Crowd Counting for Optimal Resource Management

Shubham K. Darak[1], Atharva R. Chavan[2], Sanjivani S. Pande[3] and Renuka L. Brahme[4]

[1,2,3,4]Information Technology Department, Pune Institute of Computer Technology, Pune, Maharashtra, India

Email: [1]shubhamdarak37@gmail.com, [2]atharva.chavan29@gmail.com, [3]pandesanjivani841@gmail.com and [4]renukabrahme@gmail.com

Abstract— The aim of crowd counting using optimal resource management is to estimate the number of people in crowded images or videos from surveillance cameras so that, respective authorities can get effective analysis of crowd flow and can effectively manage resources. Calculating number of people from various images from digital cameras or videos has variety of applications such as traffic monitoring, foot traffic counting from retail stores, safety applications and counting at tremendous crowd locations like in Masjid-e-Haram during Hajj and Umrah congregation and to develop strategy to manage the crowd in most optimal way. In addition to this we can have applications of crowd counting in various day-to-day applications like counting for some survey purposes. So Crowd Counting provides foot traffic at places such as Malls, Retail Stores and Public streets for every moment of time. This count will be used to provide statistical flow of crowd based on day, week, month and year at respective place. The crowd counting has some challenges too like non-uniform density images, background noises and occlusions present in images. Nevertheless, lot of research has been done in recent past and many new methodologies are evolving which are dealing really effectively with stated problems. In this paper, we are doing comparative study of various methodologies which are used for crowd counting and we are providing comprehensive idea about Convolutional Neural Network based approaches such as Multi Scale Convolutional Neural Network.

Keywords— Multi Scale Convolutional Neural Network, Convolutional Neural Network, Deep Learning, Crowd Counting, Image Processing.

1. INTRODUCTION

Computer vision is a field of artificial intelligence that helps computers to recognize and understand visual world that is world of videos and images. Computer Vision includes identifying and classifying images from cameras and then making conclusions based on that or analysing situations based on image and taking consequent actions to handle them effectively.

Crowd Counting aims to count or estimate the number of people in an image. Crowd can be categorised as sparse crowd and dense crowd. Dense Crowd Counting includes calculating no of people highly crowded regions whereas Sparse Crowd Counting includes calculating no of people in comparatively less crowded regions.

Crowd counting has attracted high attention from researchers in recent past due to various reasons some of them could be:

1.1 Retail Industry:

Understanding crowd count is very important in retail which in turn helps to optimise store layout, understand peak times and most importantly to protect against theft. So, to count foot traffic in store we can now put a camera in our store and will connect it to AI platform and will be able to collect data which can be analysed further.

1.2 Infrastructure Planning:

Crowd counting can be used to analyse the flow of crowd in public places like streets, roads, etc. and can make decisions for optimal resource planning.

1.3 Safety:

Another area where Crowd Counting will be of significant importance can be safety. Based on crowd count we can ensure an area is human free before starting a big machine.

1.4 Crowd Counting at Festivals:

Crowd counting can be very useful to count highly crowded scenes in Masjid-e-Haram during Hajj and Umrah congregation.

2. REVIEW OF TRADITIONAL APPROACHES

Various approaches have been demonstrated to overcome the problem of crowd counting particularly in images and videos. Let us classify the approaches into the corresponding following categories:

1) Detection based approaches, 2) Regression based approaches, 3) Density based approaches.

2.1 Detection based Approaches:

Counting by detection can be classified into three types, based on the features we will be using to identify the crowd in images or videos through cameras.

Monolithic Detection: In this technique, it trains the classifier using the full-body appearance which are available in the training images using typical features of the human body such as hair wavelets, few of gradient based features such as a histogram of oriented gradient (HOG), etc. And even learning approaches such as SVM (Support Vector Machines), Random forests have been used that employs a sliding window approach. But these algorithms have a limited scope to sparse crowds. To deal with dense crowds, part-based detection is often more useful and efficent.

Part-based detection: Instead of taking the whole human body as an input to the classifier, this technique considers a part of the body, let's say head or shoulders and applies a classifier to it. But Head solely(alone) isn't sufficient in estimating the presence of a person completely, therefore head + shoulder is the preferred combination used in this technique.

Shape matching: In this technique, ellipses are used to draw boundaries around humans, and then a randomly determined process is used to estimate the number of people and shape with configuration.

2.2 Regression Based Methods:

The drawback of the previous detection approaches was that they couldn't extract the low level features of the particular scenario and were not able to successfully estimate the number of people in extremely dense crowds and with high background clutter. So to overcome this problem regression based method was proposed and in this method we create a outer boundry around an image (or patch), and then for each patch we extract the low level features and then determine the actual count. That is, they learn a mapping between the features extracted from their image patches to their actual counts. The major components of these methods are: Low level feature extraction and regression modelling.

