ISSN 2395-1621

Crowd Detection and Counting using Convolutional Neural Network with Expression Detection

Niharika Arkhade, Hemant Shelar, Anup Patil, Rohit Dash

niharikaarkhade4@gmail.com
srshelar555@gmail.com
anuppatil0808@gmail.com
rohitdash08@gmail.com

Computer Department, Savitribai Phule University
AISSMS IOIT, Pune, India

ABSTRACT

In this paper, we advance the state-of-the-art performance for crowd counting in high density scenes by further exploring the idea of a fully convolutional crowd counting model. The traditionally used crowd counting techniques along with low quality detectors resulted in degrading performance. The fully convolutional neural network method is more superior as compared to traditional techniques. Developing a highly generalised counting model, which can detect the crowd and give accurate count in low as well as high density scenes, is the key aim of the model. Moreover, the model is further extended to detect human facial expressions such as happy, sad and angry. The model is also equipped with a feature that, in the case of a rally, it can scan and detect the angry expression in the crowd. And, if massive amount of angry or violent expressions are detected then automatically an alert message is sent to the police to avoid further damage.

Keywords— Crowd counting, crowd detection, convolutional neural networks, expression detection.

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

I. INTRODUCTION

Recently, the applications of crowd counting and crowd detection have gained importance. Estimating the accurate number of people in low as well as high areas is of high usage. Applications for counting and detection of crowd exist in numerous diverse areas such as mall, gardens, markets, shops, urban planning, stadium, metro-stations, managing traffics and many more. Scanning the scenes and computing the count can be of utmost importance in finding the footfalls of a shop and thus in estimating sales. Modern technology aspires to make human life simpler. The traditional methods have generally yielded poor results in highly congested scenes with several hundreds of people in frame. Here, the concept of convolutional neural network comes in picture. Convolutional neural network is an aspect of neural networks used in analysing visual images, image recognition and classification. There are a few factors that can affect the counting of crowds such as occlusion, density difference, perspective distortion, and camera angle.

The model implemented is able to detect and estimate the crowd count in low density and high density regions. It can also detect the human facial expression which is of utmost usage and has many applications. The highlight of the model is the feature that sends an alert message to police in case of increasing anger in the rally, by detecting facial expressions in the gathered crowd.

In deep learning, the crowd counting and detection methods are classified into three categories based on different features. Namely the methods are Monolithic Detection method, Part-based detection method and Shape matching method.

[1] **Monolithic Detection**: In this method, it trains the classifier by using the full-body appearance which is available in the training images using typical features such as gradient-based features such as a histogram of oriented gradient (HOG), etc. Learning approaches such as SVMs, random forests have been used that employ a sliding window approach. But these methods are applied only to sparse crowds. To particularly deal with dense crowds, part-based detection is more often used.

[2] Part-based detection: This method takes into consideration only heads or shoulders. Head merely is not enough in evaluating the presence of a person reliably,

© 2020, IERJ All Rights Reserved

therefore head + shoulder is the preferred combination in this technique.

[3] Shape matching: In this method, Ellipse shape is used to draw boundaries around human figures, and then a stochastic process is used to approximately evaluate the number and shape configuration.

• Regression-based Counting Approach:-

Regression based approach is also called as map-based approach. This method is basically used to overcome the issue of occlusion. Occlusion occurs in image processing when a human figure is hidden behind another human figure or an obstacle. Occurrence of occlusion may affect crowd count. In this method it takes into consideration the mapping between actual points to estimate the accurate count. Though this method is more sophisticated than other methods it deals with certain limitations which include not taking into account the spatial information for crowd counting. Along with this, the regression-based approach could also be time-consuming.

• Counting by estimating the density:-

The constraint in regression-based approach is overcome in this method. The spatial information is taken into consideration in this method of crowd counting by estimating the density. Spatial data is a data which has an overall reference to a particular location or geographical region. Spatial data is referred to as geospatial data or geographic information. The spatial information can be obtained by learning the mapping between object density maps and local features, which are also used in detecting density.

