International Journal of Scientific & Engineering Research Volume 11, Issue 12, December-2020
ISSN 2229-5518

Face Counting and Mask Detection using Deep Learning

Janani Prabu [a] ,Sai Saranesh B [b], Dr.S.Ajitha [c]

[a] *ITM Business School, Student, Chennai, (janani2706@gmail.com)*
[b] *ITM Business School, Student, Chennai, (bsaisaranesh@gmail.com)*
[c] *ITM Business School, Assistant Professor, Chennai, (ajitha1985@gmail.com)*

Keywords:
Face Recognition
Biometrics
Deep Learning
Machine learning
Prediction
TensorFlow

A b s t r a c t

Face is one among the foremost important human's biometrics which is used frequently in every day human communication and due to some of its unique characteristics plays a major role in conveying identity and emotion. So far numerous methods have been proposed for face recognition, but it's still remained very challenging in real world applications and up to date; there is no technique which equals human ability to recognize faces despite many variations in appearance that the face can have in a scene and provides a strong solution to all situations. The aim of this paper is to present the problems & issues that occur in a face recognition system in detecting & recognizing faces accurately due to light, aging, expressions, similarity in faces and other systematic problems like noise, image acquisition, video camera distortion etc. In this article using TensorFlow this problem can be rectified. The coronavirus COVID-19 pandemic is causing a global health crisis. One of the effective protection methods is wearing a face mask in public areas according to the World Health Organization (WHO). In this paper, a hybrid model using deep and classical machine learning for mask detection are going to be presented. The models (available at (https://github.com/JananiPrabu/)) can be used to assist this article.

.

1.Introduction

Face recognition has recently become a very active research area, partly due to the increased interest in biometric security systems in general, but also because of recent advances that have taken the state-of-the-art far beyond the initial attempts of using direct image comparison. the primary of face detection system has been developed since in early 1970. Due to the limitation of computation, system can't be satisfied the requirement of users, which is identify passport photograph real time. At the beginning of 1990's techniques are proposed focused on the face recognition on and increase the need of face detection. Face recognition has attracted much attention and its research has rapidly expanded by not only engineers but also neuroscientists, since it has many potential applications in computer vision communication and automatic access control system. it's a specific and hard case of object recognition. It has always been a very challenging problem within the field of image processing and computer vision, and intrinsically has received an excellent deal of attention over the last few years. On the one hand, its applications are very useful in various domains. On the other hand, it has always been very difficult to implement because of all different situation that a person's face is often found. Especially, face detection is a vital part of face recognition as the first step of automatic face recognition. However, face detection is not straightforward because it has lots of variations of image appearance, such as pose variation (front, non-front), occlusion, image orientation, illuminating condition and facial expression. The aim of face detection is detecting faces in any images or videos. The face is our primary focus of attention during a face recognition system similarly as in social intercourse of our day-to-day life and plays a major role in conveying our identity and emotion. Although, there have been developed so many Face Recognition systems based on numerous methods to recognize faces but not equals human ability to recognize faces despite many variations in appearance that a face can have in a scene. The human ability to recognize faces is remarkable. We can recognize thousands of faces learned throughout our life time and identify familiar faces at a look even after years of separation despite large changes in the visual stimulus due to viewing conditions, expression, aging and distractions such as glasses or changes in hairstyles or facial hair.

The trend of wearing face masks in public is rising due to the COVID-19

coronavirus epidemic all over the world. Before Covid-19, People used to wear masks to protect their health from air pollution. While other people are self-conscious about their looks, they hide their emotions from the public by hiding their faces. Scientists proofed that wearing face masks works on impeding COVID-19 transmission. COVID-19 (known as coronavirus) is that the latest epidemic virus that hit the human health within the last century. In 2020, the rapid spreading of COVID-19 has forced the planet Health Organization to declare COVID-19 as a worldwide pandemic. More than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas.

The coronavirus epidemic has given rise to an extraordinary degree of worldwide scientific cooperation. Artificial Intelligence (AI) based on Machine learning and Deep Learning can help to fight Covid-19 in many ways. Machine learning allows researchers and clinicians evaluate vast quantities of knowledge to forecast the distribution of COVID-19, to function an early warning mechanism for potential pandemics, and to classify vulnerable populations. The provision of healthcare needs funding for emerging technology like AI, IoT, big data and machine learning to tackle and predict new diseases. In order to better understand infection rates and to trace and quickly detect infections, the AI 's power is being exploited to address the Covid-19 pandemic such as the detection of COVID-19 in medical chest X-rays.

