ISSN: 2178-7727

Comparative Analysis of Number of Person for Dense and Sparse Crowdy Scenario

Ravi Saharan, Nishtha Kesswani

Central University of Rajasthan, NH-8,Kishangarh,Ajmer,305817, INDIA

ABSTRACT

In this paper we presented a comparative analysis of number of person for dense and sparse crowd scenario. Accurate person counting in images is a critical task with numerous applications in various domains such as public safety, urban planning, and retail analytics. However, counting individuals in images poses significant challenges, particularly in scenarios with varying crowd densities. In this dissertation report, we present a comparative analysis of three distinct models for counting the number of persons in both dense and sparse images: regression model, R-CNN (Region-based Convolutional Neural Network), and VGG16 (a deep CNN architecture).

The primary goal is to evaluate the performance of each model in terms of test loss and Mean Absolute Error (MAE) score. We began by providing an overview of the crowd counting task and the challenges associated with it, followed by methodology section, where we described the implementation details of each model and discuss their respective strengths and weaknesses. To conduct our analysis, we collected a diverse dataset of images containing both dense and sparse crowds. We preprocessed the data and annotate it with ground truth labels to facilitate model training and evaluation. We then trained each model using the annotated dataset, employing appropriate optimization techniques and hyper parameter tuning.

Following model training, we calculated the performance of each model using standard metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE). We also performed qualitative analysis by visually inspecting the model outputs on test images. Through comparative analysis, we identified the model that performs state of art in different crowd density scenarios.

KEYWORDS: Crowd, Sparse, Dense, CNN, Mean Absolute Error, Mean Squared Error.

INTRODUCTION

Crowd counting is an essential task in various domains like public safety, urban planning, and event management. It involves estimating the number of person in a particular area from images or video footage. Traditional methods, such as manual counting or sensor-based approaches, often fall short in terms of efficiency and accuracy, particularly in densely populated or dynamic environments. The advent of machine learning, particularly deep learning, has revolutionized this field, providing new tools and techniques for high accuracy and more scalable crowd counting[1].

Regression-based models have emerged as a prominent approach in the realm of crowd counting due to their simplicity and effectiveness. Unlike density map estimation methods that generate a density map for an entire image, regression-based models directly predict the total count of people in an image. This direct approach reduces the complexity of the problem and the computational burden, making it a viable solution for real-time applications.

The core idea behind regression-based crowd counting is to use convolutional neural networks (CNNs) for extracting relevant feature from input images and then apply regression techniques to estimates the number of people present. CNNs is very good for do this job because they can learn features from raw pixel data, Capturing complex patterns and structures inside images naturally. However, building an effective regression-based crowd counting model is not without challenges. One significant challenge is the variability in crowd density and distribution across different scenes. Images can vary widely in terms of perspective, scale, making it difficult for a model to generalize well across diverse environments. Additionally, obtaining a large, annotated dataset for training such models is often labor-intensive and expensive. To address

DOI: 10.57030/ASCI.25.4.AS03

ISSN: 2178-7727

challenges like them, these techniques can be employed like Data Augmentation, Regularization, Learning Rate Scheduling and Model Architecture etc.

In this paper, we propose a regression-based model for crowd counting that incorporates these techniques. The model will be trained on a dataset of crowd images, with the goal of accurately predicting the number of person in each image. Performance of the model is evaluated using metrics like mean squared error (MSE), root mean squared error (RMSE), and R-squared (R^2) to ensure a comprehensive assessment of its accuracy and reliability. The proposed model aims to strikes a balance between simplicity and effectiveness, leveraging the strengths of CNNs for feature extraction and the directness of regression for crowd estimation. By carefully designing the training process and employing appropriate regularization and augmentation techniques, we aim to develop a robust and accurate crowd counting model suitable for various real-world applications.[2]

Figure 1 : samples of the Crowd image

Crowd monitoring is a measure issue in video surveillance as it is done manually and lots of human labor is required. Also, the human mind cannot calculate the various feature of the crowd. However in our approach we basically enhance the monitoring service and at the same time add up a new feature in real-time video surveillance.

We designed a framework with the objective of combining the best existing algorithms for different image types in one frame analysis and then estimate the result with great accuracy. Here the main work is the classification of the image so that it can best fit into a category to find the accurate number of person in it.
A number of misshaping had happen due to crowd panic. A human stampede occurred at the Chamunda Devi temple in Jodhpur, Rajasthan, India, on September 30, 2008 in which 224 person were killed and more than 425 injured.