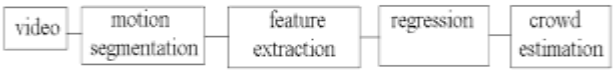
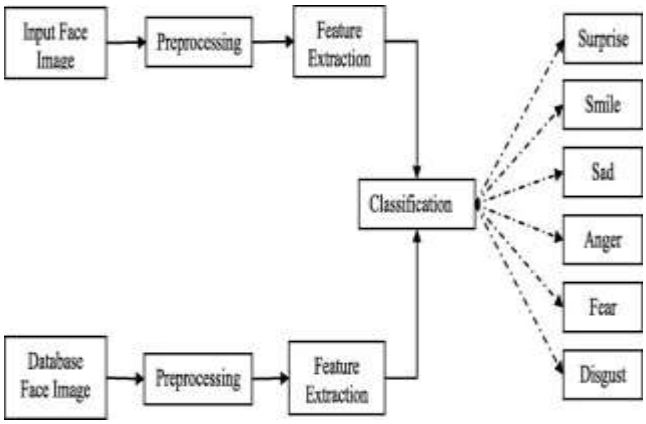


Fig.: Regression-based approach



II. LITERATURE SURVEY

[1] Deepak Babu Sam, R. Venkatesh Babu "Top-Down Feedback for Crowd Counting Convolutional Neural Network"

Typical crowd counting CNNs are trained to look for crowd patterns, instead of accounting every person and count. As a result, in many scenes, crowd like patterns are perceived as people that can lead to false predictions. In this paper, top-down feedback is proposed, which carries high-level scene context to accurate spurious detections. The architecture includes a bottom-up CNN, which has connections to another top-down CNN. The top-down CNN generates feedback in the form of gating to the lower level activations of the bottom-up CNN. This selectively passes legitimate activations and damps spurious responses. It is shown that, such a feedback model achieves better, accurate or competitive results on all basic crowd counting datasets.

[2] Deepak Babu Sam*, Skand Vishwanath Peri*, Mukuntha N. S., Amogh Kamath,R. Venkatesh Babu "Locate, Size and Count: Accurately Resolving People in Dense Crowds via Detection"

In this paper, a new dense detection framework for crowd counting is introduced and the general paradigm of density regression is shown. The proposed LSC-CNN model uses a multi-column architecture with top-down feedback processing to resolve people in dense crowds. In this model, a bounding box is drawn around a human figure. Experiments prove that the model achieves not only greater crowd counting performance than traditional regression methods, but it also has better localization considering all the merits of a detection system. It is also proved and stated that the detection based method is more superior as compared to regression method.

[3] Kang Han, Wanggen Wan, Haiyan Yao, and Li Hou "Image Crowd Counting Using Convolutional Neural Network and Markov Random Field"

In this paper, a CNN-MRF based approach to crowd counting in a still image from various scenes is proposed. The features extracted from the CNN model are trained to show a strong ability to represent crowd density. With the overlapping patches divided strategies, the adjacent local counts are highly correlated. This correlation can be used by the MRF to smooth out the adjacent local counts to obtain a more correct overall count. Experimental results signifies that the proposed approach achieve superior performance compared with several traditional methods.

[4] Preeja Priji, Rashmi S Nair "A Survey on Multiple Face Detection and Tracking in Crowds"

In this paper, it is discussed in brief about image processing. It is an ever-increasing technology with an extensive range of applications and aids. Multiple face detection is a very difficult technique. There are many methods to detect faces. Each one of them has their own advantages and disadvantages. Accuracy decreases with speed and also with image quality. Each method is used for a variety of purposes. It is impossible to decide which method is superior than other. Face detection and expression detection is a basic step in many high-end functions like emotion detection, face recognition, lie detection, etc.