Policymakers are facing a lot of challenges and risks in facing the spreading and transmission of COVID-19. People are forced by laws to wear face masks publicly in many countries. These rules and laws were developed as an action to the exponential growth in cases and deaths in many areas. However, the process of monitoring large groups of people is becoming more difficult. The monitoring process involves the detection of anyone who isn't wearing a mask. In France, to guarantee that riders wear face masks, new AI software tools are integrated in the Paris Metro system's surveillance cameras. The French startup DatakaLab, which developed the software, reports that the goal is not to recognize or arrest people who do not wear masks but to produce anonymous statistical data that can help the authorities predict potential outbreaks of COVID-19.

2. Related Work

Face counting and Analysis have a plethora of real-world applications such as planning emergency evacuations in case of fire outbreaks, calamitous events, etc. and making informed decisions on the basis of the number of people such as water, food planning, detecting congestion etc. and hence, there are many methods proposed to achieve Face count [3].

I. Face counting based on object detection mechanism Earlier approaches for Face counting have used Detection Based methods. The detection methods are candid and it make use of off-the-shelf detectors [7][8][9] to detect the target objects and count these objects in images or videos.

1) Monolithic detection [10][11][12][13] which is called as the typical pedestrian approach where training is performed by extracting the human anatomy features. Some of the common features in this method are obtained using gradient-based features [14], edge let [15], and shape lets [16]. The accuracy of human detection significantly depends on the classifier used for classification. This approach gives satisfactory detection in sparse scenes, but it does not work well in Faced scenes in presence of occlusion and clutter scene.

2) Part-Based Detection: By adopting a part-based detection method [17][18][19], there are several solutions to handle the partial occlusion problem till some extent; such as, head features are not sufficient to provide reliable results due to its shape variations, one can construct ensemble boosted classifier for specific body part considering its own set of features [20].

3) Multi-sensor detection: To resolve inter object occlusion which occurs due to the partial or full overlapping of more than one objects, the multi-view information from multiple cameras can be incorporated in the process. For instance, in [21], authors have estimated the Face count by applying the multi-view geometric constraints to its full extent. Through the use of this methodology, the speed of detection is increased significantly apart from detection accuracy improvement. These supervised methods of Face counting detect faces very well, but it fails in highly congested environments and even in surveillance applications where the resolution of the images affects its accuracy.

B. Regression based Face Counting Some of the images are captured with low resolution, it is the major performance issue of detection-based Face counting and the occluded multiple objects. Regression based counting performs better in this environment, where local features get extracted from the segmented images and then the regression model gets applied to estimates the Face count in each segment [22][23][24]. Prior to this, regression-based methods were developed [25][26][27][28][29][31] using the global image features, but these approaches 772 Ujwala Bhangale et al. / Procedia Computer Science 171 (2020) 770–779 U. Bhangale et al. / Procedia Computer Science 00 (2019) 000–000 3 cannot capture the region wise distribution of the information. One of the crucial parts of this type of methods is extracting suitable features. This approach may overestimate the prediction when the Face is less.

C. Face Counting by Density The density-based methods generate density values which are estimated using low-level features such as pixels or regions, it overcomes the drawback of regression-based methods and also maintains the location information [11]. The predicted density maps may have different characteristics as the density map estimation methods may vary depending on the selection of the loss function and type of prediction. The prediction and loss function can either be region-wise or pixel-wise. Since image-wise prediction reuse computations, they are relatively faster.

