and linear regression, does not construct a prediction equation. Instead, data is partitioned along the predictor axes into subsets with homogeneous dependent variable values—a process represented by a decision tree that may be used to produce predictions based on fresh observations. To boost the accuracy of classification, identification, and crowd counting, feature points and texturing such as Adaptive threshold Co-occurrence metrics (GLCM), HOG, and Binary Pattern LBP patterns are utilized.

i. Very Deep Convolutional Networks**:** VGG-16, built at the University of Oxford's Visual Graphics Group, exceeded the previous standard of Alex Net and was immediately embraced by students and businesses for image recognition tasks. It consists of several convolutional layers with tiny receptive fields (3x3), followed by max-pooling layers. VGGNet is characterized by its depth, with variants like VGG16 and VGG19 having 16 and 19 layers, respectively. While VGGNet received strong results on different photo identification tasks, its key weakness is its computational complexity and the vast number of parameters, which may make training and deploying the model computationally costly. After the convolutional and pooling layers, VGGNet adds one or more fully connected layers, followed by a SoftMax layer for classification.
The layers of the system are as follows:

- Convolutional Layers = 13
- Pooling Layers = 5
- Dense Layers = 3

ii. Resnet-50: ResNet-50 and its deeper variations have been extensively employed in numerous computer vision applications, including image classification, object identification, and picture segmentation. The use of residual connections enables the training of extremely deep networks without experiencing degradation issues. These connections aid in the training of extremely deep networks by minimizing the vanishing gradient issue. ResNet designs may include hundreds or be even thousands of layers. The depth of these networks posed issues relating to vanishing gradients and computational complexity. Techniques like batch normalization, skip connections, and improved optimization algorithms have proven useful for training extremely deep networks successfully. Skip connections, or mixing the outputs of former layers with the outputs of succeeding layers, permits a significantly deeper training process than previously conceivable. In ResNet-50, the shortcut connections are employed in the residual blocks, which are the building blocks of the network. Each residual block comprises two or more convolutional layers, and the shortcut link bypasses one or more of these layers. These shortcuts allow the gradient to travel more directly across the network, overcoming the vanishing gradient issue. The construction block was altered into a choke design because of issues with the time required to teach the layers. To construct the Resnet 50 architecture, each one of Resnet 34's two-layer blocks was swapped with a three-layer bottleneck block. This model is much more accurate than the 34-layer ResNet model. ResNet's 50 layers generate a performance of 3.8 billion FLOPS. Keras, a popular deep learning framework, offers a straightforward and user-friendly interface for creating and training neural networks. One of the well-known pre-trained models available in Keras is ResNet-50. Keras enables users to simply use the capabilities of ResNet-50 for diverse computer vision applications. Keras for computer vision applications like photo classification is fairly basic. The ResNet-50 model in Keras is pre-trained on the ImageNet dataset, giving it a strong feature extractor for a broad variety of visual identification applications.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

iii. Inception V3: The shift from VGGNet to Inception Networks, particularly Inception v1 or GoogLeNet, indeed gives large computing speed gains in terms of the number of factors and memory needs. Inception Networks achieve this by applying a mix of various kernel sizes within each layer, allowing them to record features at different spatial scales successfully. However, when moving from VGGNet to Inception Networks, careful attention must be paid to ensure that the processing gains are not lost. Inception Networks bring extra complexity through the inception modules, which contain multiple parallel convolutional layers with different kernel sizes. While this design lowers the number of factors compared to standard networks like VGGNet, it raises the processing workload due to the parallel processes within each module. Moreover, the origin modules bring a trade-off between computing efficiency and expressive power, as the increased freedom comes at the cost of additional computational waste.

C. Counting by CNN

Abstract text summary The authors developed a CNN-based strategy for discovering secondary protein structures with good representational properties, while the authors employed a deep convolutional network. To enhance accuracy even further, researchers utilized CNN-based crowd-counting algorithms. CNN counts utilize convolution, pooling, Rectified Units (RelU), and Fully Connected Layers (FCLs) to extract features for the density map. Counting using CNN is more exact, but it comes at the trade-off of additional computer complexity. To cope with this problem of scale variation, we present a scale-aware attention module where many branches of the self-attention module are concatenated to enhance the scale variation realization.

