International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Crowd Detection and Counting from Images using MResnet

Sai NitishaVemuri, Srinivas Kudipudi

Abstract*: Crowd Detection and counting is important for crowd control and monitoring in places like pilgrimages. Automatic crowd detection from images have several challenges. Different scale variations and viewpoints of images make it difficult for crowd detection models to generalize for broader data. Most of the existing approaches for crowd detection contains multiple columns for extracting multi-scale features. By using multiple columns through a deeper network can cause the layers to lose features as the layers get deeper. In this paper, a new Multi-Residual Network (MResnet) is proposed for crowd detection and counting. MResnet contains multiple three columns sub-networks with three receptive field variations. The advantages of the proposed network is that each sub-network has a specific receptive field for imbalanced distribution of human crowd in the image. Residual connections are utilized in each subnetwork for information propagation. The MResnet is evaluated using the ShanghaiTech dataset. Extensive experiments have shown that our proposed network achieves lower count error and high spatial localization.

Index Terms*: Crowd detection, Crowd counting, deep learning, ShanghiTech dataset.

I. INTRODUCTION

Overcrowding at places such as pilgrimages and tourist attractions can lead to stampedes. Automatic crowd detection and crowd counting helps maintain security and also analyze crowd patterns. Most approaches for automatic crowd counting suffer from the images. The crowd images have several viewpoint variations, imbalanced distribution of crowds and different illuminations. These make it difficult to approximate the crowd numbers and localize the humans the image. One part of the image has clear crowd representation while other parts has farther and minimal representations.

Most recent research in crowd counting is focused on using deep convolution neural networks. CNN Regression based approaches have become the state of the art approaches, where the count is directly calculated from the images. These networks have several drawbacks. Single column networks cannot extract the different scales of crowd from images. They have better approximation with nearer crowd than the farther crowds. To overcome these Multi-column approaches or Multi-path approaches are used. These approaches contain multiple CNN networks for different crowd densities. The multi-columns approaches take a large number of layers and parameters. As the layers are increases the information that is propagating through these layers vanishes. Although multiple columns extract blob features with different scales, they tend to be biased towards a single scale.

In this paper, they are proposing a novel network named MResnet to overcome these challenges. The proposed network contains three column sub networks for different scales in the images. At the end of each sub network all the columns are fused together and given as input to the next subnetwork. Each column in the subnetwork contains residual connections so that information is propagated throughout the network along with the multi scale feature maps. The contributions from this paper include 1) leveraging residual connection so that the vanishing gradient problem is minimized. 2) Combining all the multiscale features from each the subnetwork throughout the network to approximate all types of crowd densities equally. One of the popular dataset for crowd counting ShanghaiTech dataset was used to evaluate and experiment the proposed approach. The results obtained have shown that our proposed network has lower counting errors and can accurately predict on crowd images different scales and variations.

II. RELATED WORK

Zhang et al. [1] proposed a multicolumn convolutional neural network where the image has several layers in which the heads are of different sizes. Filters of different sizes are used to capture the crowd density at different scales. Density maps are derived by using different sized filters. A new data set which contains 1198 images were present. Ravanbakhsh [2] proposed a schematic representation approach. In this framework they used two channel representation, appearance and motion capture was used in first channel and in the second channel two cross channel tasks which include generation of optical flow images from the original image frame and in the second task an appearance information was generated starting from an optical flow. They used a fully convolutional layer which consists of convolutional layers, batch normalization and ReLu nonlinearities.

Maria et al. [3] proposed aa CNN model contains six convolutional layers. The output from the convolutional layer is fed to the softmax layers which produces the difference between the two classes of crowd and noncrowd. Except the last convolutional layers each convolutional layer in the architecture was attached to a parametric rectified linear unit. Two types of regularizations namely DA and MEB were applied on each convolutional layer from the architecture.

Manuscript published on 30 June 2019.
* Correspondence Author (s)
Sai Nitisha Vemuri*, Department of CSE, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.
Srinivas Kudipudi, Department of CSE, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.

Zhou et al. [4] proposed a Robust mobile crowd sensing architecture, which consists of 4 planes which includes sensor plane, distributed plane, medium plane and application plane.
[5] proposed a Cascaded deep learning network for detecting the anomaly in the crowd. It consists of CNN which is made up of nine layers the initial layers are used for extracting the features where the last two fully connected layers were used for predicting output probabilities and coordinates. Detection and tracking was done for detecting and keeping track of the features. Anomaly detection is used for focusing on the individual spatial and temporal features. The heat map was proposed which contains local features embedded in it. Mark [6] proposed Resnet crowd architecture for crowd counting. In this model initially 7*7 conv layer was taken then two batch normalization blocks were used and then four 3*3 conv which contain batch normalization blocks in between them was used. The whole module was divided into two 3 modules. First module consists of input image conv layer and a batch normalization layer whose output is given as an input the second block and also the third block. Second block consists of two batch normalizations and two conv blocks whose output is given to the third block as well as as the output layer.