3. Methods and data

3.1. Data description

Dataset includes the Name, two files. One is training data and the other is testing data. The training data includes Name, width,height, xmin, ymin, xmax, ymax, class. It contains co-ordinates of each head in the train image, located by the formation of a bound-box around the head. The testing data includes Name. It contains only name of testing images. The prediction file will have Name and the headcount. It contains format for a valid submission) AI based algorithms learn from the historical data to provide predictions for the future outcomes. Machine learning (ML) and deep learning (DL) algorithms can be considered as a subset of the AI. It is an area that is based on learning and improving on its own by analyzing computer algorithms. There are certain differences between machine learning and deep learning. To assess the predictive performance of each of the developed predictive models, we calculated their performance in terms of accuracy, f1-score, precision, recall, and area under roc curve (AUC). To validate the data, we both used 10-fold cross validation and 80– 20 train-test split approach

4.Deep analysis into algorithms

4.1 TensorFlow

TensorFlow is a machine learning / deep learning / multilayer neural network library from Google. Libraries using data flow graphs can be used to describe complex networks in an easy-to-understand manner. With high versatility, it can be used from the research level to real products.

This has not only greatly improved the voice recognition and image search service performance such as “Google search application” and “Google Photos”, but it also has become the core technology that supports most of Google’s services such as search engine and Gmail. However, because it was deeply dependent on the company’s infrastructure, it could not be difficult to publish outside. Therefore, “TensorFlow” was developed to solve the problem and increase versatility. One of the biggest features of TensorFlow is the ability to build a neural network. By using this neural network, machines can perform logical thinking and learning similar to humans. There are other tensors for all processing such as “data loading, preprocessing, calculation, state, output”. Originally considered not only as deep learning, but also as a library for performing tensor calculations, it is the most excellent library when considered as a deep learning framework that can also describe basic calculation processing. In addition, there is “Define and Run” which builds calculation processing once as calculation graph and then collects calculation processing. TensorFlow basically describes all calculation processes by calculation graph, no matter how simple the calculation is. Therefore, some familiarity is required to handle it. Besides, TensorFlow can perform distributed learning to work in any environment such as iOS and Android. Distributed processing allows TensorFlow to handle large amounts of data such as big data. Of course, there are plenty of additional libraries for doing calculations at high speed. From the large number of users such as the Wrapper layer of TensorFlow, various functions will be extended in the future.

4.2 Challenges of TensorFlow

First of all, because of its very high performance, hardware that uses TensorFlow is also required to have a matching high performance. In addition, although it can be said that distributed learning of data calculation can be said to be an advantage, it can be said that preparing an environment for doing that is a bit more difficult. The operating systems that support TensorFlow are 64-bit “Windows” “MacOS” “Ubuntu / Linux”. In addition, there are two types of “TensorFlow for Mobile” and “TensorFlow Lite”, which are also provided for mobile devices, and can be used with mobile devices such as Android, iOS, and Raspberry Pi. It is important to note here that when using TensorFlow in full scale, be sure to prepare a computer with high GPU performance that can perform a large amount of computation at high speed by parallel processing. The higher the performance of the GPU, the faster the learning speed of TensorFlow can be. Also, the programming languages that can be used with TensorFlow are “Python”, “C language” and “C ++”.

4.3 Image Recognition

TensorFlow can analyze the information in the image. This image recognition is a technology that has great potential for this. For example, Google has developed a function that allows a machine to recognize an image of one animal and make it into sentences. Is this useful for the blind in the future? And is greatly expected. It is also used as an “automatic brake system” that automatically brakes when it is likely to be hit by image recognition if it is more familiar. Image recognition is also used for speed control and steering wheel operation. By learning the fake image of the brand back in the net shop, etc, a system that detects and tells when similar images are found has become possible. In addition, by recognizing the face based on the image from the surveillance camera and comparing the face image with the registered database, identification is performed without using a person to obtain information on who and when they came to the store can do. In addition, a story that often talks about this image recognition accuracy is about "sorting cucumbers". I tried this to see if the former engineer could use Deep Learning to assess the grade of the cucumber. As a result, we were able to achieve a grade of about 80 percent of cucumbers that would otherwise require a veteran eyesight test. This groundbreaking idea was featured on Google's official blog and exhibited at many IT fairs.

4.4 Image Search

Speaking of Google, there should be a small number of people answering image search, and TensorFlow is also active in this image search that many people casually use. By learning the original image, appropriate images can be searched against the features learned so far. Currently, this identification performance is said to be evolving at a speed that is more than doubled in a year. It is possible to say that Google has evolved to performance almost the same as human beings in the near future, just because Google with world-class identification performance.