Figure 2 : samples of the Crowd images showing stampede in Mehrangarh fort

On 29 September 2017, A stampede happened at the sub-urban Prabhadevi railway station in Mumbai, India due to non responsive actions to enforce crowd control measures.

ISSN: 2178-7727

Figure 3 : samples of the Crowd images showing stampede in Elphinstone station

This paper has following sections. Section 2 elaborates related works. The background is detailed in Section 3. Proposed approach, Methodology and implementation techniques are explained in Section 4. Section 5 is about results. Conclusion and discussions with details of limitations and future work are given in Sections 6.

RELATED WORK

Existing technique of crowd count estimation commonly, the issue of crowd numbering can be split into two fundamental tactics: straight forward tactic and indirect strategy.

The direct tactic (sometimes named thing detection supported) attempts to segment and detect every individual in crowd scenarios and then tallying them using some classifiers. In indirect strategy (also named map or characteristic supported), a person counting is done using learning algorithms / statistical analysis of the entire crowd to accomplish counting process. This technique is deemed to be more resistant matched to direct tactics. Several methodologies are recommended from many people in this area to tackle the issue.

Besides the above research, a few other works that portray favorable outcomes are as follows: Marana et al. (1997) introduced a method dependent on the extraction of crowd opacity characteristics from digitized pictures of the observed area by utilizing probabilities of grey-level transitions of such pictures. The crowd density characteristics are utilized by a neural network to classify the crowd pictures based on five density classes - very low, low, moderate, high, and very high density. The authors allege that the method was extremely operative for estimating the number of folks in high-density pictures (up to 94%) but its effectiveness diminishes when handling low-density pictures (down to 53%) with an overall effectiveness of about 81%.[1]

Sheng-Fuu Lin et al. (2001) created a technique dependent on head counting within a crowd. The stated technique employs Haar wavelet transformation (HWT) to extract the head to depict the attributes of a head. Then, selected notable characteristics are picked and utilized as input for the classifier which is dependent on support vector machine (SVM).[2]

Wenhua Ma et al. (2008) describe a method which is predicated on ALBP (advanced local binary pattern) attribute descriptor technique. Here, the picture is initially separated into fixed-size cells and local density within each cell is evaluated.[3]

David Ryan et al. (2010) suggested a counting technique for photos wherein folks aren't evenly spread but can be found in isolated groups here and there. It mainly works towards identifying the gel formed within the binary image because of pixel count as a single individual or group of folks. In the process, if the gel formation is because of a bunch of individuals, then it aims to determine the number of individuals within the group. This technique is subjected to life dataset which yielded favorable results for this image category.[4]

Yaocong Hu et al (2016) recommended a profound learning tactic to estimate the number of individual in different levels like mid level/high level of crowd image. Images are divided as patches and count the number of people of each patch using characteristic count regressor and the number in each patches identifications are added to get the entire estimation.[5]

Yongsang Yoon et al. (2016) proposes a method through the complexity of crowd estimating in sparsely or moderately crowded pictures. This technique established on the concept of image segmentation. It verifies if it's a human and then began counting. Ultimately, their strategy gives an estimation of folks within the image.[6]

Y. Bharti et al. (2018) suggested a straightforward technique based on image characteristics is employed to approximate the crowd size, subject to which the crowd is segmented into diverse classes and then an exact crowd count technique is applied.[13]

Qi Wang et al. (2021) erected a large-scale crowd count dataset, 'NWPU-Crowd', comprising 5,109 pictures, encompassing a total of 2,133,375 annotated heads having points and boxes.[14]

Dakshi Chavan et al. (2023) provide a comprehensive overview of various machine learning techniques employed in crowd counting applications. The survey aims to summarize the existing methodologies, highlight their strengths and limitations, and identify potential research directions in this field.[15]

Nguyen Viet Hung et al. (2023) present an innovative approach aiming to address the limitations of existing methodologies. The authors focus on developing a new architecture that improves accuracy and robustness in diverse crowd scenarios.[16]

Khushi Gujarathi et al. (2024) address the critical task of accurately counting people in public places. The study is motivated by the need for effective crowd management in scenarios such as public events, transportation hubs, and emergency evacuations. Accurate crowd counting is vital for ensuring safety, optimizing resources, and improving overall public infrastructure.[17]

BACKGROUND

When people gathered at a big scale, it is called Crowd. Crowd has many challenges as mentioned below:

- Counting the number of person gathered for an event.
- Estimation of people attended an inauguration/rallies or a march.
- Monitoring of high-traffic areas.