III. METHODOLOGY

The proposed methodology includes acquiring the ShanghaiTech Dataset. Pre-Processing the dataset to make it compatible for the proposed network. Evaluating and comparing the results of the proposed and other state of the art models.

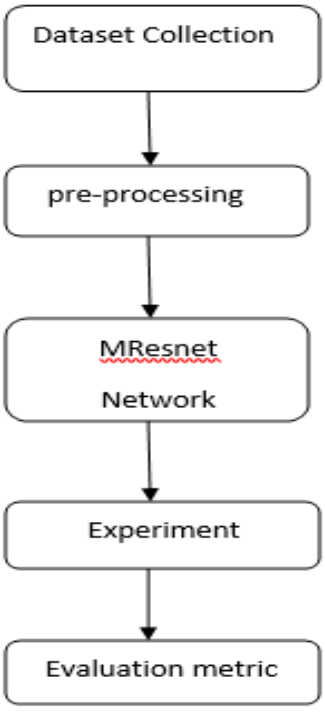

Fig 1. Proposed methodology

A. Dataset Collection

One of the popular benchmark dataset for crowd counting is ShanghaiTech dataset. It contains around 1198 annotated images. Each image has head annotations with a total of 330,165 annotations across the entire dataset. The dataset is divided into two parts, part A contains 482 images taken from the internet and part B contains 716 images taken from shanghai city. Training and testing set are pre-separated. Part-A gave an MAE of 72.5 and part -B gave MAE of 19.1.

B. Pre-processing

The model was trained on both parts of the dataset. As the model can take arbitrary image sizes, randomly cropped the images are taken and used Gaussian kernel to generate the corresponding density maps. The cropped images along with the original images are taken to train the model.

C. MResnet Architecture

The proposed architecture contains three multi column sub networks as shown in Figure 2. Each column follows a particular receptive field throughout the network. After the initial convolution the feature maps are given to three column networks. First column network follows a receptive field of 3 x 3 size. Other three follows 5 x 5 and 7 x 7 respectively. All the networks contain Residual connections. Residual connections minimize the vanishing gradient problem and increase the efficiency of the network. All the feature maps from the first sub network are concatenated to give equal importance to the receptive fields which corresponds to the equivalent distributions in the image.
The following two sub networks follow the similar structure that of the first subnetwork. The second subnetwork has increasing feature filters which are decreases by the third sub network. The features from the final subnetwork are concatenated and a convolution operation of 3 x 3 is performed along with Rectified Linear Unit (ReLU) activation. To produce the final crowd count 1 x1 convolution operation is performed. Each convolution block in the network contains a convolution block followed by Batch Normalization and then ReLU activation.

IV. EXPERIMENTS

Tensor flow framework is used to implement our model and adam[7] optimizer with a learning rate of 0.001 was used as the optimizer. 20% of random data from the training data is used for validation testing. The network converges in 60 epochs.

A. Evaluation Metrics

Mean Absolute Error (MAE) and Mean Squared Error (MSE) are generally used evaluation metrics for application such as crowd counting. By following the convention, this research also uses these metrics to evaluate our model. They are defined as follows:

$$MAE = \frac{1}{N}\sum_{1}^{N}|z_i - \hat{z_i}| \quad (1)$$

$$MSE = \sqrt{\frac{1}{N}\sum_{1}^{N}|z_i - \hat{z_i}|} \quad (2)$$

Where N is the total number of images, z is the actual value and zˆis the predicted value. The evaluation metrics of the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