4.5 Speech Recognition

Familiar with one phrase "OK, Google", speech recognition is one of the features that are now commonplace. Even now, speech recognition rates are high, and most modern devices can also be operated with speech, which is Google's top-class performance. Google's research team is thinking in the near future that it will not be possible to recognize speech using chips alone without using neural networks. What's more, it's amazing because I'm thinking of doing it with a disposable tip.

4.6 TensorFlow Object Detection API

In 2017, Google announced the TensorFlow Object Detection API, and has since released a number of new features and models. Neural architecture search, instance segmentation support, models trained on new datasets such as Open Images, etc. TensorFlow Object Detection API is used in various ways, from illegal parking detection on bus lanes in New York City to disease diagnosis in cassava in Tanzania.

4.7 TensorFlow 2.0

TensorFlow 2.0 alpha has been released. It is the first major version update for TensorFlow. Initial Release ~ 2.0 shows the progress to the official release, and introduces the outline of the new features of 2.0. The developer conference "TensorFlow Dev Summit 2019" was held on March 6–7, 2019. Release of the Alpha (Alpha) version of TensorFlow 2.0 was announced there, and the new functions and modifications of TensorFlow 2.0 were explained. Today, as part of Google's commitment to democratize computer vision, we have incorporated some requests from the research community and made some additions to the API to make this program more useful.

4.8 MobileNetV2

MobileNets are a family of neural network architectures released by Google to be used on machines with limited computing power, like mobile devices. They strive to provide state of the art accuracy, while requiring as little memory and computing power as possible. This makes them a very fast family of networks to use for image processing. MobileNets achieve this performance by reducing dramatically the number of learnable parameters, which also makes them faster and easier to train compared to more traditional networks.

5.Application Result

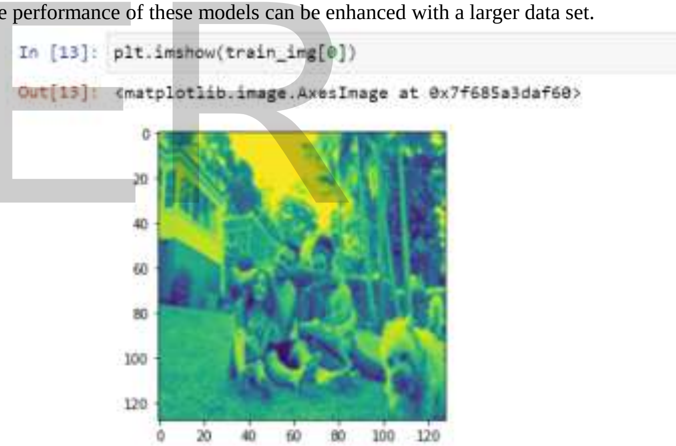
The below table shows the accuracy and other criteria when the model is developed in TensorFlow.

	Accuracy	Precision
TensorFlow	0.998	0.987

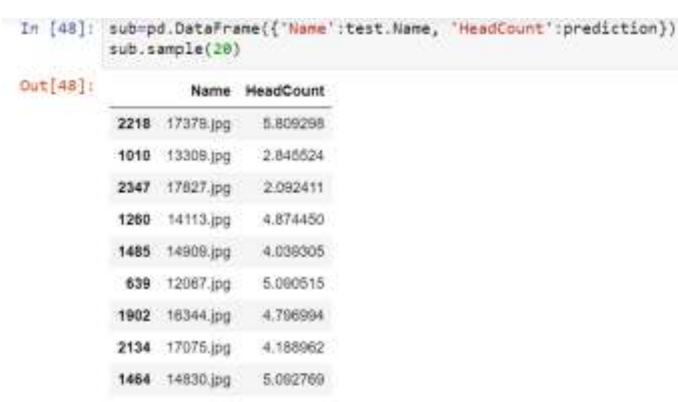
In addition to these, we tested the performance of the algorithms using 80–20 train-test split approach.

(5.1 Training the data)

The major limitation in this study is the size of the data. In addition to these, the data was imbalanced, thus we balanced the data by deleting some materials. The performance of these models can be enhanced with a larger data set.



