

A novel & Dynamic Framework for People Counting

Yogesh Kumar L^{1*}, Jaya Dinesh GR², ShanmugaSivam S³,GowthamSethupathi M⁴

SRM Institute of Science and Technology, Ramapuram Campus
yogekick@gmail.com

ABSTRACT

Counting the number of persons in a public space provides valuable intelligence for live as well as recorded-video based monitoring and surveillance applications. When we take the case of diagonal camera setup, counting is achieved by identifying individuals or by mathematically establishing connections between values of simple image properties to the number of persons. In the current System, a people's head finder could be utilized for assessing the dimensionally changing head measurement, that is the important element utilized in our people estimation strategy. Here our idea makes use of the best-in-class C-N-N for the sparse head location in thick group of people. After dividing the given image into rectangular patches, we employ S-U-R-F feature-based S-V-M binary classifier to name every single box as having people and not-having people and remove all empty boxes. In the current framework, task generally experiences numerous issues, similar to the absence of ongoing handling of the recorded recordings or the event of mistakes because of unimportant individuals being tallied. Our proposed system overcomes the mentioned problems with a state-of-the-art real-time person counting approach referred as YOLO based People Counting. In our proposed system, after the specific pre-treatment, adaptable segmentation and feature production for the people-counting data, the feature vector is leveraged as the input of the trained YOLO to segregate and provide the statistics of the overall number of the people present.

Keywords: YOLO-PC, SVM binary classifier, Real-time people counting.

Introduction

In today's modern world of rapid globalization and social era, crowding has become a regular occurrence in every sphere of livelihood especially but not limited to, urban environments such as metropolitan cities, where population density per sq.km can reach a max of 9000. Dense crowds of people create a host of problems for law enforcement, security agencies and management authorities. A critical trouble of crowd evaluation is a correct estimate of the gang length. Multiple algorithms in the past have been kept forth within the paper for crowd length approximation. Pre-designed algorithms may be extensively known as expert-based totally and knowledge-based totally approaches. Expert tactics utilize various handmade capabilities obtained from the scenario for people and crowd counting. For lower resolution pictures and densely packed crowds, specific head counting the use of photo capabilities may be very difficult as every head has no extra than 10 to 15 pixel each. A predominant downside of texture-based strategies is that they are oblivious strategies since they no longer don't forget whether or not some people are there or not and frequently lead to excessive counts. Real-time dynamic people counting and estimation from video logs, video graphs and footages are a key component for many smart city applications and whose need has not been satisfied yet. In the existing system, a head counter can be utilized to gauge the dimensionally varying head measurement which is the

characteristic key used in their people counting method. We use the next generation convolution neural network to detect heads that are scattered in a tightly packed crowd. After breaking the picture into rectangle shaped patches, we make use of a SURF technology based SVM binary classifier to name every single patch as either crowd or non-crowd divisions and remove any patches that are not from the crowd.

In the existing system, task of people counting sometimes encounter lots of issues, because of tangential individuals being counted. The planned system overcomes the above problems with a unique period of time individual's investigation approach named Y-O-L-O-PC (Y-O-L-O is mostly based on individuals Counting). Within the planned system, once the special pre-treatment, reconciling segmentation and have extraction for the people counting data, the feature vector is employed so that the input of the trained YOLO to classifies and provides statistics of the full range of the individuals. In the existing system, task of people counting sometimes encounters several problems, because of tangential individuals being counted.

The planned system overcomes the above problems with a unique period of time individual's investigation known You only look once Technology (Y-O-L-O), within the planned system, once the special pre-treatment, reconciling segmentation and have extraction for the people counting data, the feature vector is employed because the inputs of the trained YOLO to classify and statistics of the full range of the individuals.

Literature Review

A Large-Scale Crowd Density Classification using Spatio-Temporal Local Binary Pattern.

SonuLamba, Neeta Nain October 2017. They proposed a method [4] consists of an interest points detection followed by Spatio Temporal feature extraction. They state that the rotation invariant Spatio Temporal local binary (RIST-LBP) pattern/design is introduced to extract dynamic texture of the moving crowd. Moreover, a multi-class support vector regression is adopted for density classification. They also include a tracking step which monitors the selected interest points over the video frames for crowd flow estimation. They validate their approach on three different datasets such as PETS, UCF and also CUHK that vary in density which range from low to very dense. The performance of their approach is compared with most widely used pixel-based statistics. Their approach has the advantage of less computational complexity with high efficiency in real world applications of Video-surveillance [4].