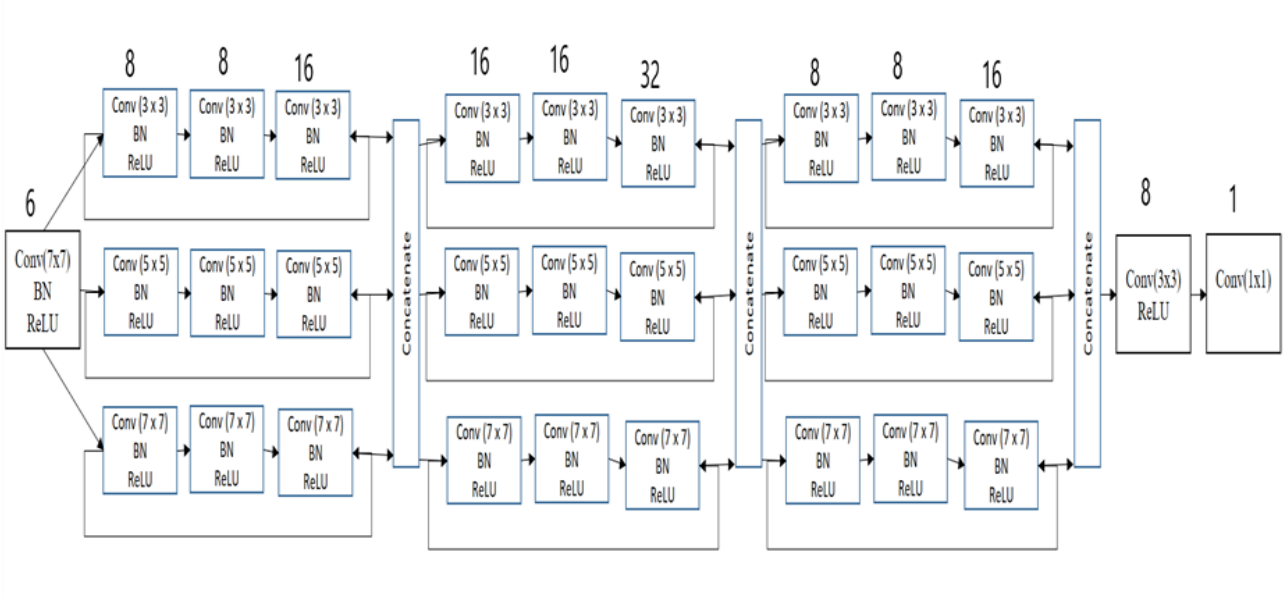

Fig 2. MResnet Architecture

MRsenet model and the comparison models are given in Table 1. The MAE and MSE shows the overall error in the predicted test set. MResnet has produced better results than all the state of the art comparison models with MAE of 72.5 and MSE of 98.7 in part A and MAE of 19.1 and MSE of 31.7 on part B.

V. RESULTS

The Heatmap predictions from the test set are given in Figure 3, 4, 5. The results have shown that our proposed model can produce human count near to the ground truth. The three receptive field columns in the network have facilitated for predicting human crowd that is farther in the image as well as crowd that is nearer. This shows the potential of our network at predicting different kinds of crowd images at different scales.

Method	MAE	MSE
Zhang	182.5	280.1
MCNN	112.7	186.9
Switch-CNN	92.4	140.8
Our Proposed	**72.5**	**98.7**

Table 1: Comparison of our proposed and state of the art methods on part A.

Method	MAE	MSE
Zhang	33.6	51.8
MCNN	26.2	42.3
Switch-CNN	22.5	33.5
Our Proposed	**19.1**	**31.7**

Table 2: Comparison of our proposed and state of the art methods on part B.

VI. CONCLUSION

The new network named MResnet for crowd counting and crowd detection from overcrowded images is proposed. Each column in the subnetwork of our model has different receptive fields that can extract features at different scales. The model can minimize the vanishing gradient problem by leveraging the residual connections. Images with different scales and distributions can be used to predict the approximate crowd count. The experiments performed on the ShanghaiTech has shown the effectiveness of our model at predicting overcrowded images with MAE of 72.5 and MSE of 98.7. This research can further improve the automatic crowd counting approaches to achieve better approximate results.

REFERENCES

1. Y. Zhang, D. Zhou, S. Chen, S. Gao and Y. Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 589-597.
2. Ravanbakhsh, Mahdyar&Sangineto, Enver&Nabi, Moin&Sebe, Nicu. (2019). Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds. 1896-1904. 10.1109/WACV.2019.00206.
3. M. Tzelepi and A. Tefas, "Graph Embedded Convolutional Neural Networks in Human Crowd Detection for Drone Flight Safety," in IEEE Transactions on Emerging Topics in Computational Intelligence.
4. Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz and J. Rodriguez, "Robust Mobile Crowd Sensing: When Deep Learning Meets Edge Computing," in IEEE Network, vol. 32, no. 4, pp. 54-60, July/August 2018.
5. Qiu, Peng & Kim, Sumi & Lee, Jeong-Hyu& Choi, Jaeho. (2018). Anomaly Detection in a Crowd Using a Cascade of Deep Learning Networks. 10.1007/978-981-10-7512-4_59.
6. M. Marsden, K. McGuinness, S. Little and N. E. O'Connor, "ResnetCrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification," 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, 2017, pp. 1-7.
7. Kingma, Diederik P. and Jimmy Ba. "Adam: A Method for Stochastic Optimization." CoRR abs/1412.6980 (2015).

Fig3: (a) Original Image with Ground Truth of 1111 people (b) Heatmap image with predicted value is 1101 people

Fig 4: (a) Original Image with Ground Truth of 169 people (b) Heatmap image with predicted value is 155 people

Fig 5: (a) Original Image with Ground Truth of 825 people (b) Heatmap image with predicted value is 817 people

AUTHORS PROFILE

Sai NitishaVemuri is currently pursuing Master of Technology in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. Her research interests include deep learning, cloud computing and Data structure.

Dr. K. Srinivas is currently working as professor in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College. His research interests include bioinformatics, data mining, data structures, database design, deep learning and big data analytics.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)