Scientific Exploration for Density Estimation and Crowd Counting of Crowded Scene

Juginder Pal Singh1, Manoj Kumar2, Ayush Arya3, Baldev Bharat Badmera5

Department of Computer Engineering & Applications,
GLA University, Mathura, India.

Email: [juginder.singh1, manoj.kumar2, ayush.arya_cs163, baldev.gla_cs164, bharat.badmera_cs165]@gla.ac.in

Abstract. Crowd density estimation model is a typical concept which compute the counting of the people in the crowded image. There are many of research papers are written in this area to solve the different kind of real world problems. This paper shows the state of art and different types of framework proposed on density estimation and crowd counting techniques. The population of our country increases year by year, so the purpose of this paper is to explore different techniques available for crowd density estimation. There are lot of applications existing based on this technique like retail system, surveillance system etc., where the foot traffic is very vital in retail system to organize merchandise in aisles, optimize store layout, understand peak times and potentially even protect against theft. In surveillance system, this is mainly used for security purpose. In this, the researcher captures the crowd using the camera and the technique of counting the crowd which is divided into two parts unsupervised and supervised learning. These techniques are further divided into parts which is shown in the paper. The survey presents the different types of datasets of the crowded image which is useful for simulating the people counting in the crowded scenes. In detection based approach, this technique counts the people with the help of face detection technique. In this researchers face the problems of count the people in crowded scenes and also face the problem in surveillance applications. The cluster based approach data works well on the scattered crowd. This method needs image frames; it does not work on still images. In regression approach the model trained is subject to the point of view map. In the model were to be used in a different scene of a changed perspective map, it will have many inaccuracies in its result. The current research work intends to provide a general idea on crowd density estimation and counting approaches employed in visual surveillance in the perception of the study of computer vision. The analysis of review categorizes and frameworks quite a lot approximation related to crowd density and calculating approaches which are beneficial in providing the analysis related to the crowd places of interest.

Keywords: Crowd-analysis, Crowd-counting, Density-estimation.

1. Introduction
The central persistence of the paper is to estimate the people count in the crowded scene images, in which we calculate the estimated density in the input image by matching the input image to the corresponding image of the crowd counting [1-5]. The problem of evaluation of density is of high
priority significance and it is used significantly for top-level cognitive skill in crowded scenes for monitoring of people and visual awareness among them [6-11] Crowd analysis is the major topic for us in the recent years by the researchers for the various points of view like increasing the world population in the recent years rapidly by this we get the crowded scenes on the different places like temples, airports, parks [5, 22, 23], sporting places, walking places, shopping malls, parties, and there are lot of places where we face the problem to handle the crowd scenes. we estimate the people count in the crowded picture by using different algorithms [9]. Our world’s population increases year by year rapidly and Urbanization has helped for rising activities such as different events like sporting events, political rally, public shows and so on, resulting in more congestion in recent years. In such cases, we analyses that to handle the crowd we need to crowd counting and density estimation for enhancing the security, safety, and better management [24, 25]. In recent years, there are lots of paper are written by researchers. In this paper, we will tell about the different methods and their problems to estimate the people count in the crowded scenes. This is useful in surveillance system; it monitors the crowd and its count, behavior, and activities, etc. For crowd behaviour analysis, Smart sensing devices play a very good role. The researchers are using the signal processing system for achieving a good success but now, sensors play a very good role in all aspects. they are playing a good role in crowd’s whether forecasting visualizing etc. But today, researcher have more focus on sensors and give very accurate value and frequently [12-18]. These thing motivate us for using this type of sensors and finds the crowd behaviour in a very easy way, also, result will be more predictive. It will be useful in surveillance system, security purpose and handling the crowd for the society as shown in the crowded scenes in Figure 1.

2. Crowd Counting
A large amount of people in a place is called a crowd. The crowd counting is a mechanism which is useful for the count of the people in the crowded scenes. There are different types of crowd counting and density estimation techniques [1, 3] which is used for counting the people and handle the crowd [6, 7, 10, 11]. These approaches are mainly divided into two measures, first one is supervised learning crowd counting and the second one is unsupervised learning crowd counting approaches.
Figure 2: Chart of Crowd counting methods

2.1 Unsupervised Learning Techniques for CrowdCounting

Unsupervised Learning is a technique in which we are not required to supervise the model [1]. However, we need to let the machine to work independently for data discovery [4, 5, 6, 7]. It deals and handles the data which is unlabeled. This algorithm permits us to complete complex processing of tasks in contrast to supervised learning algorithm. Although, this algorithm is more unpredictable in comparison to other learning methods. The unsupervised machine can find all types of unknown patterns in the data set like the crowded scene images [10, 11, 12, 22, 23]. Unsupervised methods help us to find out all type of characteristics which is useful for defining the category. It takes place in real-time, therefore, all the input values need to be examined and labeled in accordance with learners. It is simpler to take out unlabeled data than labeled data from computer, which requires manual process.