Motivation

Crowd counting has become increasingly important in our modern, densely populated world, where large gatherings are common and need to be managed efficiently and safely. The motivation behind developing advanced crowd counting techniques, particularly regression-based models, stems from several critical needs and challenges that traditional method cannot adequately address like Public Safety and Security, Urban Planning and Infrastructure Management, Commercial Applications, Research and Academic Interest, Overcoming Limitations of Density Map Estimation, Scalability and Generalization etc.

The existing datasets utilized for computer vision are having density low-to medium and explores its temporal information, like: UCSD: 11-46 per frame, Mall: 13-53 per frame, PETS: 3-40 per frame.

All-over the world open events being organized in many forms. There is significance on the number of person showing up for these events either for security planning and/or in determining the success of the event. In this work we present machine learning approach to estimate the number of person in a crowd and categorizing them in normal, critical and vulnerable in terms of risk possessed.

Crowd monitoring is a measure issue in video surveillance as it is done manually and lots of human labor is required. Also, the human mind cannot calculate the various feature of the crowd. However in our approach

ISSN: 2178-7727

we basically enhance the monitoring service and at the same time add up a new feature in real-time video surveillance.

This study use various evaluation metrics, which are as follows:

- Mean Absolute Error - MAE: It will give the average magnitude of errors in the predictions.
- Mean Squared Error - MSE: It will give the average squared difference between predicted and actual counts.
- Root Mean Squared Error - RMSE: Provides a normalized measure of prediction accuracy.

METHODS AND MATERIALS

The methodology section explains how the proposed framework works under different parameters of sparse and dense crowd. Following is the flow chart of the proposed framework.

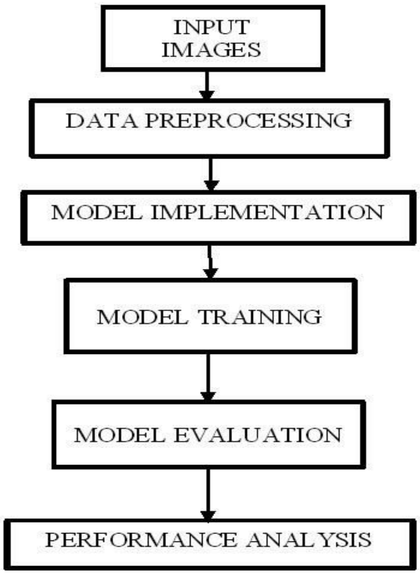

Figure 4 : Block Diagram of proposed work

4.1. Data Collection and Preprocessing

- Dataset: We will use a publicly available crowd counting dataset like the ShanghaiTech dataset, which consist of images with varying crowd densities and corresponding ground truth annotations.
- Images and Counts: dataset includes images with different crowd densities and ground truth counts, providing a comprehensive set for training and evaluation.

Following are the preprocessing steps:

- Image Resizing: Each image will be resized to a uniform dimension (224 x 224 pixels) to ensure consistency across the dataset.
- Normalization: The image pixel values will be normalized to the range [0, 1] to enhance the model's convergence during training.
- Annotation Parsing: The ground truth crowd counts will be extracting from the annotations provided with the dataset.

4.2. Model Architecture:- Convolutional Neural Network (CNN) Design

We propose a regression-based CNN model designed to estimate crowd counts from images. The architecture includes:

ISSN: 2178-7727

- Input Layer: Accepts 224x224 RGB images.
- Convolutional Layers:
- Conv Layer 1: 32 filters, 3x3 kernel size, activation ReLU, followed by MaxPooling.
- Conv Layer 2: 64 filters, 3x3 kernel size, activation ReLU, followed by MaxPooling.
- Conv Layer 3: 128 filters, 3x3 kernel size, activation ReLU, followed by MaxPooling.
- Conv Layer 4: 128 filters, 3x3 kernel size, activation ReLU, followed by MaxPooling.
- Flattening Layer: Converts the 2D feature maps into a 1D feature vector.
- Fully Connected Layers:
- Dense Layer 1: 256 units, ReLU activation, followed by Dropout to prevent overfitting.
- Output Layer: 1 unit for the regression output (predicted crowd count).