Enhanced People Counting System based Head-Shoulder Detection in Dense Crowd Scenario

Mohammed Abul Hassan, IndratnoPardiansyah, AamirSaeed Malik, Ibrahima Faye, WaqasRasheed October 2016. They proposed [2]an enhanced technique to count number of people in any crowd scenarios. This system is proposed by using an integrated feature vectors of two feature extraction methods, A Histogram of Oriented Gradients (HOG) and completed Local Binary Pattern (CLBP), to find head-shoulder region in crowd scenarios. This technique expressively improves the detection speed of the system. Moreover, an enhanced fused features technique based on confidence measure is applied to improve the accuracy performance of SVM classifier [2].

Detecting Humans in Dense Crowds using Locally-Consistent Scale Prior and Global Occlusion Reasoning

Haroon Idrees, Khurram Soomro, Mubarak Shah, October 2015, they proposed. They proposed an approach that [1] bridges the gap between holistic methods for crowds and isolated analysis of individuals from non-crowded scenes. The contribution of this paper is summarized [1] as:1)

usage of [1]locally consistent scale prior for human detection and also an approach for its own application in dense crowds, 2) a method to generate detectors comprising several parts without requiring annotations of those parts, made possible with the use of Latent SVM, 3) occlusion reasoning in crowds with a worldwide solution, 4) a fresh and challenging dataset of dense crowd pictures with tens of thousands of humans[1].

Scale-adaptive Real-time crowd detection and counting for drone images.

[3] Markus Kuchhold, Maik Simon, Volker Eiselein and Thomas Sikora, February 2018, they proposed an idea of [2]a scale-adaptive crowd detection and counting approach for drone images. Based on local feature points and density estimation considering the image scale, they detect dense crowds within multiple distances and introduce an extremely fast and quick counting strategy with high accuracy for their detected crowd regions. They compare their results with a recent CNN-based state-of-the-art approach and validate both methods for different scaling factors on a novel crowd dataset. The results show that their method outperforms the pre-trained CNN-based approach and receives very accurate counting results for different zoom factors, resolutions and crowd sizes. Their method's low computational complexity makes it highly suitable for real-time analysis or embedded systems [2].

Tracking People in Dense Crowds using Supervoxels.

ShotaTakayama, Teppei Suzuki, Yoshimitsu Aoki, ShoIsobe, Makoto Masuda, October 2018, they proposed a concept of [5] combination of supervoxels and optical flow tracking. The SLIC based supervoxel algorithm adaptively estimates the border between a person and a background. Thus, the combination of supervoxels and optical tracking becomes an extremely reliable approach for crowd tracking. In tracking experiments, higher performance is achieved for the UCF crowd dataset [5].

Existing System

The current framework works best with RGB pictures with conservative group i.e., more than 400 heads at an image. The current framework includes four fundamental segments.

1.The chief part is a CNN-based head identifier which gives you a deficient space of heads and their sizes in the photographs.

2.The second fragment is a component classifier the image is first confined into identical size rectangular patches, which are masterminded as gathering or not gathering by a Surf-Vector-Machine classifier on speeded up astonishing qualities SURF features.

3.The third part is a backslide module which give us the check of the head for each gathering fix reliant upon its spatial puts together doubtlessly have zero counts it's conceivable that the head identifier may disregard to perceive a segment of the heads several the gathering patches.

4.This is tended to by the fourth fragment where it implies these gathering patches are surveyed from the spatially dependent weighted typical of the counts from the abutting eight patches' spots. The last development is to aggregate the sum of the individual fix evaluations to get the entire mean the whole picture.

The current structure doesn't acknowledge that the gathering fills the whole picture. That is because we follow a fix system, very few of the patches may not contains any gathering. It's important to recognize such kind of patches to thwart over evaluation. To deal with this, they introduced a matched gathering/not-swarm classifier. The current structure simply makes saves acknowledgment, which in this way prompts a huge load of patches with no heads perceived in

them at all while SVM may portray them as gathering fix. If there aren't any heads perceived in a given fix by CNN, the surveyed head size would be zero.