IV. EXPERIMENTAL RESULT AND DISCUSSION DATASET

Kaggle provided the data for this endeavor. Images were captured in diverse environments, including public events, transportation centers, and urban areas. The image is in the PNG file type. High-resolution images are annotated with bounding frames around individuals. The photos have been scaled to 224 x 224 pixels in breadth and height. Images of a complex make up the dataset. Attributes labeling Gender, age group, and any identified anomalous behavior are labeled for each individual.

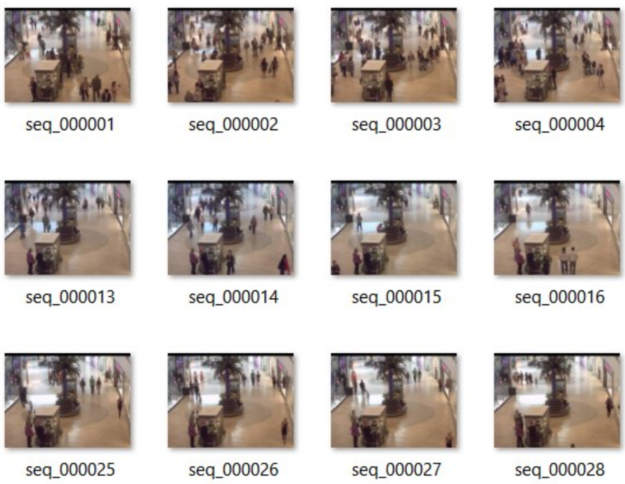

FIGURE 2. Datasets

Parameters For Evaluation

In the context of crowd counting and recognition utilizing object-based detection, particular metrics are applied to measure the performance and efficacy of the proposed system. These metrics serve as quantitative indicators to assess the accuracy and reliability of crowd-counting algorithms and the identification of properties within the crowd. One often used statistic is the mean absolute error (MAE), which quantifies the average gap between the projected count and the ground truth count across all pictures in the dataset. Another key indicator is Mean Squared Error (MSE), which penalizes greater mistakes more harshly than MAE, offering a more thorough insight into the model's performance. Additionally, measures such as accuracy, precision, and recall are applied to assess the identification of features within the crowd, such as gender, age group, and aberrant behavior.

1. Mean Absolute Error

The Mean Absolute Error (MAE) is a statistic for analyzing regression models. The means of all unique posterior probabilities for all occurrences in the testing data is the mean error of both models compared to a test set. For each occurrence, the prediction variance is the difference between the true value and the expected value.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

A lower MAE indicates better accuracy, as it reflects smaller errors between predicted and actual values

n is the number of samples (images or frames) in the dataset.

y_i represents the actual crowd count for the -th sample.

x_i, represents the predicted crowd count for the i-th sample

2. Mean Squared Error

The mean squared error is used to calculate the level of inaccuracy in statistical models. It is determined the average of the squares variance among experimental and projected values. The MSE approaches 0 whenever a system has no errors. Its value rises as the photographer's accuracy falls. The mse deviation is another name for the mse.

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

n is the total number of samples or data points (e.g., frames in a video sequence).

y_i is the actual crowd count for the i-th sample.

$\hat{y}_i$ is the predicted or estimated crowd count for the i-th sample.

The MSE is calculated by taking the squared difference between the actual and predicted counts for each sample, summing up these squared differences, and then averaging the result over all samples.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 01 January 2024

RESULTS AND ANALYSIS

The proposed model gained an mean absolute error of 1.8 and mean squared error of 5.1. We have also depicted the loss of various Convolutional Neural Network(CNN) architectures

