

A Novel and Dynamic Framework for People Counting

Prabhu.D, Roshni Musaddi, Anny Jaiswal, Nitheesh Varma



Abstract: A real time person counting from video records is a primary structure coalition for some applications in dense urban areas. Practically speaking, this undertaking typically experiences numerous issues, such as the absence of real time processing of the recorded videos or the event of errors because of insignificant individuals being counted. To defeat the above issues, we propose a novel and dynamic framework for real time people counting named YOLO-PC (People Counting based on YOLO).

Keywords: YOLO-PC, People Counting, CNN.

I. INTRODUCTION

Estimating the number of persons in public places provides useful information for Video based surveillance and monitoring applications. In this work we alter the existing object detection system YOLO by proposing the so-called YOLO-PC (People Counting based on YOLO). YOLO-PC extends the original YOLO system using a deep learning approach to achieve a higher accuracy in people counting. Compared to other existing object detection systems such as R-CNN, Fast R-CNN, and Faster R-CNN, YOLO has been chosen as the base method in our YOLO-PC, because of its low computation overhead and its ability to detect objects in real-time. In terms of Performance in real time, YOLO-PC retrains a deep convolutional neural network to detect people at more than 40 fps (frames per second) with the support of a GPU. Regarding the people counting process, the image is divided into a 9*9 grids and boundaries are noted. This leads to more detected grids and a greater average confidence value. Moreover the YOLO-PC chooses a different boundary area according to the actual application scenario to count people in a contrapuntal manner which further improves the counting accuracy. YOLO-PC can also ignore the insignificant individuals who may be in billboards or other irrelevant areas. Experimental results show that YOLO-PC can count people quickly with a higher accuracy at the entrance or exits of places like escalators, stadiums, malls etc.

Revised Manuscript Received on April 18, 2020.

* Correspondence Author

Prabhu.D*, SRM Institute of Science and Technology, Chennai
Department of Computer Science and Engineering prabhud1@srmist.edu.in

Roshni Musaddi, SRM Institute of Science and Technology, Chennai
Department of Computer Science and Engineering
roshnimusaddi1997@gmail.com

Anny Jaiswal, SRM Institute of Science and Technology, Chennai
Department of Computer Science and Engineering
annyjaiswal11@gmail.com

Nitheesh Varma, SRM Institute of Science and Technology, Chennai
Department of Computer Science and Engineering
nitheesh.w.7@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

I. RELATED WORK

There are for the most part three sorts of strategies to count the number of individuals in a video. The first type utilizes a factual strategy to evaluate the quantity of individuals in an area. This strategy typically relates the pixel or on the other hand different highlights of the moving region and the number of individuals in the region, and afterward prepares a capacity to gauge the quantity of individuals. Lee et al (2007) and Kim et al. (2008) utilized the hidden highlights such as closer view pixel measurements and movement vectors for individuals tallying. The quantity of pixels was related with the quantity of individuals and the movement vector was used to recognize the headings[1]. In these papers, the basic component was utilized to compute the part work and the effect was found to be satisfying. In any case, the best approach to ascertaining the portion work was profoundly pertinent to the specific scene. The second kind of strategy comprises consolidating object discovery with object following. This sort for the most part has a preprocessing system. During this technique, the moving zone is separated and the individuals are recognized in the moving zone. Normal object identification techniques incorporated the segmentation technique which was based on projection histogram proposed by Zhang et al. (2007) and Ma et al. (2008)[2], the technique in view of format coordinating was proposed by Hsieh et al. (2007) and the technique is dependent on factual classification which was proposed by Li et al. (2008). Bat calculation is a novel populace based transformative and bio-propelled stochastic streamlining calculation and Cuckoo search is a late created meta-heuristic (2018). These calculations have likewise been applied in object following because of their high calculation overhead and long preparation time, these techniques are very difficult to convey progressive frameworks. The third sort of strategies utilizes the feature point tracking to obtain information regarding the path followed and then performs cluster analysis. Rabaud et al. (2006) proposed straightforward methods for spatially and also transiently molding the trajectories. Given this portrayal, they incorporated it with a learned object descriptor to accomplish a division of the constituent movements. In the field of people counting, Convolutional Neural Networks(CNN) can naturally gain proficiency with the significant level semantics of highlighting the individuals through an arrangement of convolution activities and continually correcting parameters by back engineering. One of the agent works was proposed by Zhang et al. (2015),

