B.JOURNAL PAPER

***Deep learning based dense object counting with density map***

**ABSTRACT:** Assessing the quantity of people in a public spot gives valuable data to video-based reconnaissance and checking applications. On account of sideways camera arrangement, tallying is either accomplished by distinguishing people or by measurably setting up relations between estimations of basic picture highlights to the quantity of individuals. Ongoing individuals tallying from video records is a fundamental structure block for some applications in shrewd urban areas. In the current System, a head indicator can be utilized to appraise the spatially fluctuating head size, which is the key component utilized in our mind tallying methodology. We influence the best in class convolutional neural organization for the meager head identification in thick group. After sub-partitioning the picture into rectangular patches, we first utilize a SURF highlight based SVM double classifier to name each fix as group/not-swarm and kill all not-swarm patches. In the current framework, task as a rule experiences numerous issues, similar to the absence of continuous handling of the recorded recordings or the event of mistakes because of superfluous individuals being tallied. The proposed framework defeats the above issues with a novel continuous individuals tallying approach named YOLO-PC (YOLO based People Counting). In the proposed framework, after the extraordinary pre-treatment, versatile division and highlight extraction for the human checking information, the element vector is utilized as the contributions of the prepared YOLO to characterize and measurements of the absolute number individuals.

***Keywords: People Counting , YOLO-PC, CNN***

##### 1 INTRODUCTION

Dependably assessing the quantity of individuals present in a public spot (for example road, shopping center, tram station) over the long run might be of basic significance for both wellbeing and monetary reasons. For example, a specific number of people in a given setting may mirror an abnormal and possibly risky circumstance. Then again, observing the quantity of individuals in a territory of a shopping center will give important data not exclusively to streamlining the working hours of the shops yet additionally for assessing the appeal of the shopping region. With the approach of savvy cameras and the expanding opportunities for robotized reconnaissance and checking, the mechanization of the human tallying task additionally turns out to be mechanically doable. In this work we modify the current article location framework YOLO by proposing the purported YOLO-PC(People Counting dependent on YOLO). YOLO-PC broadens the first YOLO framework utilizing a profound learning way to deal with accomplish a higher exactness in individuals checking. Contrasted with other existing article location frameworks, for example, R-CNN, Fast R- CNN, and Faster R-CNN, YOLO has been picked as the base technique in our YOLO-PC, as a result of its low calculation overhead and its capacity to identify objects continuously. Regarding Performance continuously, YOLO-PC retrains a profound convolutional neural organization to distinguish individuals at in excess of 40 fps (outlines each second) with the help of a GPU. As to individuals tallying measure, the picture is partitioned into a 9\*9 networks and limits are noted. This prompts more recognized frameworks and a more noteworthy normal certainty esteem. Besides the YOLO-PC picks an alternate limit territory as indicated by the genuine application situation to include individuals in a contrapuntal way which further improves the tallying exactness. YOLO-PC can likewise overlook the immaterial people who might be in announcements or other unimportant territories Experimental outcomes show that YOLO-PC can tally individuals rapidly with a higher exactness at the passageway or ways out of spots like lifts, arenas, shopping centers and so on

The initial step comprises in setting the identification edge and changing the camera. The location results underneath the limit, which is generally set from 0.2 to 0.4, won't be tallied. For straightforwardness, we utilize the default estimation of 0.2 in this work. In the real scene, the camera ought to be acclimated to the proper tallness and point. In the subsequent advance, we distinguish individuals through re-preparing a convolutional neural organization. YOLO partitions the picture into a 7\*7 framework and for every network cell predicts two jumping boxes just as the certainty esteem for those crates. We accept that this division isn't adequate and we point that our calculation will be more effective in distinguishing individuals to accomplish higher tallying exactness. At the end of the day, YOLO-PC works better with more recognized boxes and higher certainty esteems. To this end, YOLO-PC utilizes 9\*9 lattice and 3 bouncing boxes. We set up three arrangements of examinations of 4 minutes video each utilizing various edges (i.e., 0.2, 0.3 and 0.4). The acquired outcomes are promising since YOLO- PC distinguishes more boxes and accomplishes higher certainty esteems for those containers contrasted with YOLO. All the more explicitly, YOLO-PC distinguishes in excess of 10% of the containers, the certainty normal worth is in excess of 50% higher when the limit is 0.2. The third step comprises in recognizing and checking individuals dependent on a suitable limit determination. YOLO-PC chooses at least one lattice cells as the territory limit from 243(9\*9\*3) cells and picks an alternate limit as per the real circumstance. In the event that individuals turn left by some place, the limit of the left region of the video ought to be chosen, the estimation of the limit is around 113 and individuals through that limit will be tallied. Essentially, if individuals turn directly by some place, the estimation of the limit is around 129.