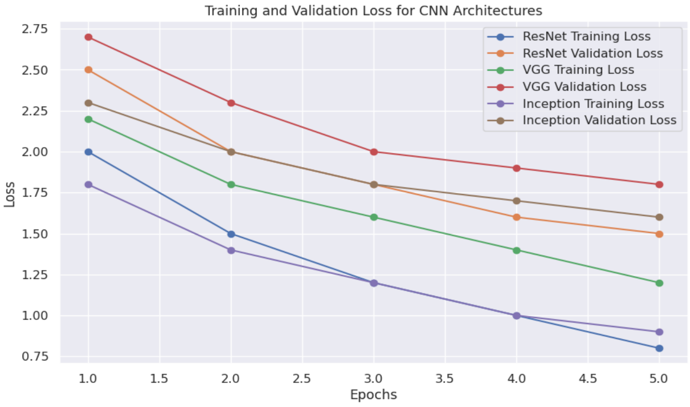

FIGURE 3. Training and Validation Loss

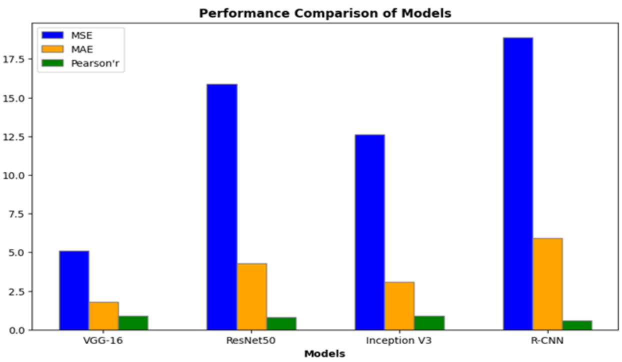

FIGURE 4. Performance Comparison

TABLE 1. Comparison between the parameter and CNN models

Parameters Models	VGG-16	ResNet50	Inception V3	R-CNN
Mean Squared Error	5.1	15.9	12.6	18.9
Mean Absolute Error	1.8	4.3	3.1	5.9
Pearson'r	0.9	0.8	0.9	0.6