© 2020, IERJ All Rights Reserved

[5] Kaustubh Kulkarni, Ciprian Adrian Corneanu, Ikechukwu Ofodile, Sergio Escalera, Xavier Bar´o, Sylwia Hyniewska, J¨uri Allik, and Gholamreza Anbarjafari, "Automatic Recognition of Facial Displays of Unfelt Emotions"

In this paper, a method inspired from action recognition is proposed and extended to perform facial expression of emotion detection and recognition. It combines the feature maps computed from the EMNet CNN with a facial landmark detector to compute spatio-temporal TPF descriptors. It encodes these descriptors with Fisher vectors to get a single vector representation per video. The feature vector per video issued to train a linear SVM classifier. It outperforms the state of the art performance on the publicly available CK+ and Oulu-CASIA both containing posed FEEs, and show competitive results on the BP4D dataset for facial action unit recognition. In addition, it provides several baselines on our SASE-FE dataset.

[6] Nithya Roopa.S "Emotion Recognition from Facial Expression using Deep Learning"

In this paper, numerous researches and studies about Emotion Recognition, Deep learning techniques used for detecting the emotions are conducted. In this paper, it also stated that it is required in future, to have a model like this which is much more reliable and also which has unlimited possibilities in a variety of fields. In this project, an inception for solving emotion recognition problem is used. A variety of databases have been explored in this research. Kaggle's and Karolinska's Directed Emotional Faces (KDEF) is used as a dataset for carrying out the research. Tensor Flow is used to train the model.

III. DISCUSSION

Concept of expression detection in crowd:-

The research in crowd counting and crowd detection field can be further enhanced to great horizons. The application of expression detection can be of great utilization and advantage all over the world. Scanning the human facial expression using biometric markers can gain immense uses. The facial expression recognition software is a technology which uses sentiment analysis tool and can basically detect the six main expressions:- happiness, sadness, anger, surprise, fear, and disgust. Many algorithms are of great use in expression detection. Some of them are HoG, Haar cascade classifier, and CNNs. Local Binary Patterns Histograms (LBPH) algorithm is one of the easiest algorithms whereas CNN is considered to be the most accurate one.

Following listed and explained are some of the innovative research experiments which can be incorporated with crowd counting and detection can yield high successful results:-

[1] The human facial expression detection mechanism incorporated with crowd counting can be used in detecting happy, sad or angry faces in crowded scenes.

[2] Crime rates can be reduced by detecting suspicious expressions. Moreover, further measures could be taken to avoid heinous crimes.

[3] The crowd detection incorporated with face detection can also be of utmost usage in case a person goes missing in mall, market, stations or any other crowded public places.

[4] Number of angry and violent people in a rally or protest can be detected by using crowd counting and expression detection. In case of large number of people with angry or violent expression, an alert message could be immediately sent to police and government who can take preventive actions and can steer clear further damage. The alert message could also be sent to other citizens for the purpose of security and safety.

[5] The crowd detection incorporated with face detection can also be of utmost usage in case a person goes missing in mall, market, stations or any other crowded public places.

[6] The government can use this mechanism of facial expression detection for computing the societal wellbeing metrics. It can be aided in calculating how happy and content the citizens are.

IV. METHODOLOGY

In our approach, for crowd count and expression detection, we have used following features and mechanisms:-

- The dataset used is either a live dataset or a Kaggle dataset.

[1] Box classification:-

In this, LSC-CNN (Location, Size and Count) is used to locate individuals in a crowd. The MRF module locates people and puts bounding boxes on the detected heads. While in some cases bounding boxes are drawn around head + shoulder, in this case we consider only the heads.

[2] Count heads:-

In this, the mechanism for counting of bounding boxes is initiated and a total count is computed after applying Non-Maximum Suppression (NMS) in cases of massive overlapping. This standard NMS technique is majorly important in areas of high density.

