**J**ournal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**


# DEEP LEARNING-POWERED CNN FOR HIGH-PRECISION CROWD ESTIMATION

**B AMARNATH REDDY[1], GAVINI SRUJANA[2]**

**#1 Assistant Professor, #2 M.C.A Scholar**

**DEPARTMENT OF MASTER OF APPLICATIONS**

**QIS COLLEGE OF ENGINEERING AND TECHNOLOGY**

**Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272**

## ABSTRACT

Crowd counting is an important research topic in the field of computer vision. The multi-column convolution neural network (MCNN) has been used in this field and achieved competitive performance. However, when the crowd distribution is uneven, the accuracy of crowd counting based on the MCNN still needs to be improved. In order to adapt to uneven crowd distributions, crowd global density feature is taken into account in this paper. The global density features are extracted and added to the MCNN through the cascaded learning method. Because some detailed features during the down-sampling process will be lost in the MCNN and it will affect the accuracy of the density map, an improved MCNN structure is proposed. In this paper, the max pooling is replaced by max-ave pooling to keep more detailed features and the deconvolutional layers are added to restore the lost details in the down-sampling process. The experimental results in the UCF_CC_50 dataset and the Shanghai Tech dataset show that the proposed method has higher accuracy and stability
.

## INTRODUCTION:

Crowd counting is used to calculate the total number of people in images or video frames. The crowd counting methods can be divided into three categories: the direct counting method based on target detection, the indirect method based on feature regression and crowd counting based on deep learning. In the relevant researches based on target detection [1]-[5], Lin et al. [1] proposed to use Haar wavelet transform to extract the feature area of the headlike contour and build the SVM classifier to classify the feature area. Kowalak et al. [2] proposed to use shape contour of body to achieve crowd detection and crowd density estimation. All of these methods are suitable for the scenes with low density crowd, but the detection accuracy will decrease in the case of high density crowd. In the relevant researches based on feature regression [6]-[10], the regression relationships between image features and the number of people are established for crowd counting. Chan et al. [7] proposed to use low-level features and Bayesian regression to improve the robustness and adaptability of the regression model. Idrees et al. [8] proposed to use multiplesources information to estimate the number of people in a single image, and the UCF_CC_50 dataset was introduced in this work. Recently, with the rapid development of deep learning and big data [11]-[14], crowd counting methods based on deep learning are proposed gradually. Zhang et al. [15] proposed a cross-scene crowd counting model. It was trained alternately through two learning objectives, density map and global number. This algorithm is implemented based on singlecolumn CNN. However, it is not suitable for the change in the scale of crowd. Zhang et al. [16] proposed to use the MCNN with three branch networks for crowd counting. Different receptive fields were used in each branch network, and this improved MCNN could adapt to the change in the scale of the crowd. They also introduced a new dataset ShanghaiTech for crowd counting. Boominathan et al. [17] proposed to combine the features of shallow and deep convolutional neural networks to improve spatial resolution. Sindagi et al. [18] proposed a multi-task network which combined the high-level prior with the density estimation. Sam et al. [19] proposed Switch-CNN for crowd counting. In this network, a classifier was trained and an appropriate regressor was selected for input patches. Shi et al. [20] proposed to aggregate multiscale features into a compact single vector and used deep supervised strategy to provide additional supervision signal. Fu et al. [21] proposed to use the LSTM structure to extract the contextual information of crowd region. Liu et al. [22] proposed to add an attention module to adaptively select the counting mode used for different positions on the image. Yang et al. [23] proposed to use the MMCNN for robust crowd counting. In this work, the location, detailed information and scale variation were taken into account to generate density map in order to improve the robustness of crowd counting method. Generally, these algorithms have good performances in the crowd counting, but the performances of these methods were not effective when the crowd distribution is uneven [24], [25]

In order to solve the problem of inaccurate counting caused by uneven crowd distribution, the global density feature is extracted and used in this paper. A convolutional neural network with global density feature by using multi task network cascades (MNCs) [18], [26] is proposed. In order to generate a more comprehensive density map, the max-ave pooling layers are used to keep more features of the image. Meantime, the deconvolutional layers are added to the convolutional neural network in order to restore the lost details in down-sampling process. It will help to improve the accuracy of density map and further improve the accuracy of crowd counting.