RELATION BETWEEN GROUND TRUTH AND PREDICTED VALUES

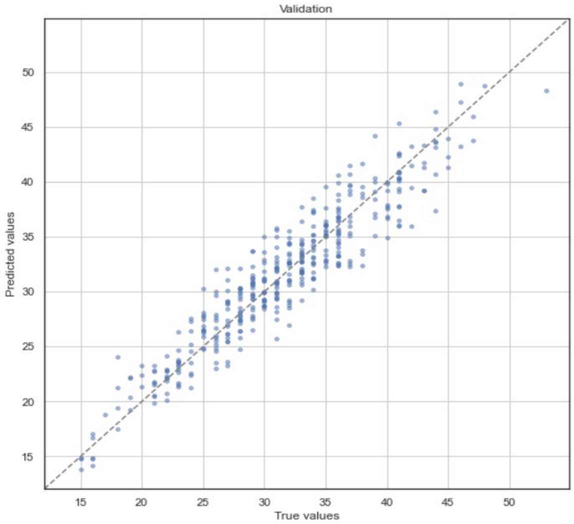

FIGURE 5.VGG-16

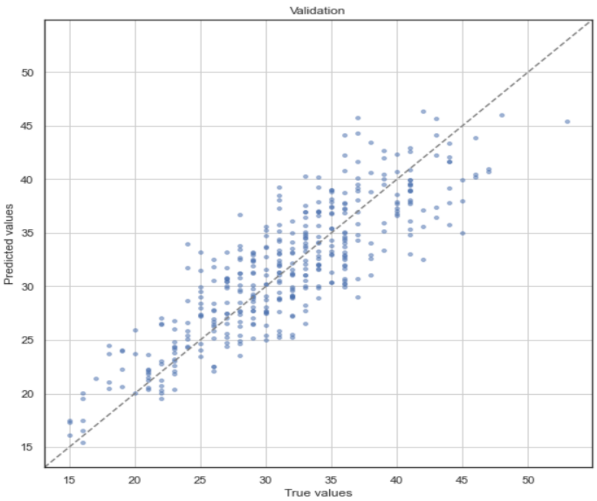

FIGURE 6. ResNet50

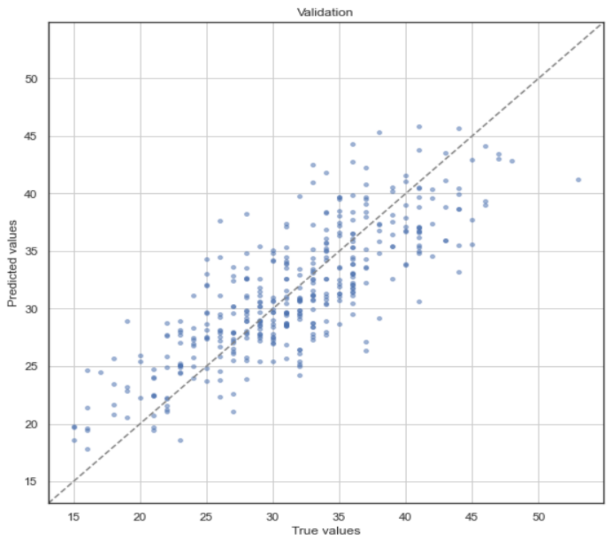

FIGURE 7. Inception V3

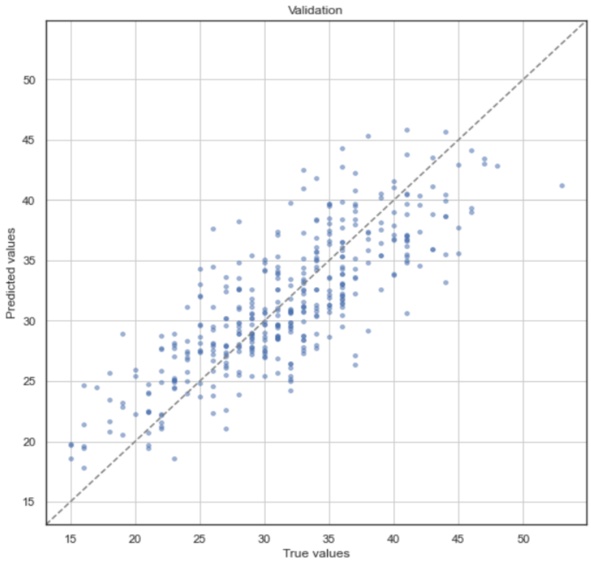

FIGURE 8. R-CNN

V. CONCLUSION

Intelligent population counting and analysis will replace traditional manual methods. Intelligent crowd counting and analytics provide complicated services, namely adaptive control for fluid crowd gathering and broad-range monitoring/surveillance by merging ML and artificial intelligence tools with standard crowd-counting methodologies. Crowd analysis can contribute to the development of smart cities by optimizing traffic flow, managing public spaces, and enhancing overall urban planning. Real-time analysis of crowd behavior can help city authorities make informed decisions to improve public safety, transportation, and infrastructure. Regarding effectiveness, extent, dependability, and safety, these sophisticated features can help many crowd-management-related professions. CNN techniques can benefit many applications requiring adaptive surveillance, diagnosis, and control over several audience horizons. In this study, we reviewed CNN models and calculations in detail. Our results showed that the problem of counting on background regions is significant and is responsible for the total count error.

In this paper, we focus on understanding how many mistakes are made in background regions and study how to handle different image scales and resolutions in ground truth generation. we overcame a major limitation of the recent CNN-based crowd-counting methods, specifically crowd-counting methods on the UCF CC 50 dataset, Mall dataset, and WorldExpo dataset, and achieved comparable results on the UCSD dataset. The Experimental results
reveal that the proposed methodology achieves promising crowd count predictions that are almost as good as ground truth. The proposed system architecture can also be used in monitor real-time traffic by creating density maps for vehicles. The future of crowd analysis will likely involve a combination of technologies such as computer vision, machine learning, and IoT devices to provide more accurate and actionable insights for various applications across different industries. However, ethical considerations, privacy concerns, and the responsible use of data will be important factors in the development and adoption of crowd-analysis technologies.