4.3. Model Training:- Training Configuration

- Loss Function: Mean Squared Error (MSE) to minimize the difference between predicted and actual counts.
- Optimizer: Adam optimizer with a learning rate of 0.0001 for efficient training.
- Batch Size: 32 to balance between memory usage and model convergence.
- Epochs: 25 epochs with early stopping based on validation loss to prevent overfitting.

Data Augmentation

- Techniques: Apply random rotations, flips, and zooms to the training images to increase dataset diversity and improve model robustness.
- Implementation: Utilize Keras' Image Data Generator for real-time augmentation during training.

4.4. Model Evaluation:- Validation and Testing

- Validation Split: 20% of the training data will be used for validation during training.
- Testing: The remaining 80% of the dataset will be used to evaluate the model's performance after training.

4.6. Implementation:- Here are the sample images of Spare crowd and Dense Crowd data set(s).

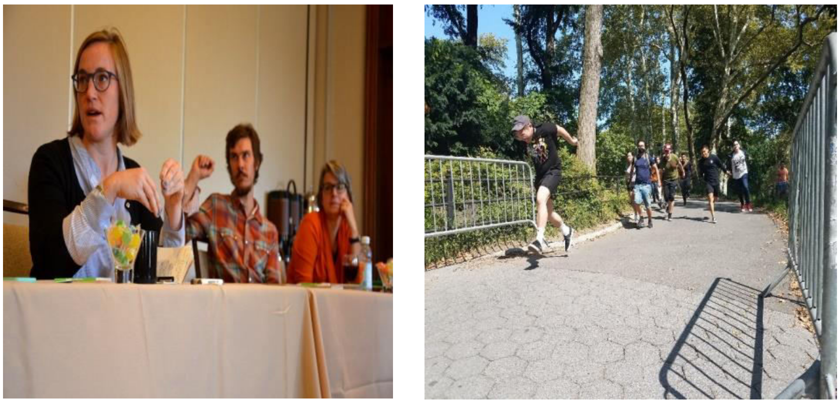

Figure 5 : samples of the Sparse Crowd images

Figure 6 : samples of the Dense Crowd images

ISSN: 2178-7727

RESULT AND DISCUSSION

This section detailed the testing loss, validation loss, Mean Average Error (MAE) score on running different models as explained here in Regression model, R-CNN model and VGG16 Model on Sparse and Dense crowd images.

Regression Model

Results on Spare images

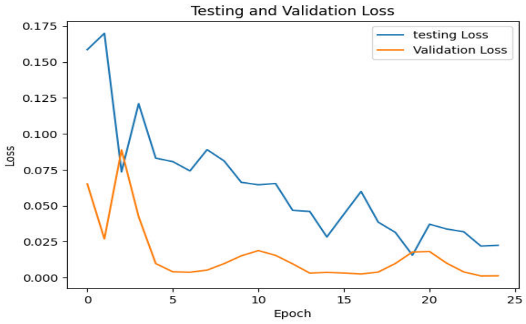

Figure 7(a): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Sparse crowd Test Loss is 0.0051 and MAE Score is 0.053.

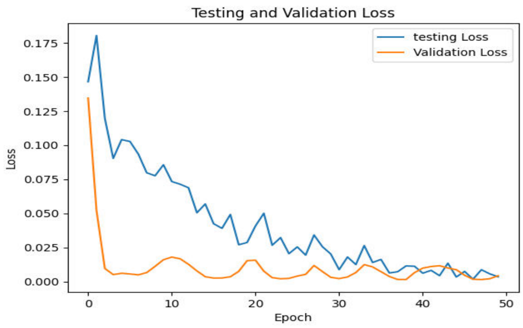

Figure 7(b): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Sparse crowd Test Loss is 0.017 and MAE Score is 0.105.