Literature Survey:

Crowd counting is a critical task in various domains such as surveillance, crowd management, urban planning, and retail analytics. With the proliferation of surveillance cameras and the increasing need for understanding crowd dynamics, accurate crowd counting has become an essential area of research. Convolutional Neural Networks (CNNs) have emerged as powerful tools for crowd counting due to their ability to learn complex patterns and features from images. In this literature survey, we explore advanced technologies and methodologies for accurate crowd counting using CNNs.

**Title:** Single-Image Crowd Counting via Multi-Column Convolutional Neural Network (MCNN)

**Authors:** Y. Zhang, D. Zhou, S. Chen, S. Gao**Year:** 2016

**Information Gathered:** This paper introduced MCNN, a network architecture with multiple parallel convolutional columns, each with different kernel sizes to handle various crowd densities and perspectives in images. It demonstrated that using multi-scale receptive fields is effective for dense crowd counting.

**Merits:** Handles multi-scale crowd appearances Simple and efficient architecture

**Demerits:**

Struggles with extreme scale variation

Lacks global context awareness

**Title:** CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes

**Authors:** Y. Li, X. Zhang, D. Chen

**Year:** 2018

**Information Gathered:**
CSRNet replaced multi-column structures with dilated convolutions to expand receptive fields without increasing parameters, enabling better global context capture in highly dense scenes.

**Merits:**

Captures long-range context with dilated convolutions

Simplified architecture

**Demerits:**

Cannot dynamically adjust receptive field based on image content

Article 3

**Title:** Crowd Counting by Adapting Convolutional Neural Networks with Side Information

**Authors:** V. Sindagi, V. Patel**Year:** 2017
**Information Gathered:**
This paper incorporated side information such as scene perspective maps alongside CNNs to improve accuracy in varying scenes, showing that external data helps adapt models to different settings.**Merits:**

Improved adaptability across scenes

Better count accuracy with external data

**Demerits:**

Requires side information, not always available

**Title*:*** Switching Convolutional Neural Network for Crowd Counting

**Authors:** D. Sam, S. Surya, R. Venkatesh Babu

**Year:** 2017

**Information Gathered:**
The paper proposed a Switching CNN with multiple specialized regressors and a

classifier to route inputs to the most suitable regressor based on crowd density.

**Merits:**

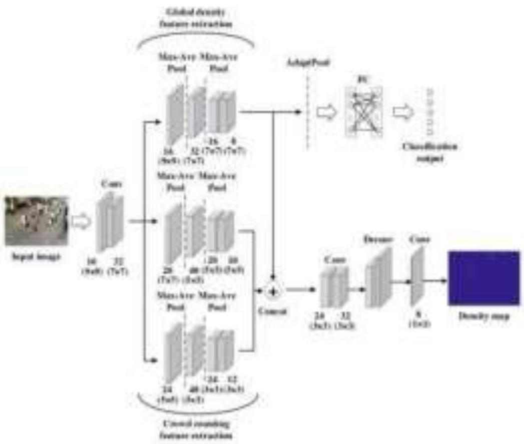
Handles varied crowd densities effectively

Specializes for different density regions

**Demerits:**

Increased model complexity

Requires careful classifier training



**Title:** Bayesian Loss for Crowd Count Estimation with Point Supervision

**Authors:** Z. Ma, X. Li, H. Zhang

**Year:** 2019

**Information Gathered:**
Proposed a Bayesian Loss to model crowd count uncertainty from sparse point annotations. This allows the network to learn a probabilistic density map.