In the event that individuals go straight by some place, the estimation of the limit is around 121. YOLO-PC can be more precise in recognizing the progression of individuals as some garbage obstruction, like individuals in the boards and disconnected backs, can be overlooked as a result of the limit choice. The checking data of the past advance is currently handled in the fourth step. In the chose limit zone, the cases number collects and continually refreshes, we allude to this number as S. The estimation of S at a second t in a video addresses the quantity of identified individuals at that point, which is exact. The estimation of S in a timeframe addresses the complete number of identified individuals. Since it requires some investment for individuals to move in the limit region, the estimation of S is rehashed, in other words, a similar individual has been distinguished commonly. As per the examinations, each individual has been distinguished around multiple times when going through the chose limit zone, so the anticipated number is S/18 at the default edge. In the fifth and last advance, we yield the constant individuals tallying data. YOLO-PC can straightforwardly show the real-time individuals including data in the video pictures, including the current number, FPS, certainty esteem and so forth YOLO-PC can likewise save continuous data and proceed to refresh, and afterward yield them through certain interfaces.

##### II RELATED WORK

The problem of counting people has been handled by many examination endeavors, utilizing a few heterogeneous methodologies. In any case, an investigation in important writing uncovers that each approach is on a basic level persuaded by the uncommon states of the current issue. Numerous strategies utilize face/individual discovery calculations, while others join mass/object location and following plans. The issue is normally alluded to as "individuals tallying" or "swarm checking" when it is applied to huge, open zones. Regular

applications center chiefly around reconnaissance or traffic. The intrigued peruser may discover an investigation zeroing in on checking. To assess person on foot stream checking, ***Hsieh et al. [1]*** introduced a framework that utilizes Kinect. Their methodology depended on morphological handling and extraction of associated parts to extricate districts of interest. They guaranteed that their framework delivered continuous outcomes, having wonderful precision. ***Ryan et al. [2]*** proposed a methodology that pre- owned neighborhood includes and worked on independent closer view mass sections.

Thusly, they got a complete group gauge as

the amount of the gathering sizes. They asserted that their methodology was adaptable to concealed group volumes, while it required a little preparing informational index. ***Zhang et al. [3]*** introduced a framework that utilized a vertically mounted Kinect. They abused profundity data as a methods for eliminating the impact of the appearance variety. They saw that since the head is in every case nearer to the Kinect sensor than different pieces of the body, individuals tallying task equivalents to locate the reasonable nearby least locales.

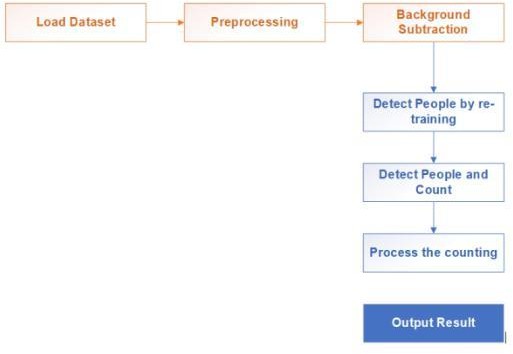
Accordingly, they built up a solo water filling strategy, ready to discover these locales while being powerful and scale-invariant. ***Zhao et al.***