which utilized CNN to extract the population density feature in pictures to gauge the individual's number in different situations. It performed end-to-end training and recognition without the complex steps such as foreground segmentation, feature extraction and artificial design. Article discovery dependent on deep learning primarily incorporates two kinds of techniques. The first type comprises a two-organize discovery calculation that partitions the issue into two phases: (i) producing district propositions, and (ii) characterizing the applicant locales. The agent of such calculations was R-CNN (2014), Fast R-CNN (2015), and Faster R-CNN (2015). Rather Than these methodologies, the second kind of strategies is one-organize identification calculation which utilizes the CNN system to ascertain just a single time[6]. Redmon et al. (2016) proposed YOLO, which is the spearheading work of this kind of strategies. Along these lines, SSD (2016), Mask R CNN (2017) and RetinaNet (2017) were proposed. YOLO changed the article identification issue into a relapse issue, which straightforwardly restored the position of the limit confine and the class likelihood the yield layer. Since the entire procedure of YOLO employments a similar system, it is anything but difficult to execute the end-to-end identification. Yolo has been the subject of a few upgrades, YOLO9000 (2017) and YOLO V3 (2018) were proposed to look for incredibly quick and exact item discovery calculations. With its high exactness and high speed, YOLO is progressively appropriate for real time applications than other algorithms.

II. PROPOSED SYSTEM

YOLO-PC separates a picture into 9×9 cells, every one of which was tried independently. More cells lead to more recognition boxes and higher confidence. YOLO-PC utilizes the cell to define the limit, picks different limits as indicated by the real scene for individuals checking and overlooks insignificant and uninterested regions, making the individuals discovery increasingly relevant and improving the counting exactness. The model of YOLO is exceptionally complicated, and the structure of YOLO-PC is progressively sectioned, coming about in the expansion of the quantity of parameters and model extra room. In this way, the expense of model transmission and capacity on the Internet is colossal.

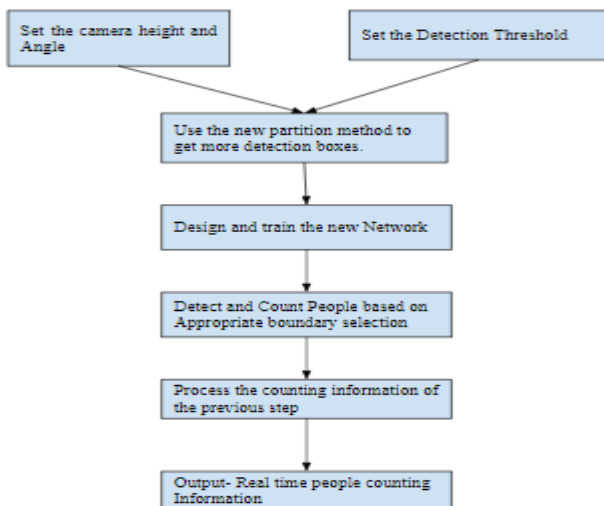


Fig 1. Process of People Counting

As shown in figure 1, YOLO PC includes six steps. The initial step comprises setting the limit of identification and altering the point and height of the camera. The limit is as a rule somewhere in the range of 0.2 and 0.4. On the off chance that the certainty estimation of identified individuals is beneath the limit, the individuals won't be counted. Right now, the default limit estimation of 0.2 is utilized for simplicity.