(5.2 Implementing the TensorFlow Algorithm))



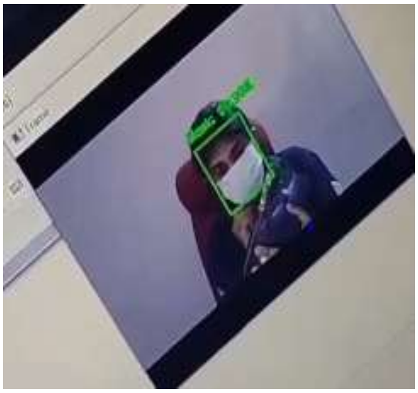
```
In [48]: sub=pd.DataFrame({'Name':test.Name, 'HeadCount':prediction})
         sub.sample(20)
```

Out[48]:

	Name	HeadCount
2218	17378.jpg	5.809298
1010	13308.jpg	2.846524
2347	17827.jpg	2.092411
1260	14113.jpg	4.874450
1485	14908.jpg	4.039305
639	12067.jpg	5.090515
1902	16344.jpg	4.796994
2134	17075.jpg	4.188962
1464	14830.jpg	5.092769
1036	13386.jpg	3.036891

(5.3 Prediction using TensorFlow)

"Mask Detection" is implemented by using Keras, TensorFlow, MobileNet and OpenCV. The mask detector didn't use any morphed masked images dataset. The model is accurate since the MobileNetV2 architecture is used. It is also computationally efficient and thus making it easier to deploy the model to the embedded systems.



(5.4 Mask Detection using TensorFlow)

6.Conclusion

In this work, we propose a contribution that improve the effectiveness of TensorFlow. The efficacy of this technique was evaluated in the domain of visual object counting. In this paper, we explored the basic premise of the TensorFlow. However, there are several improvements that we would like to explore in the future. These include an adaptive parameterization for the sample selection parameters: high and low thresholds. This can be based, for example, on the relative contribution of each sample to the weight updates. The model is accurate since the MobileNetV2 architecture is used. When comparing to the other machine learning algorithms, the accuracy of the Tensor is much higher and appropriate. This project can be integrated with embedded systems for application in railway station, airports, offices, colleges, schools, public places to ensure that the public safety guidelines are followed.

Conflict of interest

On behalf of all authors, the corresponding author (Dr.S.Ajitha) states that there is no conflict of interest.

References

[1] Zhang, Yingying, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. (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.

[2] Simonyan, Karen, and Andrew Zisserman. (2014) "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556.

[3] Boominathan, Lokesh, Srinivas SS Kruthiventi, and R. Venkatesh Babu.(2016), "Crowdnet: A deep convolutional network for dense crowd counting." In Proceedings of the 24th ACM international conference on Multimedia, pp. 640-644. ACM,

[4] Peng, Lei, Lei Wu, Yalan Ye, Fengqi Yu, and Hai Yuan. (2007), "CPN modeling and analysis of HMIPv6." In 2007 IEEE International Conference on Integration Technology, pp. 68-73. IEEE, 2007.

[5] Image datasets, https://cdn-images-1.medium.com,[online] [accessed on December 2018]

[6] Li, Min, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection." In 2008 19th International Conference on Pattern Recognition, pp. 1-4. IEEE, 2008.

[7] Leibe, Bastian, Edgar Seemann, and Bernt Schiele. "Pedestrian detection in crowded scenes." In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 878-885. IEEE, 2005.

[8] Wang, Lu, and Nelson HC Yung. "Crowd counting and segmentation in visual surveillance." In 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 2573-2576. IEEE, 2009.

[9] Gavrila, Dariu M., Jan Giebel, and Stefan Munder. "Vision-based pedestrian detection: The protector system." In Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy. 2004.

[10] Mahadevan, Vijay, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. "Anomaly detection in crowded scenes." In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1975-1981. IEEE, 2010.

[11] Tuzel, Oncel, Fatih Porikli, and Peter Meer. "Pedestrian detection via classification on riemannian manifolds." IEEE transactions on pattern analysis and machine intelligence 30, no. 10 (2008): 1713-1727.

[12] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.

[13] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In international Conference on computer vision & Pattern Recognition (CVPR'05), vol. 1, pp. 886-893. IEEE Computer Society, 2005.