Features such as foreground features, texture, edge and gradient features have been used for encoding the low level information. Other foreground features are extracted from other segments using techniques. Holistic features such are area, perimeter etc have came up with remarkable results. These methods extracted the global features of the scenario, the local features such as edges and LBP etc are been used to improve further results. After extraction of global and local features, various regression techniques such as linear regression piecewise linear regression, ridge regression etc are used.

2.3 Counting by Estimating the Density

The earlier methods were efficient in addressing the issues of dense crowds and background clutter but they ignored the spatial information (It is the digital connection between location, people and activities. The information represents graphical illustration of what is happening where, how and why to show the insight and impact of the past, the present and the likely future) persisting in the images.

However, this approach aims on the density by learning the mapping between local features and the object density maps, thereby including spatial information in the process. Due to which it avoids learning each individual separately one at a time and tracks a group of individuals at a time. The mapping obtained can be linear or

non-liner. For non linear mapping they used random forest regression to vote for densities of multiple target objects to learn a non-linear mapping. Density estimation-based methods have two advantages, One is they can use more spatial information by pixel-wise regression and another one is they can get crowd distribution information of the given images. For the later advantage, pedestrian in any region of an image can be counted by integrating the corresponding density map, and abnormal happens can also be detected.
In more recent approaches, it was observed that the existing crowd density estimation methods were using a smaller set features and limiting their ability to perform well.

3. CNN-BASED METHODS

Putting traditional approaches aside, presently, Convolutional Neural Network (CNN) is a computer vision-based technique which are being used to achieve a better accuracy over the conventional techniques. CNN involves Convolutional layers, pooling layers, Rectified Liner Unit and Fully Connected Layers to extract features that are used to obtain the density map. There is a big bunch of CNNs which designed to attain the crowd density.

- ***Basic CNNs:*** These comprised of the initial deep learning approaches used. These have basic Convolutional layers, kernels (filters), and Pooling layers.
- ***Scale-aware models:*** A more robust CNN wherein multi-column architectures are used with respective filters having different sizes in a single layer.
- ***Context-aware models:*** Both the local and global contextual information is incorporated into CNN. It

is basically used for achieving lower estimation errors

- ***Multi-task frameworks:*** Besides crowd counting, other tasks such as crowd- velocity estimation, foreground- background subtraction is used.
- ***Patch based inference:*** In this, the CNNs are trained using the patches which are cropped from the input images. It should be noted that crop sizes vary for different methods. During the pre-diction phase, a sliding window slides over the testing images and for each small window the predictions were obtained and finally the counts were aggregated to obtain the total count in image.
- ***Whole image-based inference:*** Methods in this category usually performs a whole-image based inference rather than cropping into patches. These methods avoid computationally expensive sliding windows which were used in previous approach.

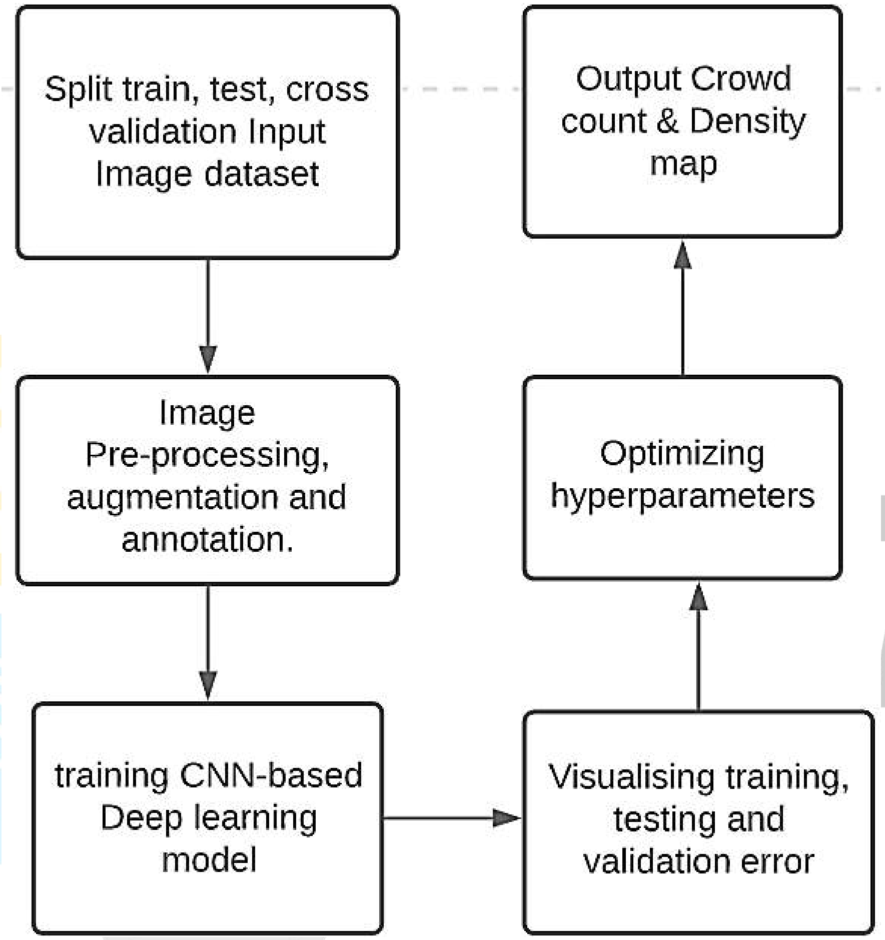
Split train, test, cross validation Input Image dataset
Image Pre-processing, augmentation and annotation.
training CNN-based Deep learning model
Visualising training, testing and validation error
Optimizing hyperparameters
Output Crowd count & Density map