1. utilized an ordinary face location calculation, upgraded by profundity data procured by a Kinect sensor. At that point, they followed individuals and checked directions. Among their perceptions we ought to stress the affectability of face recognition to changing lighting conditions. ***Brostow et al[5]*** followed straightforward picture includes and assembled them into bunches which addressed autonomously moving elements with a probabilistic methodology. ***Rabaud et al [6]*** proposed an exceptionally parallelized adaptation of the notable KLT tracker. A given video was at first handled into a bunch of highlight directions. at that point, a strategy for spatially and transiently molding was applied on the last mentioned. This portrayal was at last taken care of to an ordinary article descriptor. ***Celik et al. [8]*** explored a few methodologies for point of view contortion adjustment and proposed a strategy that depended on forefront object extraction, a viewpoint rectification and a certainty rate that guided a weighted middle channel to refine the tallies. ***Denman et al. [9]*** proposed a scene invariant group tallying calculation, whose objective was to work on different aligned cameras. Highlights between perspectives were standardized and locales of covering were redressed. They additionally explored a few highlights, for example, object size, shape, edges and key points and a few relapse models like neural organizations, K-closest neighbors and so forth They accomplished best outcomes by joining every accessible component. ***Chan et al. [10],*** expecting to ensure the security of guineas pigs, proposed a two-venture calculation, where the group was fragmented into parts of homogeneous movement, utilizing the combination of dynamic surfaces movement model. At that point a bunch of straightforward all-encompassing highlights was extricated from each fragmented district. The correspondences among highlights and the quantity of individuals per portion were gotten the hang of utilizing Gaussian Process relapse. Thusly, they didn't utilize neither article acknowledgment nor following. At last he proposed an ongoing tallying calculation. They utilized an observation camera with ordinary mounting. Their procedure joined component coordinating and point line distance approaches at the item and the identification line. They characterized a recognition line and checked individuals that enter or exit

##### III EXISTING SYSTEM

The current framework works best with shading (RGB) pictures containing thick group for example in excess of 500 heads in a picture. The current framework includes four significant segments. 1. The primary part is a CNN-based head indicator that gives a scanty area of heads and their sizes in the pictures. 2. The subsequent segment is a component classifier the picture is initial separated into equivalent size rectangular patches, which are arranged as group or not group by a Support Vector Machine SVM classifier on speeded up hearty highlights SURF highlights. 3. The third part is a relapse module that assesses the head mean each group fix dependent on its spatial facilitates and assessed head sizes. While not crowd fixes clearly have zero checks, it is conceivable that the head finder may neglect to recognize any head in a portion of the group patches. 4. This is settled by the fourth segment in which the means these group patches are assessed by the spatially reliant weighted normal of the tallies from the adjoining eight patches. The last advance is to whole all the individual fix appraisals to get a complete mean the whole picture. The current framework doesn't accept that the group fills the whole picture. Since we follow a fix approach, a portion of the patches may not contain any group. It is imperative to distinguish such fixes to dodge over assessment. To address this, we present a parallel group/not-swarm classifier. The current framework just delivers inadequate discoveries, which prompts numerous patches having no heads identified in them at all while SVM may arrange them as group fix. In the event that there are no heads identified in a fix by CNN, the assessed head size would be zero.

##### IV PROPOSED SYSTEM

The proposed framework sets the location edge and changing the camera. The discovery results underneath the edge, which is generally set from 0.2 to 0.4, won't be tallied. For straightforwardness, we utilize the default estimation of 0.2 in this work. In the genuine scene, the camera ought to be changed in accordance with the fitting tallness and point.

###### FIG 1 OVERVIEW OF THE PROPOSED SYSTEM

The proposed framework distinguishes individuals through re-preparing a convolutional neural organization. YOLO isolates the picture into a 7\*7 network and for every framework cell predicts two jumping boxes just as the certainty esteem for those cases. We accept that this division isn't adequate and we point that our calculation will be more effective in distinguishing individuals to accomplish higher tallying precision. The proposed framework is a powerful foundation deduction module is first considered to portion moving items from each caught video outline. To defeat light varieties, a powerful edge esteem related with recognizing areas of interest from the separated picture is iteratively determined by the conveyances of foundation and closer view pixels in each casing. Subsequent to acquiring the closer view areas, four states including new, leaving, blended and split are allocated to the recognized moving articles as per their appearances in the current casing. Specifically, targets recognized as conditions of consolidation and split further pass through in reverse following for calming the impediment impacts by exploring the centroid distances among objects in the past edge. At last, focuses in four states are labeled to yield the aftereffects of individuals following and tallying.