2.1.1 Cluster Based Approach (Unsupervised Learning)

The cluster-based technique for crowd counting depends on the assumption [4, 5] that all types of visual features and separate motion fields are identical, therefore related characteristics are gathered into several sections. It is a new integrated approach for the people analysis that includes clustering and analyzing the distribution of pixel values within clusters over one or more images [8, 9]. It assumes that the way an individual move [12, 13, 14], or other visual features (such as hand-carry bags or clothing) [16, 17, 20], is comparatively persistent and uniquely based on the trajectory of these features, grouping of cluster is done to represent independently moving entities. In the static position it gives us wrong results estimation or give a repeatedly the same type of data sets. So it clear that it is useful only continuous image data sets for the count the people in the crowded images. It is required for the target to be applied in the continuous movement of the data. The training data which we provide to our machine does not have to be labeled and works well on the sparse crowd. This technique will not work on still images; it needs imageframes.

2.2 Supervised Learning Techniques for CrowdCounting

In supervised learning [1, 2, 3, 4], as the name, it is clear that a supervisor is always present for the guidance [6, 7, 8]. Basically, in this, we train the machine with the help of data like crowded scenes image and the data which and provides well labeled with the correct input and output values so that
technique of supervised learning evaluates the training data (like crowded images) [18, 19, 20, 21] and produces accurate results from labeled data.

### 2.2.1 Detection Based Approach

Detection based counting is defined as a method for calculating the distraction of information of image and local decisions for identification of characteristics of particular type at any instant. An author proposed CNN based hybrid hidden Markov model (HHMM)[3, 5, 9]. This model is used to generate fixed dynamic characteristics that is useful for utilization of detection of exception in analysis of crowd. A resolution for building full-body locomotion will be carried out in analysis of 3D crowd. The research of earlier times generates on this approach to compute the people count in a scene. Common techniques of detection use consistent detection. In this method, a classifier is prepared by taking various features, including Histogram Oriented Gradient (HOG). The way of function detection is so strong in low-density crowds, but its execution decreases in high density. Accordingly, researchers were dedicated to direct this problem by dividing the part based techniques that uses advanced divisions for particular parts of body, consisting of head and shoulders, to compute the area count.

### 2.2.2 Regression Based Approach

Regression based computation is carried out to generate an accurate function with the help of input images and outputs [3, 5]. A method of building a complete movement of the body which can also be encapsulated in analysis of 3D crowd and detection of anomaly is proposed [12, 15, 16, 18, 20]. Davies et al. firstly utilized the regression based crowd density estimation technique. Lower degree characteristics are visualized through video set. With this method, model of linear regression is proposed for building the competency in between the expected vis-a-vis actual computation. The detectors based on the form and shape didn’t result into approximation for the high clutter backgrounds or for high density crowd’s. Low-level feature extraction and regression modeling are the main elements that construct regression counting pipeline. Different characteristics, like foreground slope, edge characteristics, surface consistencies utilized for concealing the lower-grade information. However, standardized techniques useful for background subtraction supports for the mining of foreground characteristics separated from forefront elements.Holistic characteristics, like area, boundary, and boundary–area proportion has defined outcomes. However, [22, 23, 24], these methods concentrate on the global features of the scene. Local characteristics and textures like grey Level Co-Occurrence Metrics (GLCM), HOG, and Local Binary Pattern (LBP), are used to increase the efficiency of detection, classification and counting of crowd. After the reduction of local and global properties, a variety of regression techniques, including Gaussian, linear and ridge regression and NNs are useful for the mapping amid low level characteristics and real crowd counting.