In the second step the new partition method is known to produce more bounding boxes of individuals. The calculation is intended to be increasingly proficient in distinguishing individuals with high exactness[13]. YOLO partitioned the picture into a 7×7 lattice and two bouncing boxes were anticipated for every lattice cell. Each bouncing box included the facilitated data of the box and the certainty of the individuals. By playing out a lot of examinations, we discovered that individuals counting works better with additional identified boxes and higher certainty esteems. Therefore, YOLO-PC partitions the picture into a 9×9 framework rather than 7×7 and recognizes three bounding boxes.

In the third step, new system structure joined with the fire layer is structured and prepared. The second step is to make more segments, which results in an expansion in the quantity of model parameters. The new organized structure is made out of 21 convolutional layers also, 3 fire layers, and the new model of individuals location is prepared on the informational indexes of VOC-2007 and VOC-2012.

The fourth step intends to identify and check individuals in light of a fitting limit choice. YOLO PC chooses at least one lattice cell from $243(9 \times 9 \times 3)$ cells and picks a limit as per the genuine circumstance. In the event that individuals turn left by some place, the limit of the left region of the video is chosen. The area of the limit is about the area of the 113th cell. Individuals who go through the limit will be checked. So also, if individuals turn directly by some place, the area of the limit is about the area of the 129th cell.

On the off chance that individuals go straight by some place, the area of the limit is about the area of the 121th cell. YOLO-PC can be progressively relevant and exact in recognizing the progression of individuals since some immaterial impedance can be stayed away from. For instance, if individuals are in the boards and random backs, they can be overlooked since they are not in the chosen limit.

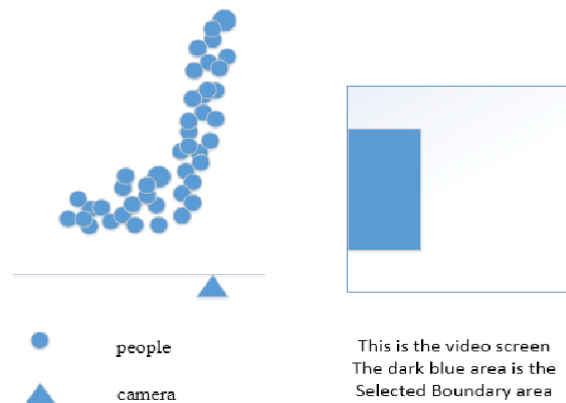


Fig 2. Boundary Selection Sketch Map

The data of the fourth step will be handled in the fifth step. In the chosen limit zone, the boxes number gathers and continually refreshes as time goes on, we characterize the quantity of boxes as S . The estimation of S right now t speaks to the absolute number of recognized individuals from the earliest starting point until time t . Since it requires some investment for individuals to move in the limit zone, a similar individual might be identified over and over again. Concurring to the measurement result, each individual has been identified around multiple times when going through the chosen limit region, so the anticipated number is $S/18$ at the default limit. In the last advance, the checking data is yielded. YOLO-PC shows the present tallying number, FPS what's more, certainty esteem in the video. YOLO-PC can additionally spare constant data and keep on refreshing the data, and afterward yield them through a few interfaces.

III. EXPERIMENTAL RESULTS

In this section the number of model parameters and the average precision of YOLO and YOLO PC are compared. Contrasted with the structure referenced in the paper Redmon et al. (2016), the past form referenced is called YOLO-OLD. YOLO-PC is proposed in light of the YOLO-OLD.

The precision rate primarily relies upon the exactness of the individual's recognition and the precision of counting. Individual recognition is a sort of article identification issue also, the reference purpose of estimating precision in object recognition is the mean normal precision (mAP).

Improvement of more partitions brings about an expansion in parameters and extra space for the model. YOLO PC makes improvement as per this issue. The technique for figuring parameters number and the level of diminished extra space is utilized to look at the impact of model pressure. The quantity of parameters is decreased by 17202912 on the first premise, and the model size is decreased by 36 % compared to YOLO-OLD and 11.5% compared to YOLO.