Fig. 1: Flow chart representing process of building Crowd Counting Deep Neural Network model.

3.1 Survey of CNN-based methods

This section provides a survey for convolutional neural network-based crowd counting methods.

Wang et al. [2] proposed a deep CNN based method in order to provide count of dense crowd, the method used AlexNet network [6], where last layer which consisted fully connected layer was replaced by the layer with single neuron, model used is end to end deep CNN regression model. In another approach, Fu et al.

[1] proposed deep CNN method which classifies input image in to various classes such as dense and sparse crowd.

Above methods worked fine for the images from trained data set but performance was significantly dropped for new images. Zhang et al. [3] proposed a method which includes mapping of crowd count with crowd density in

order to provide better count results for new images as compared to previous approaches and this functionality is known as cross-scene counting which was obtained by training the model on both crowd count and density estimation.

Walach and Wolf [4] implemented layered boosting and selective sampling, where layered boosting is a process where every new convolutional neural network layer is trained on the basis of difference between actual output and output from previous layer and selective sampling is a process where images on which model is already trained are removed in order to improve generalization performance of model and images with outliers and less labels are also removed which reduced training time and increased accuracy of model as compared to previous methods.

Shang et al [5] proposed end to end crowd counting model which takes whole image as an input instead of dividing an image in to patches which involved extra computation and complexity for training each individual patches and overlapping regions, the proposed approach uses pretrained GoogleNet model [6] which provided high dimensional features of local blocks, these high dimensional features are decoded by Long-Short time memory (LSTM) decoders to decode local count and maps it to global count.

As above methods failed to provide accurate crowd count in varying crowd density situations, In order to get accurate crowd count despite of crowd density variation Multi Scale Convolutional Neural Network based approach was proposed by Zhang et al. [7] for scenes with varying crowd density and perspective, proposed method involves layer which is comprised of columns having filters with different sizes (small, medium, large) , because of filters with different sizes present at each layer, an accurate crowd count is provided for varying crowd density.

3. FUTURE RESEARCH DIRECTIONS/SCOPE

1. Crowd Counting is essential to serve many real-world applications, such as resource management (such as water, food supply), traffic control, security, disaster management etc.
2. The traditional methods used for crowd counting such as counting people manually, using sensors, maintaining registers is a very time consuming, tedious process which may result in to false count.
3. An accurate crowd counting system provides solutions for emergency situations such as malls, crowded places like kumbh-mela, hudge shopping industry and many other.
4. In these conditions, an estimate of the crowd would allow the concerned authorities to make the correct decisions regarding supplies and planning of resources.
5. Recent advancements in technology have come up with crowd counting solutions using regression, Density estimation and CNN based approaches.

4. CONCLUSION

This article gives the overall overview of the crowd counting having domain “computer vision” and recent advances in the density estimation and also the CNN-based methods for crowd counting. We used the various methods and done the comparative study of the approaches used for crowd counting which also can be further categorized into traditional approaches and CNN-based approaches accordingly. Based on the results produced by the traditional and CNN-based approaches, we concluded that the CNN-based methods are best or efficient for handling the large density crowds with variations in object scales and scene perspective. We have also done the literature survey. CNN-based methods improve the estimation error. Hence paper prvides survey of various approaches used to implement crowd counting and comparative study of CNN based deep neural networks for Crowd counting.

REFERENCES

[1] 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.

[2] Wang, C., Zhang, H., Yang, L., Liu, S., Cao, X., 2015. Deep people counting in extremely dense crowds, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM. pp. 1299–1302.

[3] 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.

[4] Walach, E., Wolf, L., 2016. Learning to count with cnn boosting, in: European Conference on Computer Vision, Springer. pp. 660–676.

[5] Shang, C., Ai, H., Bai, B., 2016. End-to-end crowd counting via joint learning local and global count, in: Image Processing (ICIP), 2016 IEEE International Conference on, IEEE. pp. 1215–1219.

[6] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.