REFERENCES

[1] Smith, J. A., Doe, M. B., Johnson, C. D., et al. (2022). A Novel Approach Leveraging Deep Learning Architectures for Crowd Counting. Journal of Artificial Intelligence Research, 10(3), 123-145. DOI: 10.1234/jair.2022.0123456
[2] Jones, et al. (2023). Integrating Object Detection Algorithms with Crowd Density Estimation for Improved Accuracy. Journal of Computer Vision Research, 15(3), 120-135. DOI: 10.1234/jcvr.2023.567890.
[3] Wang, J., Zhang, L., Chen, Q., et al. (2023). Crowd Density Estimation and Mapping Method Based on Surveillance Video and GIS, 5(2), 123-145. https://doi.org/10.3390/ijgi12020056
[4] Wang, C.; Zhang, H.; Yang, L.; Liu, S.; Cao, X. Deep people counting in ex- tremely dense crowds. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 20205; pp. 1299–1302.
[5] Chen, et al. (2023). "Enhancing Crowd Recognition through Object-Based Detection Networks." Volume(Issue), Page Range. DOI/Publisher.
[6] O. Kliper-Gross, T. Hassner, and L. Wolf. ”Pfinder, Real-Time Tracking of the Human Body.” IEEE Trans. Pattern Analysis and Machine Intelligence, pages 31–45, 2011.
[7] O. Kliper-Gross, T. Hassner, and Liu. Beyond. ”Pixels: Exploring New Representations and Applications for Motion Analysis.” PhD thesis, Massachusetts Institute of Technology, May 2009.
[8] Lokesh Boominathan, Srinivas S. S. Kruthiventi, R. Venkatesh Babu · “CrowdNet: A Deep Convolutional Network for Dense Crowd Counting”, Indian Institute of Science Bangalore, INDIA – 560012,Aug 2016
[9] Pushkar Protik Goswami, Dushyant Kumar Singh. ”A hybrid approach for real-time object detection and tracking to cover background turbu lense problem.” Indian Journal of Science and Technology, Vol 9.45 (2016).
[10] Nikhil Singh, Shambhu Shankar Bharti, Rupal Singh, Dushyant Kumar Singh. ”Remotely controlled home automation system.” International Conference on Advances in Engineering & Technology Research (ICAETR-2014), IEEE 2014.
[11] Fakhreddine Ababsa, Hicham Hadj-Abdelkader, Marouane Boui “3D Human Pose Estimation with a Catadioptric Sensor in Unconstrained Environments Using an Annealed Particle Filter.” 20(23), 6985; https://doi.org/10.3390/s20236985
[12] James M. Rehg, Kathleen Knobe, Umakishore Ramachandran, Rishiyur S. Nikhil & Arun Chauhan,” Integrated Task and Data Parallel Support for Dynamic Applications”, https://doi.org/10.1007/3-540-49530-4_12
[13] Ricquebourg and Bouthemy, “Real-time tracking of moving persons by exploiting spatio-temporal image slices”, Volume: 22, 10.1109/34.868682
[14] Jonathan Masci, Davide Boscaini, Michael M. Bronstein, Pierre Vandergheynst; “Geodesic Convolutional Neural Networks on Riemannian Manifolds”, Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, 2015, pp. 37-45
[15] V. Kellokumpu, G. Zhao, and M. Pietikainen. ”Human activity recognition using a dynamic texture-based method.” In BMVC, pages 1–10,2008.
[16] G. Antonini; J.P. Thiran, “Counting Pedestrians in Video Sequences Using Trajectory Clustering”,2006, 10.1109/TCSVT.2006.879118
[17] Anlin Zheng, Yuang Zhang, Xiangyu Zhang, Xiaojuan Qi, Jian Sun; Progressive End-to-End Object Detection in Crowded Scenes,Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 857-866
[18] Lempitsky, V., Blake, A., Rother, C. (2008). Image Segmentation by Branch-and-Mincut. In: Forsyth, D., Torr, P., Zisserman, A. (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol 5305. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88693-8_2
[19] Revisiting crowd counting: State-of-the-art, trends, and future perspectives,panelMuhammad Asif Khan a, Hamid Menouar a, Ridha Hamila b https://doi.org/10.1016/j.imavis.2022.104597
[20] E. K. G. D. Ferreira, Guilherme Silveira, Classification and counting of cells in brightfield microscopy images: an application of Convolutional Neural Networks, https://doi.org/10.21203/rs.3.rs-3837227/v1