[3] Feature Extractor:-

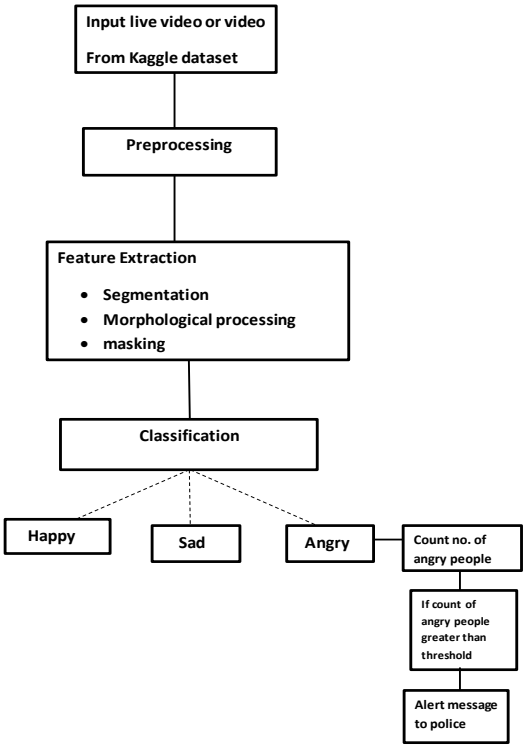
At this phase image segmentation, morphological processing and masking mechanisms are used for emotion detection and recognition.

[4] Classification of emotions:-

Once the feature extraction stage is completed, the detected emotions or expressions are classified into happy, sad or angry expressions.

ARCHITECTURE:-

© 2020, IERJ All Rights Reserved



V. CONCLUSION

In this paper, Convolutional Neural Network is the vital framework to learn prominent features for crowd detection and crowd counting. It is an end-to-end training method which performs a whole image based inference. Enhanced innovative research topics for facial expression recognition incorporated with crowd counting and crowd detection are discussed in brief. The model implemented is able to detect the crowd and give accurate crowd count. The model, equipped with facial expression detection mechanism, can detect happy, sad and angry expressions. The key highlight of the implemented model is an automatic alert message feature which is of utmost importance and usage and has great scope.

Future study can be involved around gender detection in the crowd. The study can be further enhanced to 3D image processing and feature extraction and detection.

VI. ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on A Study on Crowd Detection Fully Convolutional Crowd Counting on Low-Resolution Crowded Scenes".With due respect and gratitude we would like to take this opportunity to thank our internal guide Prof. A. G. Said for giving us all the help and guidance we needed.

REFRENCES

1. He D.; Hujic D.; Mills J.K.; Benhabib B. "Moving object recognition using premarking and active vision"; Robotics and Automation, 1996. Proceedings. 1996 IEEE International Conference.

2. Diplaros, A.; Gevers, T.; Patras, I.;"Combining color and shape informationfor illumination-viewpoint invariant object recognition", Image Processing, IEEE Transactions on Jan. 2006 Volume: 15,Issue:1

3. Jae-Han Park; Seung-HoBaeg; JaehanKoh; Kyung-Wook Park; Moon-Hong Baeg; "A new object recognition system for service robots in the smart environment" Control, Automation and Systems,2007. ICCAS '07.

4. Chyi-Yeu Lin; Setiawan, E. "Object orientation recognition based on SIFT and SVM by using stereo camera", Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference.

5. Chensheng Wang; Fei Wang; "A Knowledge- Based Strategy forObject Recognition and Reconstruction "; Information Technology and Computer Science,2009.ITCS 2009.

6. Jen-Shiun Chiang; Chih-Hsien Hsia; Shih-Hung Chang; "An efficient object recognition and self-localization system for humanoid soccerrobot", SICE Annual Conference 2010.

7. YinghuaXue; Guohui Tian; Rongkuan Li; Haitao Jiang; " A new object search and recognition method based on artificial object markin complex indoor environment ", Control Conference, 2008. CCC2008.

8. Seung-HoBaeg; Jae-Han Park; JaehanKoh; Kyung-Wook Park; Moonhong-Baeg; "An object recognition system for a smart home environment on the basis of colour and texture descriptors", Intelligent Robots and Systems, 2007.IROS 2007.

© 2020, IERJ All Rights Reserved