Crowd Counting Using Density Map

Reeda Saeed (CS1946)
Faculty of Computer Science and Engineering
Ghulam Ishaq Khan Institute of Institute of Engineering Sciences and Technology, Pakistan
E-mail: reeda.saeed@gmail.com

Abstract

This paper performs a comparison of different models used for crowd counting. In MCNN, multi-column parallel convolutional neural network structure is used that generates population density maps by adapting crowd changes caused by camera view-points and resolution using filters with different size receptive fields. In CNN with transfer learning structure of a multi-column convolutional neural network, is abandoned using the first ten layers of VGG-16 as the front part and the convolutional neural network as the latter part.

1. Introduction

Crowd Counting is a technique used to estimate the number of people in an image or a video. Estimating crowds from images or videos has became an increasingly important application of Computer Vision. Crowd counting use cases include: 1) Counting crowds in forbidden areas in a manufacturing unit to enforce safety rules and minimizing health risks. 2) Managing high traffic roads and public spaces. 3) Counting attendance in educational institutions. 4) Urban Planning. 5) Video Surveillance.

Earlier Methods for Crowd Counting. 1) Detection Style Framework scans a detector over two consecutive frames of video sequence. Its limitation is occlusion. 2) Feature-based Regression is the most popular method for crowd counting. It is done by segmenting the foreground, extracting the various features and then applying regression function to estimate crowd count. Its limitation is that it is very difficult to get accurate estimate. estimate. 3) CNN-based method requires perspective map on both training and testing scenes. Perspective maps are not readily available and that is its only limitation.

2. Multi-Column Convolutional Neural Network

The main objective of this neural network is accurate crowd counting from an arbitrary still images with arbitrary camera perspective and crowd density. Challenges while training the network are 1) Foreground segmentation is very difficult task to perform when the geometry of the scene is unknown. It became impossible to segment the crowd from its background. To cope this challenge estimation of crowd should be done without segmenting the foreground. 2) Very great occlusion for most people in images causes great variance in density and distribution. 3) Variation of scale causes difficulty in feature tracking so by using the methods that can learn features automatically will benefit the network.

Solution to all the above-mentioned challenges is multi-column convolutional neural network. It has following structure.

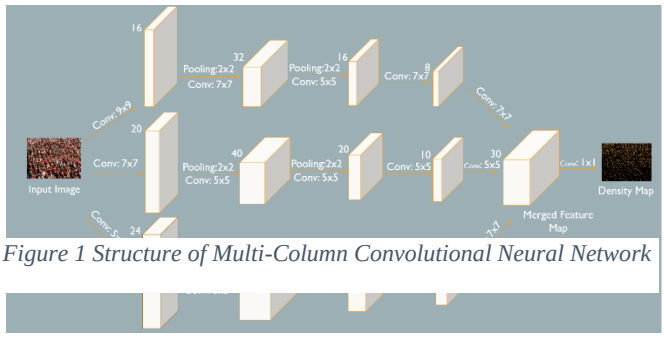

Figure 1 Structure of Multi-Column Convolutional Neural Network

3. Experiments

1)MCNN (results generated by me using author pretrained weights) 2) MCNN (trained from scratch by me) 3) CNN with transfer learning (on pretrained VGG16 weights).

4. Evaluation Metric

MAE: Mean Absolute Error (determines accuracy of estimates)
MSE: Mean Squared Error (indicates the robustness of estimates)

$$\text{MAE} = \frac{1}{N}\sum_{1}^{N}|z_i - \hat{z}_i|$$

$$\text{MSE} = \sqrt{\frac{1}{N}\sum_{1}^{N}(z_i - \hat{z}_i)^2}$$

- N: Number of test images
- z_i: actual number of people in ith image

  $\hat{z}_i$: estimated number of people in ith image

5. Results

Method	MAE	MSE
MCNN (results generated by author)	110.2	173.2
MCNN (results generated by me using author pretrained weights)	110.19	169.37
MCNN (trained from scratch by me)	189.63	246.96
CNN with transfer learning (on pretrained VGG16 weights)	75.7	112.33

Table 1 Experiments Evaluation

6. Comparison

MCNN (results generated by me using author pretrained weights)