### 2.2.3 Density Estimation Approach

Density estimation computation is proposed to generate estimation with the help of recognized data of a probability-density function which is unobservable [2, 9]. This method has resulted to defeat the problem of confusion and occlusion with the help of spatial learning with an approach of density estimation. Like, spatial learning between local characteristics and estimated density (ED) maps by incorporating linear mapping. The tedious job of recognizing and confining single objects has been reduced by subtracting density of image whose integral gives the expected count of that region in any confined region. Convex optimization jobs are solved by cutting plane optimization method with the help of a risk-based quadratic cost function.

### 2.2.4 Convolutional Neural Network (CNN)

Deep learning has been at the center of attention in current years, deep learning is common that ongoing exploration on crowd counting has moved to deep learning approaches which appear to help favorable results. Because of the idea of neural networks, significant features of the image that can assist with accomplishing the ultimate objective of people count is autonomously discovered during
training. This implies designs that we don't instinctively observe as well as are difficult to manage can also be set up and additionally considered during preparing, which permit the picture to be better spoken to in the system when compared to other methods mentioned Figure 3. it shows the working of CNN which divide the input image into the patchesthen perform pooling operation and after that check the error rate and then perform the merge operation using maps and pass the output result to the density map.

![Diagram](image.png)

Figure 3 Overview of single image crowd counting via multi-column network[39].

In the detection, clustering and crowd counting approaches based on estimation of density perform accurately to some extent, [17, 19, 22, 26] for evaluation of motion and crowd analysis, and the 3D body parts, different methods have been proposed of CNN- and LSTM- based. In particular, the author incorporated LSTM based and CNN based descriptor network to get information regarding appearance and motion with the tracks of parts of human body. Contrarily, with the help of 2D view of the face author constructs a 3D face-model [8, 9]. Further, the evaluation of the deep learning system for the driver actions classification is proposed abstractive summary of text with the help of a generative adversarial network was done and CNN based method to generate high representational characteristics for the generation of secondary protein system. Researchers used CNN based crowd counting methods to generate enhancement for sensors accuracy.

3. Existing approaches overview for Density Estimation on Crowded Image

Table: 1 Existing approaches for crowd counting along their models and dataset
### Authors | Approaches | Dataset | Challenges
--- | --- | --- | ---
Zitouni et al [30] | Motion Flow Model | PETS 2009 | People counting
Shen et al [31] | Light Effect Suppression (LES) | PETS 2009 | Occlusion problem
Fu et al [32] | ConvNet Classifier | PETS_2009 & Subway | Crowd density Estimation
Hashemzadeh et al [33] | SURF Gaussian Regression Model | PETS 2009 | Crowd density estimation
Q.H et al [34] | Count-net Based Model | UCF & AHU-CROWD | Crowd Density Estimation
Liu et al [35] | Crowd motion flow Social Network Model (SNM) | Real World crowd dataset | Crowd behavior
Chakrer et al [37] | End-to-end deep CNN regression model | UCF | Anomaly Detection
Wang et al [38] | Count-net based Model | UCF & AHU-CROWD | Crowd density estimation using deep CNN
Zhang et al [39] | Joint learning of crowd density and velocity | UCF & AHU-CROWD | Crowding background and people's head
Zhao et al. [40] | Conditional Generative Adversarial Networks | UCF | Crowd Counting with Two-Phase Deep Neural Networks
Tao Xu et al [47] | Spatial-/channel-wise attention regression networks | ShanghaiTech, UCF_CC_50 | Crowd counting in high density crowded image
Junyu Gao et al [48] | Compact convolutional neural network | Shanghai Tech Part A/B, GCC, and UCF_CC_50 | Focus on the local appearance features of crowd scenes for crowd counting
Xiaowen Shi et al [49] | Density-aware convolutional neural network (DensityCNN) | ShanghaiTech dataset, WorldExpo’10 dataset | Crowd counting to deal with the lack of real-time performance
Xiaoheng Jiang et al [50] | Density-aware convolutional neural network (DensityCNN) | ShanghaiTech, UCF CC 50, UCF-QNCF and WorldExpo‘10 | Crowd counting task in various crowded scenes

### 4. General problems and evaluation

People counting are for the most part centered around overcrowded areas for both security and safety purposes [27]. Individuals counting or density estimation is a main issue to describe the level of a crowd as dense or sparse. People counting can be applied on static pictures and video frames in both outside and inside scenes. Now days, people counting considering can be organized as: counting by location, checking by grouping and counting by regression [28]. The crowd counting is playing big role for surveillance system for controlling the abnormal crowd [4, 6, 10-12]. A good system should be robust, real-time to directly or efficiently catch the both moving and static crowd in a scene. So, all of the technique those are not performing well, need to improve to satisfy all the properties. There are some general problems that approaches have in giving proper result.