[14] Wu, Bo, and Ramakant Nevatia. "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors." In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, pp. 90-97. IEEE, 2005.

[15] Sabzmeydani, Payam, and Greg Mori. "Detecting pedestrians by learning shapelet features." In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8. IEEE, 2007.

[16] Felzenszwalb, Pedro F., Ross B. Girshick, David McAllester, and Deva Ramanan. "Object detection with discriminatively trained part-based models." IEEE transactions on pattern analysis and machine intelligence 32, no. 9 (2010): 1627-1645.

[17] Lin, Sheng-Fuu, Jaw-Yeh Chen, and Hung-Xin Chao. "Estimation of number of people in crowded scenes using perspective transformation." IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 31, no. 6 (2001): 645-654.

[18] Wu, Bo, and Ram Nevatia. "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors." International Journal of Computer Vision 75, no. 2 (2007): 247-266.

[19] Li, Min, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection." In 2008 19th International Conference on Pattern Recognition, pp. 1-4. IEEE, 2008.

[20] Ge, Weina, and Robert T. Collins. "Crowd detection with a multiview sampler." In European Conference on Computer Vision, pp. 324-337. Springer, Berlin, Heidelberg, 2010.

[21] Chan, Antoni B., Zhang-Sheng John Liang, and Nuno Vasconcelos. "Privacy preserving crowd monitoring: Counting people without people models or tracking." In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-7. IEEE, 2008.

[22] Ryan, David, Simon Denman, Clinton Fookes, and Sridha Sridharan. "Crowd counting using multiple local features." In 2009 Digital Image Computing: Techniques and Applications, pp. 81-88. IEEE, 2009.

[23] Chan, Antoni B., and Nuno Vasconcelos. "Counting people with low-level features and Bayesian regression." IEEE Transactions on Image Processing 21, no. 4 (2012): 2160-2177.

[24] Idrees, Haroon, Imran Saleemi, Cody Seibert, and Mubarak Shah. "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. 2013.

[25] Kong, Dan, Douglas Gray, and Hai Tao. "A viewpoint invariant approach for crowd counting." In 18th International Conference on Pattern Recognition (ICPR'06), vol. 3, pp. 1187-1190. IEEE, 2006.

[26] Cho, Siu-Yeung, Tommy WS Chow, and Chi-Tat Leung. "A neural-based crowd estimation by hybrid global learning algorithm." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29, no. 4 (1999): 535-541.

[27] Chen, Ke, Shaogang Gong, Tao Xiang, and Chen Change Loy.(2013) "Cumulative attribute space for age and crowd density estimation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2467-2474.

[28] Chen, Ke, Chen Change Loy, Shaogang Gong, and Tony Xiang. (2012) "Feature mining for localised crowd counting." In BMVC, vol. 1, no. 2, p. 3.

[29] Li, Yuhong, Xiaofan Zhang, and Deming Chen (2018).. "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

[30] Kang, D., Ma, Z., & Chan, A. B. (2018). Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks—Counting, Detection, and Tracking. IEEE Transactions on Circuits and Systems for Video Technology, 29(5), 1408-1422.

[31] Deepak Babu Sam, Shiv Surya, and R. Venkatesh Babu. Switching convolutional neural network for crowd counting. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4031–4039, 2017. 10 U. Bhangale et al./ Procedia Computer Science 00 (2019) 000–000

[32] Vishwanath A Sindagi and Vishal M Patel. Generating high-quality crowd density maps using contextual pyramid cnns. In IEEE International Conference on Computer Vision, pages 1879–1888, 2017.

[33] Xinkun Cao, Zhipeng Wang, Yanyun Zhao, and Fei Su. Scale aggregation network for accurate and efficient crowd counting. In Proceedings of the European Conference on Computer Vision (ECCV), pages 734–750, 2018

[34] Di Kang and Antoni B. Chan. Crowd counting by adaptively fusing predictions from an image pyramid. In British Ma-chine Vision Conference, page 89, 2018.

[35] Lingbo Liu, Hongjun Wang, Guanbin Li, Wanli Ouyang, and Liang Lin. Crowd counting using deep recurrent spatial-aware network. In Proceedings of the International Joint Conference on Artificial Intelligence, pages 849–855, 2018.