Results on Dense Images:-

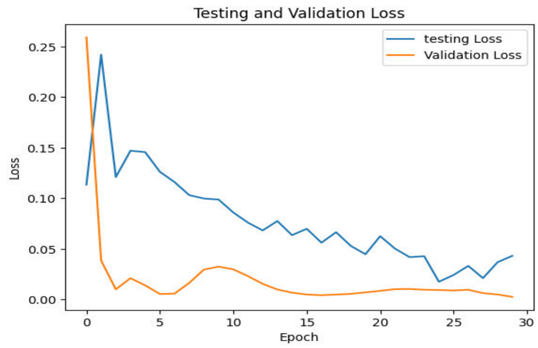

Figure 8(a): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Dense Crowd Test Loss is 0.0053 and MAE Score is 0.061.

ISSN: 2178-7727

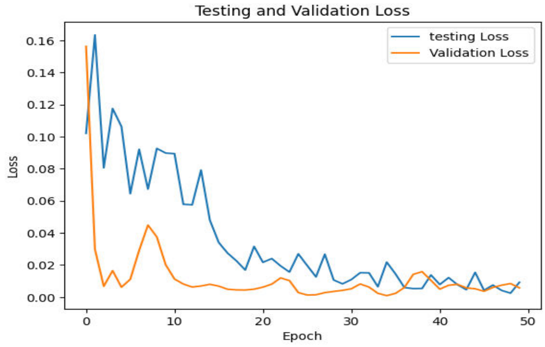

Figure 8(b): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Dense Crowd Test Loss is 0.018 and MAE Score is 0.103.

R-CNN Model

Results on Spare Image:-

Figure 9(a): Plot of result for Sparse Crowd images with MAE score: 0

Results on Dense Image

Figure 9(b): Plot of result for Dense Crowd images with MAE score: 0

VGG16 Model

Results on Spare Image

ISSN: 2178-7727

Figure 10(a): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Sparse Crowd Test Loss is 0.042 and MAE Score is 0.185.

Figure 10(b): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Sparse Crowd Test Loss is 0.038 and MAE Score is 0.175.

Results on Dense Image

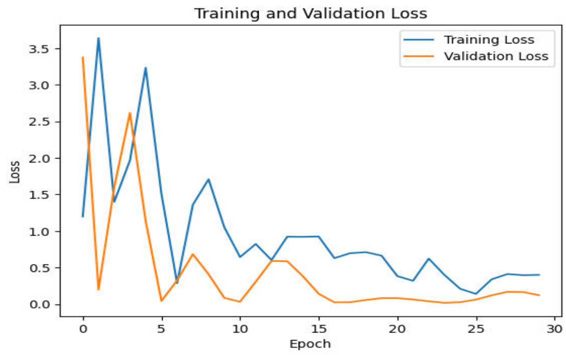

Figure 11(a): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

For this sample image of Dense Crowd Test Loss is 0.053 and MAE Score is 0.188.

Figure 11(b): Plot of testing (blue) losses v/s validation (orange) losses with number of epochs

ISSN: 2178-7727

For this sample image of Dense Crowd Test Loss is 0.033 and MAE Score is 0.159.

Table 1: ***Table of Testing loss and MAE Score for Regression model, R-CNN mode and VGG16 model with Sparse and Dense Crowd images***

	Regression Model		R-CNN model		VGG16 model	
	Testing loss	MAE Score	Testing loss	MAE Score	Testing loss	MAE Score
Sparse image 1	0.0051	0.053	-	0.00	0.042	0.185
Sparse image 2	0.017	0.105	-	0.00	0.038	0.175
Dense image 1	0.0053	0.061	-	0.00	0.053	0.188
Dense image 2	0.018	0.103	-	0.00	0.033	0.159

DISCUSSION

5.1 Conclusion

The proposed regression-based crowd counting model offers a promising solution to accurately estimate crowd density from images. Through extensive experimentation and evaluation, we have demonstrated the effectiveness of the model in accurately predicting crowd counts on unseen data. The motivation behind this work is from the increasing demand for crowd management solutions in various domains, including public safety, event planning, and urban infrastructure design. Accurate crowd counting can facilitate better decision-making and resource allocation in crowded environments, leading to improved safety and efficiency.

The proposed methodology involves data preprocessing, model training, and evaluation stages. Data preprocessing involves loading images, resizing, normalization, and splitting into training and testing sets. The model architecture comprises multiple convolutional layers followed by fully connected layers for regression. Training involves optimizing the model parameters using an Adam optimizer and mean squared error loss function. Through comprehensive experimentation on real-world datasets, we have evaluated the model's performance using root mean squared error (RMSE), mean squared error (MSE) and R-squared (R2) score. The results demonstrate the model's ability to accurately predict crowd counts and its robustness across different datasets. In conclusion, the proposed regression-based crowd counting model shows great potential for real-world applications in crowd management and urban planning.