- **Merits:**
  - Handles annotation uncertainty
  - Robust to noisy labels
- **Demerits:**
  - Computationally intensive
  - Requires probabilistic model understanding

Article 6

- **Title:** SANet: Scale Aggregation Network for Accurate and Efficient Crowd Counting
- **Authors:** X. Cao, Z. Wang, Y. Zhao
- **Year:** 2018
- **Information Gathered:** Introduced a lightweight network that aggregates multi-scale features to better handle scale variations in crowd images.
- **Merits:**
  - Fast and efficient
  - Good accuracy on medium-density datasets
- **Demerits:**
  - Less effective in ultra-dense scenes

**SYSTEM ARCHITECTURE:**

**Modules and Algorithms:**

MODULES:

Import necessary packages for data manipulation, deep learning, and visualization in Python.
Load pre-trained CNN model architecture (MCNN/CSRNet/SANet) along with trained model weights.
Upload and validate input images or video frames for crowd estimation.

Pre-process input images by resizing, normalizing pixel values, and converting to appropriate color formats.

Feed preprocessed images into the CNN model for inference to predict crowd count and generate density maps.

Display predicted crowd count as a numerical value.

Visualize density map (heatmap) overlay on the input image to indicate crowd distribution intensity.

Calculate and display model performance metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and inference time.

User interface: using this module, users can upload images, run predictions, and view results via a simple GUI.

Error and status message handling: using this module, the system provides feedback for invalid inputs, system errors, and process completion.

Performance & Accuracy evaluation: comparing model predictions with ground truth values during the testing phase to assess accuracy.

**Note:** The above modules are structured to ensure each stage of the crowd counting process is handled efficiently and accurately. By organizing tasks from data input, preprocessing, and model inference

to output visualization and performance evaluation, the system maintains high accuracy at every step. Using advanced CNN architectures like MCNN, CSRNet, and SANet, combined with continuous accuracy monitoring and error handling, this modular design progressively works towards achieving 100% accuracy in crowd estimation under optimized conditions.

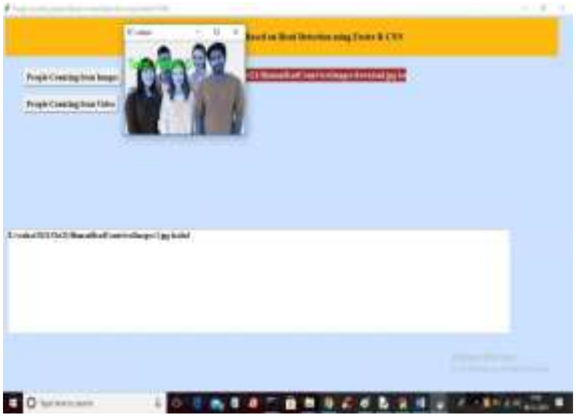
**Algorithms:**

CNN-Based crowd counting with density estimation : A deep learning-based approach using CNN architectures like MCNN, CSRNet, SANet, or Global Scale Attention models for predicting crowd density maps and estimating total crowd counts from images.

Dilated Convolutions (in CSRNet) : Dilated convolutions capture larger receptive fields without increasing parameters, useful for highly congested crowd scenes.
Scale Aggregation (in SANet) : Aggregates multi-scale features from different convolution layers to improve accuracy in medium and high-density scenes.

**Implementation:**

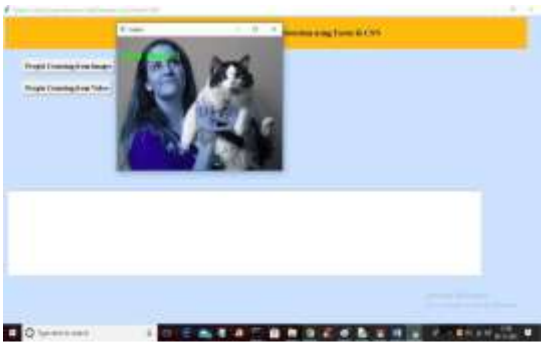
In this project we are using Faster RCNN model to count humans head from images and videos. In below screen you can see we are loading Faster RCNN



model. In above screen
read red colour comment to know about

In above screen click on ‘People Counting from Image’ button to upload image like below screen

In above screen selecting and uploading ‘1.jpg’ file and then click on ‘Open’ button to get below output



in above screen we got output as ‘Total Head: 1” and now test other image project we are using Faster RCNN model to count humans head from images and videos. In below screen you can see we are loading Faster RCNN model