- Detection-based approach performs well for the detection of faces, but less well for foot-traveler because the images of foot-traveler are varied (continuous change in body pose and there clothing) [5]. It also faced problems in crowded scenes were occlusion and scene confusion are inevitable. Even it performs worse in surveillance applications, and for the scene, those have low
resolution.
- Cluster-based data that have used for training does not need to be labeled [5] and works well on the scattered crowd. However, it needs the target to be applied on to have regular motion and if people in a scene imprecision can arise, or objects those share the same trajectories at a time. This method needs image frames; it does not work on still images.
- Regression model trained is dependent to the point of view map [5]. If the model were to be utilized in another scene of an alternate point of view map, it will have numerous erroneousness in its outcome, but whenever tried on a similar scene yet a completely new frame, it will work with fine performance.
- In the neural network, we can transfer the trained model very easily onto another scene with a different sense and people size on the scale. In other words, unlike the models [5, 17, 19, 26] that we have discussed above, it can be used for other scenes. Typically, it can boost accuracy, with some additional combination of the fine layers of the design for the particular view. However, without fine-tuning, it should still give sensible results.

5. Datasets for the Crowd Counting
We have types of the dataset available on the internet these datasets have been created over a few years. It helps the researchers to make modules with better abilities [2, 3, 7, 8]. Previous datasets mainly contain low-density crowded scenes images but now the recent datasets are focusing on high-density crowded scenes images. These types of large datasets motivate to create and develop different methods for better development for the improvement of crowd counting techniques.

5.1 Mall dataset for the Crowd Counting
Mall dataset is the dataset that was produced by Chen et al. in 2012 for crowd counting and density estimation research [2, 3, 8 and 9]. This was created by clicking the images in the shopping malls using the surveillance system camera. This dataset contains 2000 frames collections and a size 320 × 240 from a shopping mall. In this, the first 800 frames are useful for training purposes and the rest 1200 frames are useful for the computation. If we compare Mall dataset with UCSD dataset, then Mall dataset is higher density crowded scenes images.

5.2 UCSD dataset for the Counting of Crowds
Chan et al created this dataset in 2008 for crowd counting and density estimation purpose. This dataset of pedestrians was taken by a video camera UCSD walkaway [2, 3, 10]. This dataset consists of 2000 frames collections and the video camera system of 238x138 sizes. This contains the data which is moving in two different ways which are taken by a camera. The 49,880 pedestrian instances are presented in this dataset which is partitioned into test and training dataset. 6.3 UCF_CC_50 datasets for the Counting the Crowds

5.3 UCF_CC_50 dataset
This is the type of dataset which gives a very high-density image. This dataset has extremely dense crowded images are available [2, 3, 8]. This dataset was taken by the web images (public images) which are available publicly [10]. Mainly the Flickr is the main source where we collect the images. This dataset only takes 50 images. In each image the average 1280 individual for count person range between 94 and 4543. In this, the persons are indicated by the dots and in these maps are not provided for denoting the persons. In this, there is a challenging problem, in which, there are limited numbers of images accessible for the training images.
5.4 Grand Central Dataset for CrowdCounting
Grand Central Dataset, in 2012 for the crowd counting and sees the behavior of the crowd [44][46]. It was generated from Grand Central Station of New York [3]. This data set was taken like a grey scale video. This dataset contains 50010 frames; it has 25fps frame rate and 720 × 480 size.

5.5 WorldExpo ’10 dataset for CrowdCounting
In 2010, this dataset was created by Zhang et al in [45-46] which is made with the help of 108 surveillance cameras that consist of 1,132 video flows at the time of Shanghai Expo in 2010[3, 5 and 10]. 3380 annotated frames are present in the training of 103 scenes. The dataset which is for cross crowd counting scene introduced by [45], a total of 3980 frames is available with the 199923 labeled walkers in the dataset of size 576 × 720. The dataset of crowd counting are divided into 2 sections in which training set contain1, 127-minute video in the 130 crowded areas and testing sets contains video clip which is 5 hours long. 120 labeled frames are present in every scene.