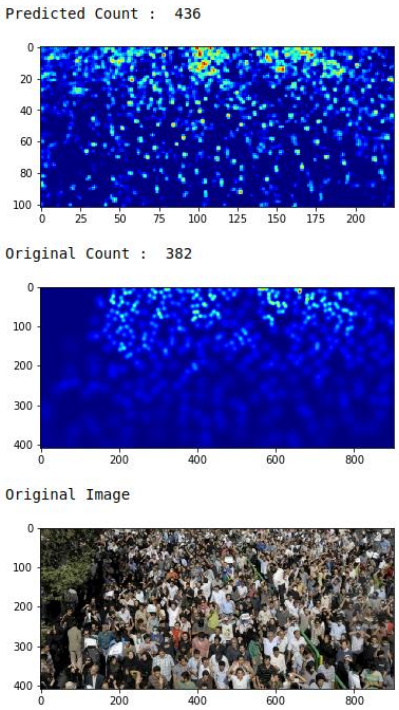

Figure 2MCNN (results generated by me using author pretrained weights)

MCNN (trained from scratch by me)

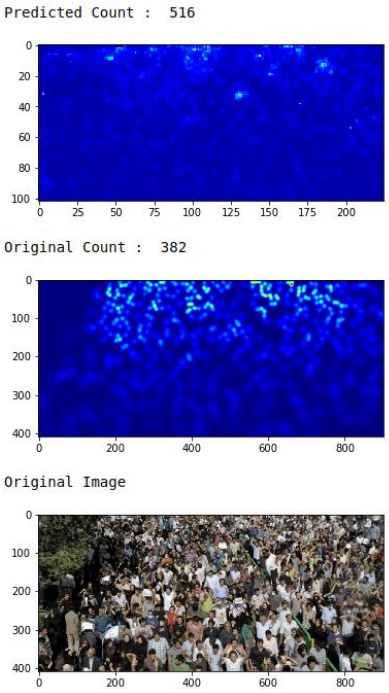

Figure 3 MCNN (trained from scratch by me)

CNN with transfer learning (on pretrained VGG16 weights)

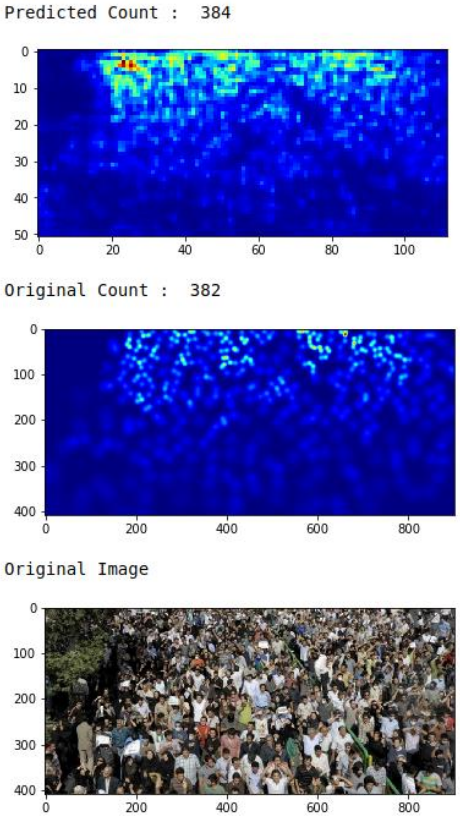

Figure 4CNN with transfer learning (on pretrained VGG16 weights)

7. CONCLUSION

By comparing all the results, it is concluded that CNN with transfer learning performed better because it is trained on VGG16 model. It has 14 million images, so it gives better results. MCNN using pretrained data set gives 110.9 MAE, MCNN trained from scratch gives 189.63 MAE. On the other hand, CNN with transfer learning gives 75.5 MAE outperforming all the other models.

6. References

[1] Zhang Y, Zhou D, Chen S, Gao S, Ma Y. Single-image crowd counting via multi-column convolutional neural network. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 589-597).

[2] https://nanonets.com/blog/crowd-counting-review/#counting-by-estimating-the-density

[3]https://boominathanlokesh.wordpress.com/collaborative-learning-application-in-crowd-counting/

[4] Wang Z, Deng Q, Zhao Y. The Comparison of Crowd Counting Algorithms based on Computer Vision. InJournal of Physics: Conference Series 2019 Apr (Vol. 1187, No. 4, p. 042012). IOP Publishing.