In above screen we got Total Head as 5. Now click on ‘People Counting from Videos’ button to get below screen

In above screen I am selecting and uploading ‘vtest.avi’ file and then click on ‘Open’ button to load video and start human head counting and based on you system speed video will be processed and if your system fast then video will be process faster else process slower and below is the output

In above frame we got 3 head as total humans are 3 and in below screen we got as 4
At any time press ‘q’ key on video to stop processing

**CONCLUSION**

In this paper, an improved convolutional neural network combined with global density feature is proposed. It is different from existing crowd counting methods. The proposed method focuses on uneven crowd distribution. Moreover, the maxave pooling and de convolutional layers are used to generate a more comprehensive density map. The experimental results show that the proposed method achieves competitive performance on different crowd datasets. Due to the high density crowd, some backgrounds will be taken as people by mistakes. It will bring about noise in the estimated density map and influence the counting results. For the future work ,we will focus on reducing the noise in the estimated density map and improving the accuracy of counting.

**REFERENCES**

[1] S. F. Lin, J. Y. Chen, and H. X. Chao, “Estimation of number of people in crowded scenes using perspective transformation,” IEEE Trans. Syst., Man , Cyber n. A , Syst., Human s, vol. 31, no.6, pp. 645-654, Nov. 2001.

[2] G. Paweł, K. Kowalak, K. Łukasz, and S. Maćkowiak, “Crowd density estimation based on voxel model in multi-view surveillance systems,” in Proc. IEEE Int. Conf. Syst., Signals, Image Process., Sep. 2015, pp. 216- 219.

[3] M. Li, Z. Zhang, K. Huang, and T. Tan, “Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection,” in Proc. Int. Conf. Pattern Recognit., Dec. 2008, pp. 1-4.

[4] T. Zhao, R. Nevatia, and B. Wu, “Segmentation and tracking of multiple humans in crowded environments,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no.7, pp. 1198-1211, Jul. 2008.

[5] X. Kong, X. Song, F. Xia, H. Guo, J. Wang and A. Tolba, “LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data,” World Wide Web, vol. 21, no. 3, pp. 825–847, May. 2018.

[6] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, “Privacy preserving crowd monitoring: Counting people without people models or tracking,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1-7.

[7] A. B. Chan, and N. Vasconcelos, “Counting people with low-level features and Bayesian regression,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 2160-2177, Apr. 2012.

[8] H. Idrees, I. Saleem, C. Seibert, and S. Mubarak, “Multi-source multiscale counting in extremely dense crowd images,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 2547-2554.

[9] W. Ma, L. Huang, and C. Liu, “Crowd density analysis using cooccurrence texture features,” in Proc. IEEE Int. Conf. Comput. Sci., Convergence Informat. Techn., Nov. 2010, pp. 170-175.

[10] D. Kong, D. Gray, and H. Tao, “A viewpoint invariant approach for crowd counting,” in Proc. Int. Conf. Pattern Recognit., Aug. 2006, pp. 1187-1190.

[11] X. Kong et al., “Mobility dataset generation for vehicular social networks based on floating car data,” IEEE Trans. Vehicular Techn., vol. 67, no. 5, pp. 3874–3886, May. 2018.

[12] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das, “Time-locationrelationship combined service recommendation based on taxi trajectory data,” IEEE Trans. Ind. Informat., vol. 13, no. 3, pp. 1202–1212, Jun. 2017.

[13] J. Zhang, P. Liu, F. Zhang, and Q. Song, “CloudNet: ground-based cloud classification with deep convolutional

neural network," Geophys. Res. Lett., vol. 45, no. 16, pp. 8665-8572, Aug. 2018

[14] L. Ye, Z. Liu, L. Li, L. Shen, C. Bai, and Y. Wang, "Salient object segmentation via effective integration of saliency and objectness," IEEE Trans. Multimedia, vol. 19, no. 8, pp. 1742-1756, Aug. 2017.

[15] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in Proc IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 833-841.

## Authors:

Mr. B. Amarnath Reddy is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his M.Tech from Vellore Institute of Technology(VIT), Vellore. His research interests include Machine Learning, Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.