##### V MODULES DESCRIPTION

###### PREPROCESSING

Numerical activities: By thinking about the picture number juggling tasks, use of one of the standard numerical or intelligent tasks to at least two pictures is the choice. The administrators are applied in a step by-step way. That is, the estimation of the yield pixel relies just upon the estimation of the info pixel. Consequently, the size of the picture should be the equivalent. The significant bit of leeway of utilizing numerical activities is that, it is exceptionally quick and easy to execute. The same numerical activities, legitimate tasks are oftentimes used to consolidate at least two double pictures. In the situation of advanced pictures, the consistent administrator is typically applied in somewhat insightful way Convert shading pictures to grayscale to lessen calculation intricacy: in specific issues you'll see it valuable to lose pointless data from your pictures to diminish space or computational intricacy.

For instance, changing your shaded pictures over to grayscale pictures. This is on the grounds that in numerous items, shading isn't important to perceive and decipher a picture. Grayscale can be adequate for perceiving certain articles. Since shading pictures contain more data than highly contrasting pictures, they can add pointless intricacy and occupy more room in memory (Remember how shading pictures are addressed in three channels, which implies that changing it over to grayscale lessens the quantity of pixels that should be handled). One significant requirement that exists in some AI calculations, like CNN, is the need to resize the pictures in your dataset to a bound together measurement. This infers that our pictures should be preprocessed and scaled to have indistinguishable widths and statures before took care of to the learning calculation.

Another regular pre-handling strategy includes expanding the current dataset with bothered variants of the current pictures. Scaling, revolutions and other relative changes are run of the mill. This is done to amplify your dataset and uncover the neural organization to a wide assortment of varieties of your pictures. This makes it almost certain that your model perceives objects when they show up in any structure and shape.

###### EDGE DETECTION

Edges are critical nearby changes of force in an image. Edges commonly happen on the limit between two distinct areas in an image. The clear edge in the picture is the vertical line between the dark paper and the white paper. To our eyes, there is a very abrupt change between the dark pixels and the white pixels. Yet, at a pixel-by-pixel level, is the progress actually that unexpected? On the off chance that we focus in on the edge all the more intently, as in this picture, we can see that the edge between the highly contrasting zones of the picture is anything but an obvious line.

Our edge discovery technique in this workshop is Canny edge recognition, made by John Canny in 1986. This technique utilizes a rogression of steps, some joining different sorts of edge recognition. The skimageskimage.feature.canny() work plays out the accompanying advances: A Gaussian haze (that is portrayed by the sigma boundary, see presentation) is applied to eliminate commotion from the picture. (So in the event that we are doing edge recognition through this capacity, we ought not play out our own obscuring step.) Sobel edge identification is performed on both the x and y measurements, to discover the force slopes of the edges in the picture. Sobel edge discovery processes the subsidiary of a bend fitting the angle among light and dull territories in a picture, and afterward finds the pinnacle of the subordinate, which is deciphered as the area of an edge pixel.

Pixels that would be featured, yet appear to be excessively far from any edge, are eliminated. This is called non-most extreme concealment, and the outcome is edge lines that are more slender than those delivered by different techniques. A twofold limit is applied to decide expected edges. Here incidental pixels brought about by clamor or milder shading variety than wanted are killed. On the off chance that a pixel's inclination esteem – in light of the Sobel differential – is over the high limit esteem, it is viewed as a solid contender for an edge. In the event that the inclination is beneath the low edge esteem, it is killed. In the event that the slope is in the middle, the pixel is viewed as a feeble possibility for an edge pixel. Last identification of edges is performed utilizing hysteresis. Here, frail up-and-comer pixels are inspected, and in the event that they are associated with solid up-and-comer pixels, they are viewed as edge pixels; the excess, non- associated feeble up-and-comers are killed.

###### PREDICTION

A natural property of articles on the planet is that they just exist as significant elements over specific scopes of scale. A straightforward model is the idea of a part of a tree, which bodes well just at a scale from, say, a couple of centimeters to probably a couple of meters. It is pointless to examine the tree idea at the nanometer or the kilometer level. At those scales it is more applicable to discuss the atoms that structure the leaves of the tree, or the woodland where the tree develops. Likewise, it is simply important to discuss a cover over a specific scope of coarse scales. At better scales it is more fitting to think about the individual drops, which thusly comprise of water particles, which comprise of iotas, which comprise of protons and electrons and so on .