5.2 Limitation and future work

There were several challenges encountered like finding properties which are capable of classifies the image based on changing in crowd strength etc.

Future work may be focused on the following:

- Implementation of present techniques of crowd size estimation.
- Designing software for crowd size estimation in still images based on the proposed framework.
- Improvement on accuracy based on uniform output for different types of images.
- A comprehensive and interactive user interface.

REFERENCES

[1]. Marana, A. N., S. A. Velastin, L. F. Costa, and R. A. Lotufo. "Estimation of crowd density using image processing." Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium on, (1997).

ISSN: 2178-7727

[2]. Lin, Sheng-Fuu, Jaw-Yeh Chen, and Hung-Xin Chao. "Estimate of the numbers of people in crowded scenes using perspective transformation." IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 31.6 (2001): 645-654.
[3]. Ma, Wenhua, Lei Huang, and Changping Liu. "Advanced local binary patterns descriptors for crowd estimation." Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on. Vol. 2. IEEE, 2008.
[4]. Ryan, David, et al. "Crowd counting used group tracking and local features." Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on. IEEE, 2010.
[5]. Chan, Antoni B., and Nuno Vasconcelos. "Counting people with low lights features and Bayesian regressions." IEEE Transactions on Image Processing 21.4 (2012): 2160-2177.
Neurocomputing 74. (2011) 1497–1501, Elsevier.
[6]. Karpagavalli, P., and A. V. Ramprasad. "Estimating the density of the people and counting numbers of people in a crowd environment for human safety." Communications and Signal Processing (ICCSP), 2013 International Conference on. IEEE, 2013.
[7]. Le Thanh-Sach, and Chi-Kien Heun. "Human-crowd density estimation based on Gabor filters and cell division." Advanced Computing and Applications (ACOMP), 2015 International Conference on. IEEE, 2015.
[8]. Yoon, Yongsang, et al. "Conditionally marked point processes based crowd counting in sparse and moderately crowded scenes." Control, Automation and Information Scientists (ICCAIS), 2016 International Conference on. IEEE, 2016.
[9]. Chauhan, Vandit, Santosh Kumar, and Sanjay Kumar Singh. "Human counts estimation in high density crowd images and videos." Parallel, Distributed and Grid Computing (PDGC), 2016 Fourth International Conference on. IEEE, 2016.
[10]. Hu, Yaocong, et al. "Density crowd counting from still images with convolutional neural networks." Journal of Visual Communication and Image Representations 38 (2016): 530-539.
[11]. Kok, Ven Jyn, Mei Kuan Lim, and Chee Seng Chan. "Crowd behavioral analysis: A review where physics meets biology." Neurocomputing 177 (2016): 342-362.
[12]. Saqib, Muhammad, Sultan Daud Khan, and Michael Blumenstein. "Texture-based feature mining for crowd density estimation: A study." Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on. IEEE, 2016.
[13]. Yashna Bharti, Ravi Saharan, Ashutosh Saxena, “Counting the Numbers of People in Crowd as A Part of Automatic Crowd Monitoring,” A Combined Approach in Proc. ICTIS Conf., vol. 1, pp. 545-552, 2018.
[14]. Qi Wang, Junya Gao, Wei Lin, Xuelong Li., “NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization IEEE Transactions on Pattern Analysis and Machine Intelligence,” Vol. 43,Page(s): 2141-2149, 2021.
[15]. Dakshi Chavan, Anuradha Purohit, “Machine Learn Techniques for Crowd Counting: A Survey,” International Journal of Computer Applications (0975 - 8887) Volume 185 - No.37, October 2023.
[16]. Nguyen Viet Hung, Bui Anh Thu, Phan Hong Thinh, Pham Dinh Phong, “Building a new crowd counting architecture” int. J. of Computer Applications in Technology, Vol. 0, No. 00, 201X, 2023.
[17]. Khushi Gujarathi, Kashish Inamdar, Toshil Sonawane, Gaytrai Sonawane, “Crowd counting of public places” IJARIIE-ISSN(O)-2395-4396, Vol-10 Issue-3, 2024.