IV. CONCLUSION AND FUTURE SCOPE

Right now, we have presented a packed real time people counting approach named YOLO-PC. YOLO-PC improves the first convolutional structure of YOLO, and utilises the streamlined fire layer to supplant

the 3 x 3 convolutional layer and the packed model with less parameters is acquired through preparation by more divisions of cells, YOLO-PC accomplishes additional bounding boxes and higher recognition certainty. Joined with a boundary selection technique, people counting turns out to be increasingly appropriate with higher location, as well as counting precision. In the rundown, YOLO-PC decreases the quantity of required parameters and packs the models while accomplishing high precision and constant execution, making it progressively reasonable for application situations. As a future work, we plan to attempt further research on planning low power object detection algorithms with high precision and high performance in real time.

REFERENCES

1. K. Venkatesh and A Bansal, "People counting in large density of crowds from still images" arXiv preprint arXiv: 1517.08465, 2015.

2. Y. Bengio, Y. LeCun, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
3. L. Bottou, Y. Bengio, Y. LeCun and P Haffner "Gradient based study applied for recognition of documents" IEEE vol 85 no 11 pp 2378–2124, 1998.
4. H Jacobs "To count a crowd," Columbia Journalism Review vol 6, no 1, p. 37, 1967.
5. A. Osokin, T.-H. Vu, and I. Laptev, "Context-aware cnns for person head detection," in ICCV, 2015, pp. 2893–2901.
6. R. T. Collins and W. Ge and R. T. Collins, "Marked point processes for crowd counting," in CVPR. IEEE, 2009, pp. 2913–2920.
7. K. Huang, M. Li, Z. Zhang, and T. Tan, "Estimating the number of people in crowded scenes by shoulder and head detection" in Pattern Recognition International Conference held on IEEE 2007
8. S. Belongie and V. Rabaud and "Counting moving objects in heavy crowds" in CVPR vol 1 IEEE 2006 pp 705–711.
9. R. Cipolla, and G. J. Brostow, "Unsupervised bayesian detection of independent motion in crowds," in CVPR vol. 2. IEEE, 2006, pp.595–602.
10. J. H. Yin, A. C. Davies, and S. A. Velastin, "Monitoring of Crowds using image processing" Electronics and Communication Engineering Journal vol 7 no 1 pp 39–49 1994.
11. Z.-S. J. Liang, A. B. Chan, and N. Vasconcelos, "Crowd monitoring and Privacy Preserving Counting people without people models or tracking" in CVPR IEEE 2007 pp 1–9.
12. T. W. Chow, S.-Y. Cho, and C.-T. Leung "A neural based crowd estimation by using hybrid global learning algorithm" IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics) vol 24, no. 4 pp. 535–541, 1999.
13. D. Gray, D. Kong, and H. Tao, "Counting pedestrians in crowds using viewpoint invariant training." in BMVC, 2005, pp. 1–6.
14. S. Denman, C. Fookes, D. Ryan and S. Sridharan, "Crowd counting by using various local features." in Digital Image Computing: Techniques and Applications, 2008. DICTA'09. IEEE, 2008, pp. 82–89.
15. A. Zisserman and V. Lempitsky "Learning to count objects in images," in NIPS, 2010, pp. 1324–1332.

AUTHOR'S PROFILE



Prabhu.D is an Assistant Professor of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He Completed his M.E in Computer and Communication Engineering at AnnaUniversity. His areas of Expertise include Data Structures, Design and Analysis of Algorithms, Operating Systems, Computer Networks etc. He is an Author of various International and National Paper Publications.



Roshni Musaddi is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. She has an expertise in the areas of M.L & A.I. Her research interests include Cryptography, Deep Learning, AI and its applications in the field of Computer Science.



Anny Jaiswal is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. She has expertise in the areas of Neural Networks, Machine Learning, Motion and Path Planning. She is also an author of various paper publications in reputed National Journals. Her research interests include Data Mining, Privacy Protection etc.

A Novel and Dynamic Framework for People Counting



Nitheesh Varma is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He has an expertise in areas of M.L, Neural Networks and Computer Vision. He is also an author of Various Paper Publications in the areas of Machine Learning.