The scale-invariant component change (SIFT) is an element location calculation in PC vision to identify and portray neighborhood highlights in images. Applications incorporate article acknowledgment, mechanical planning and route, picture sewing, 3D displaying, motion acknowledgment, video following, singular ID of natural life and match moving. Also, start the checking cycle.

##### VI CONCLUSION

In the proposed work introduced a pressed ongoing individuals tallying approach named YOLO-PC. YOLO-PC improves the first convolutional construction of YOLO, and uses the smoothed out fire layer to supersede the 3 x 3 convolutional layer and the loaded model with less boundaries is procured through planning by more divisions of cells, YOLO-PC achieves extra bouncing boxes and higher acknowledgment sureness. Gotten together with a limit choice method, individuals excluding goes to be progressively proper with higher area, just as checking accuracy. The proposed framework is a YOLO based constant individuals checking approach utilizing limit determination. YOLO- PC beats YOLO as it re-trains YOLO organization, which empowers it to distinguish more boxes and accomplish higher normal certainty esteem. The limit determination in YOLOPC makes the checking more focused on and its outcome precise and quick. Taking everything into account, this strategy is exceptionally compelling and it is additionally ready to perceive immaterial individuals and overlook them in the tallying cycle. YOLO-PC has a wide scope of uses as it can help the advancement of numerous parts of the keen urban communities.

##### REFERENCES

1. C.T. Hsieh,H.C. Wang, Y.K. Wu, L.C.

Chang and T.K. Kuo. “A Kinect based people- flow counting system” In Proc. of Int. Symp. on Intelligent Signal Processing and Communications Systems (ISPACS), IEEE, 2012.

1. D. Ryan, S. Denman, C. Fookes and S. Sridharan. “Crowd counting using multiple local features”. In Proc. of Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2009.
2. X. Zhang, J. Yan, S. Feng, Z. Lei, D. Yi and

S.Z. Li. “Water filling: Unsupervised people counting via vertical Kinect sensor”. In Int’l Conf. on Advanced Video and Signal-Based Surveillance (AVSS), IEEE, 2012.

1. G. Zhao, H. Liu, L. Yu, B. Wang and F. Sun.

“Depth-Assisted Face Detection and

Association for People Counting”. In Pattern Recognition, Springer, 2012.

1. G.J. Brostow and R. Cipolla. “Unsupervised Bayesian detection of independent motion in crowds”. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2006.
2. V. Rabaud and S. Belongie. “Counting crowded moving objects”. In Proc. of Conf. on Computer Vision and Pattern Recognition, IEEE, 2006.
3. P. Kadew Tra Ku Pong and R. Bowden. “An improved adaptive background mixture model for real-time tracking with shadow detection.” In Proc. of European Workshop on Advanced Video-Based Surveillance Systems, 2001
4. H. Celik, A. Hanjalic and E.A. Hendriks. “Towards a robust solution to people counting”. In Proc. of Int’l Conf. of Image Processing, IEEE, 2006.
5. D. Ryan, S. Denman, C. Fookes and S.

Sridharan. “Scene invariant multi camera crowd counting”. Pattern Recognition Letters vol. 44, pp. 98–112, 2014.

1. A.B. Chan, Z.-S.J. Liang and N. Vasconcelos. “Privacy preserving crowd monitoring: Counting people without people models or tracking”. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008. L. Atzori, A. Iera, and G. Morabito, The internet of things: A survey, Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.
2. Mqtt protocol specification. [Online]. Available: [http://docs.](http://docs/) oasis- open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html
3. K.Y. Yam, W.C. Siu, N.F. Law and C.K.

Chan. “Effective bi-directional people flow counting for real time surveillance system”. In ICCE Proceedings, vol. 11, pp. 863-864, 2011.

1. C. Akasiadis, E. Spyrou, G. Pierris, D. Sgouropoulos, G. Siantikos, A. Mavrommatis,

C. Vrakopoulos and T. Giannakopoulos “Exploiting Future Internet Technologies: The Smart Room Case” In Proc. of International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), 2015.

1. C.C. Loy, K. Chen, S. Gong and T. Xiang. “Crowd counting and profiling: Methodology and evaluation”. Modeling, Simulation and Visual Analysis of Crowds, pp. 347-382, Springer, 2013.