5.6 Shanghai Tech Dataset for CrowdCounting
The dataset which is used for the crowd counting system was introduced by Zhang [45]. This data type consists of two types Part A and Part B in which there are a total of 1198 crowded images with 330,165 people heads. In Part A there are 482 images in which any picture is selected randomly and another part contains those images which are taken from the urban areas. There is a total of 716 images available in Part B [9, 10, 12]. Part A has large density crowded images available in comparison to the other part. Part A consists of 300 images for the training and 182 images for the testing and Part B consists of 400 images for the training and 316 images are for the testing. In this, we can perform a large number of tests on the image because of the varying on the density level. There are some sample images of the different datasets shown in Figure 4 (A) Mall Dataset, (B) UCSD Dataset, (C) UCF_CC_50 datasets, (D) Grand Central Dataset, (E) WorldExpo ’10 dataset (F) Shanghai Tech Dataset.
Figure 4: some datasets of the images which is used to count the number of people in image. In
which (a) show the dataset of Mall created by Chan et al [41] (b) show the UCSD dataset created by Chan et al [42] (c) show the dataset of UCF_CC_50 [43] (d) Show the Grand Central dataset created by Zhou et al [44] (e) show the WorldExpo ’10 dataset created by Zhang et al [45] (f) show the dataset Shanghai Tech created by Zhang [46].

6. Results and discussion

In few last years, various data sets have been created. Previously, data sets were containing low density crowded scenes data sets. Currently, researchers are focusing on high density crowd as multiple posing challenges like scale variations, clutter and severe occlusion. Large scale datasets creation has encouraged recently approaches to develop methods which involves this type of challenges. This paper, discusses six types of datasets [2, 3, 5, 7, 8, 10]. In table 4, includes the different datasets used by different authors with different approaches. In this paper, includes the Mall dataset, UCSD dataset, UCF_CC_50 datasets, Grand Central dataset, WorldExpo ’10 datasets, Shanghai Tech Dataset. Mall datasets made by 2000 frames collections in a shopping mall which is working on high density dataset, UCSD have 2000 frames collection, this dataset which is partitioned into test and training dataset, UCF_CC_50 dataset contains 2000 frames which is working on high density datasets, Grand Central data contains 50010 frames which is generated from Grand Central Station of New York. WorldExpo ’10 contains 3980 frames, 120 labeled frames are present in every scene. Shanghai Tech datasets contains 1198 crowded images with 330,165 people heads.

7. Summary and Future Scope

In this survey, tried to generate a study on the crowded scene images using estimation of density and people counting techniques on crowded scene image which will help in surveillance application. Here, mainly two different approaches are discussed that are supervised and unsupervised summarization of various traditional and CNN based approach has been done and then further CNN divides in two different categories. Obviously, it is not easy to cover all the literature on crowd counting so latest paper has been chosen to work on crowd counting which help us in review and analysis in detail. It is done in order to identify the common lacunae of the existing techniques and to overlay a path for the further research in this area. We have also reviewed the different result from different traditional approach have also been reviewed where it is found that CNN technique has pretty nice results in comparison to other approaches.

References
[1] Carlotto, M. J. (2005). A cluster-based approach for detecting man-made objects and changes in imagery. IEEE Transactions on Geoscience and Remote Sensing, 43(2), 374-387.
[2] Lamba, S., & Nain, N. (2017). Crowd monitoring and classification: a survey. In Advances in Computer and Computational Sciences (pp. 21-31). Springer, Singapore.
[3] Dong, B., Yan, D., Li, Z., Jin, Y., Feng, X., & Fontenot, H. (2018, October). Modeling occupancy and behavior for better building design and operation—A critical review. In Building Simulation (Vol. 11, No. 5, pp. 899-921). Tsinghua University Press.
[4] Li, Yuhong, Xiaofan Zhang, and Deming Chen. "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1091-1100. 2018.
[5] Liu, L., Wang, H., Li, G., Ouyang, W. and Lin, L., 2018. Crowd counting using deep recurrent spatial-aware network. arXiv preprint arXiv:1807.00601.
[6] Xu, Mingliang, Zhaoyang Ge, Xiaoheng Jiang, Gaoge Cui, Pei Lv, Bing Zhou, and Changsheng Xu. "Depth information guided crowd counting for complex crowd scenes." Pattern Recognition Letters 125 (2019): 563-569.
[7] Sam, Deepak Babu, Shiv Surya, and R. Venkatesh Babu. "Switching convolutional neural network for crowd counting." In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4031-4039. IEEE, 2017.
[8] Sam DB, Sajjan NN, Maurya H, Babu RV. Almost unsupervised learning for dense crowd counting. In Proceedings of the AAAI Conference on Artificial Intelligence 2019 Jul 17 (Vol. 33, pp. 8868-8875).

[9] Bour, Pierre, Emile Cribelier, and Vasileios Argyriou. "Crowd behavior analysis from fixed and moving cameras." In Multimodal Behavior Analysis in the Wild, pp. 289-322. Academic Press, 2019.

[10] Zhang, Youmei, Chunluan Zhou, Faliang Chang, and Alex C. Kot. "Multi-resolution attention convolutional neural network for crowd counting." Neurocomputing 329 (2019): 144-152.

[11] Bharti, Yashna, Ravi Saharan, and Ashutosh Saxena. "Counting the number of people in crowd as a part of automatic crowd monitoring: A combined approach." In Information and Communication Technology for Intelligent Systems, pp. 545-552. Springer, Singapore, 2019.

[12] Ryan, David, et al. "An evaluation of crowd counting methods, features and regression models." Computer Vision and Image Understanding 130 (2015): 1-17.

[13] Li M, Zhang Z, Huang K, Tan T. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In2008 19th International Conference on Pattern Recognition 2008 Dec 8 (pp. 1-4). IEEE.

[14] Zhu, L., Li, C., Yang, Z., Yuan, K. and Wang, S., 2019. Crowd density estimation based on classification activation map and patch density level. Neural Computing and Applications, pp.1-12.

[15] Loy, C. C., Chen, K., Gong, S., & Xiang, T. (2013). Crowd counting and profiling: Methodology and evaluation. In Modeling, simulation and visual analysis of crowds (pp. 347-382). Springer, New York, NY.

[16] Leggett, R. (2004). Real-time crowd simulation: A review.

[17] Rodriguez, M., Laptev, I., Sivic, J., & Audibert, J. Y. (2011, November). Density-aware person detection and tracking in crowds. In 2011 International Conference on Computer Vision (pp. 2423-2430). IEEE.

[18] Hu, Yaocong, Huan Chang, Fudong Nian, Yan Wang, and Teng Li. "Dense crowd counting from still images with convolutional neural networks." Journal of Visual Communication and Image Representation 38 (2016): 530-539.

[19] Han, Kang, Wanggen Wan, Haiyan Yao, and Li Hou. "Image crowd counting using convolutional neural network and markov [random field." Journal of Advanced Computational Intelligence and Intelligent Informatics 21, no. 4 (2017): 632-638.

[20] Wang, Li, Wei yuan Shao, Yao Lu, Hao Ye, Jian Pu, and Yingbin Zheng. "Crowd counting with density adaption networks." arXiv preprint arXiv:1806.10040 (2018).

[21] Fradi, H., & Dugelay, J. L. (2015). Towards crowd density-aware video surveillance applications. Information Fusion, 24, 3-15.

[22] Bansal, A., & Venkatesh, K. S. (2015). People counting in high density crowds from still images. arXiv preprint arXiv:1507.08445.

[23] Boominathan, Lokesh, Srinivas SS Kruthiventhi, and R. Venkatesh Babu. "Crowd net: A deep convolutional network for dense crowd counting." Proceedings of the 24th ACM international conference on Multimedia. 2016.

[24] Zeng, Lingke, Xiangmin Xu, Bolun Cai, Suo Qiu, and Tong Zhang. "Multi-scale convolutional neural networks for crowd counting." In 2017 IEEE International Conference on Image Processing (ICIP), pp. 465-469. IEEE, 2017.

[25] Saqib, M., Khan, S. D., & Blumenstein, M. (2016, November). Texture-based feature mining for crowd density estimation: A study. In 2016 International Conference on Image and Video Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[26] Karpagavalli P, Ramprasad AV. Estimating the density of the people and counting the number of people in a crowd environment for human safety. In2013 International Conference on Communication and Signal Processing 2013 Apr 3 (pp. 663-667). IEEE.

[27] Yoon, Y., Gwak, J., Song, J. I., & Jeon, M. (2016, October). Conditional marked point process-
based crowd counting in sparsely and moderately crowded scenes. In 2016 *International Conference on Control, Automation and Information Sciences (ICCAIS)* (pp. 215-220). IEEE.

[28] Fruin, J. J. (1971). Pedestrian planning and design (No. 206 pp).

[29] Chen, K., Loy, C. C., Gong, S., & Xiang, T. (2012, September). Feature mining for localised crowd counting. *In BMVC* (Vol. 1, No. 2, p. 3).

[30] Chauhan, V., Kumar, S., & Singh, S. K. (2016, December). Human count estimation in high density crowd images and videos. In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 343-347). IEEE.

[31] Zitouni, M. S., Bhaskar, H., Dias, J., & Al-Mualla, M. E. (2016). Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques. *Neurocomputing*, 186, 139-159.

[32] Shen, M., Sun, T., Jiang, X., & Xu, K. (2016, October). Crowd counting estimation in video surveillance based on linear regression function. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 60-65). IEEE.

[33] Fu, M., Xu, P., Li, X., Liu, Q., Ye, M., & Zhu, C. (2015). Fast crowd density estimation with convolutional neural networks. *Engineering Applications of Artificial Intelligence*, 43, 81-88.

[34] Hashemzadeh, M., & Farajzadeh, N. (2016). Combining keypoint-based and segment-based features for counting people in crowded scenes. *Information Sciences*, 345, 199-216.

[35] Al-Zaydi, Z. Q., Ndzi, D. L., Yang, Y., & Kamarudin, M. L. (2016). An adaptive people counting system with dynamic features selection and occlusion handling. *Journal of Visual Communication and Image Representation*, 39, 218-225.

[36] Setti, F., Conigliaro, D., Rota, P., Bassetti, C., Conci, N., Sebe, N., & Cristani, M. (2017). The S-Hock dataset: A new benchmark for spectator crowd analysis. *Computer Vision and Image Understanding*, 159, 47-58.

[37] Liu, W., Lau, R. W., & Manocha, D. (2016). Robust individual and holistic features for crowd scene classification. *Pattern Recognition*, 58, 110-120.

[38] Chaker, R., Al Aghbari, Z., & Junejo, I. N. (2017). Social network model for crowd anomaly detection and localization. *Pattern Recognition*, 61, 266-281.

[39] Wang, C., Zhang, H., Yang, L., Liu, S., & Cao, X. (2015, October). Deep people counting in extremely dense crowds. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1299-1302).

[40] Zhang, Y., Chang, F., Wang, M., Zhang, F., & Han, C. (2018). Auxiliary learning for crowd counting via count-net. *Neurocomputing*, 273, 190-198.

[41] Zhao, W., Zhang, Z., & Huang, K. (2018). Gestalt laws based tracklets analysis for human crowd understanding. *Pattern Recognition*, 75, 112-127.

[42] Mall Dataset. URL (http://www.eecs.qmul.ac.uk/~ccloy/downloads_mall_dataset.html).

[43] http://www.svcl.ucsd.edu/projects/anomaly/UCSD_Anomaly_Dataset.tar.gz

[44] Idrees, H., Saleemi, I., Seibert, C., & Shah, M. (2013). Multi-source multi-scale counting in extremely dense crowd images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2547-2554).

[45] Zhou, B., Wang, X., & Tang, X. (2012, June). Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2871-2878). IEEE.

[46] Zhang, C., Li, H., Wang, X., & Yang, X. (2015). Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 833-841).

[47] Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 589-597).
[48] Xu, T., Duan, Y., Du, J., & Liu, C. (2019, November). Crowd Counting via Conditional Generative Adversarial Networks. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) (pp. 453-462). Springer, Cham.

[49] Gao, J., Wang, Q., & Yuan, Y. (2019). SCAR: Spatial-channel-wise attention regression networks for crowd counting. Neurocomputing, 363, 1-8.

[50] Shi, X., Li, X., Wu, C., Kong, S., Yang, J., & He, L. (2020, May). A Real-Time Deep Network for Crowd Counting. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2328-2332). IEEE.

[51] Jiang, X., Zhang, L., Zhang, T., Lv, P., Zhou, B., Pang, Y., Xu, M. and Xu, C., 2020. Density-Aware Multi-Task Learning for Crowd Counting. IEEE Transactions on Multimedia.

[52] Bouwmans, Thierry, et al. "On the role and the importance of features for background modeling and foreground detection." Computer Science Review 28 (2018): 26-91.