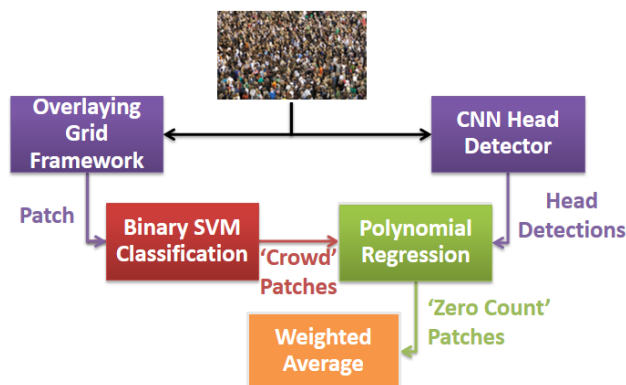


Figure 1.Existing Head Counting Algorithm

Components used are listed in the following:

- Convolutional Neural Networks
- SVM classifier

DISVANTAGES OF EXISTING SYSTEM

- It can able to keep the temporary information of moving crowd
- Some impediment actually occurred in a thick packed circumstance situation.
- Viewpoint impact and area of a camera which is by and large over your head.
- Scale between different frames may fluctuate because the drone.
- Double-person Detector raises the rate of incorrect detection.

PROPOSED ALGORITHM

The proposed system sets the ID edge and changing the camera. The acknowledgment results under the cut-off, which is ordinarily set from 0.2 to 0.4, won't be counted. Considering a real worry for ease, we use the default worth of 0.2 in this method. In the veritable scene, the camera ought to be adjusted to the reasonable point and height. The proposed structure perceives people through re-setting up a convolutional neural association. YOLO isolates the image into a 7*7 system and for each organization cell predicts two skipping boxes despite the sureness a motivating force for every last one of those containers. We acknowledge that this division isn't satisfactory and hence we point that our computation will be more capable in distinctive people to achieve higher counting precision. The proposed system is an amazing establishment allowance module which is first considered to partition moving articles out of each recorded video layout. To beat light assortments, a one-of-a-kind edge regard related to distinguishing

districts of interest from the perceived picture is iteratively controlled by the appointments of establishment and closer view pixels in each packaging. In the wake of gaining the front facing region areas, four states including new, leaving, merged and split are having been designated to the perceived moving things reliant upon their appearances in the current edge. Specifically, targets recognized as states of union and split further pass-through in turn around following for facilitating the obstacle impacts by analysing the centroid distances among objects in the past layout. At last, centres in four states have been marked to yield the delayed consequences of people checking and following.

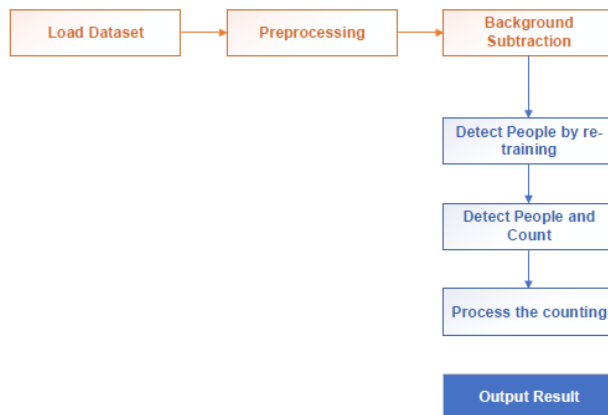


Figure 2.Algorithmic Flow Diagram

Modules

Dataset Processing

Tqdm packages among the more comprehensive packages to display progressive bars with using python and is useful for all those cases that you would like to create scripts that keep the users informed about the status of your program. Tqdm |operates on almost multiple platforms and operating systems such as Windows OS, Linux distribution, Macintosh OS etc., and any device having a Graphic User Interface.

The train-test split Procedure is proper whenever you have a tremendous dataset, an expensive model to prepare, or require a decent gauge of model usefulness quickly. The interaction includes taking a dataset and partitioning it into two subsets. The underlying subset is utilized to fit the model and can be known as the preparation dataset. The subsequent subset isn't used to prepare the model; all things being equal, the information component of this dataset is given to the model, and afterward expectations are made and contrasted with the normal qualities. This second dataset is known as the test dataset.

Train Dataset: Used to Fit the AI model.

Test Dataset: Used to Assess the fit AI model.

The Aim is to measure the exhibition of the AI model on new information: information not used to prepare the model. By default, the program ignores the initial order of data. It randomly selects data to form the test and training set, which is typically, a desirable characteristic in real world applications to prevent possible artifacts existing in the data preparation procedure. To disable this feature, just set the shuffle parameter as False (default = True).

The skimage.io image Package can be used to read the image from the file. Rescale action resizes a Picture with a specified scaling variable. The scaling variable can either be one Floating-point value, or several values - one along each axis. Resize acts as the exact same purpose, but enables to define an output image shape rather than a Scaling factor.

Pre-processing

Some sort of improvement is needed to separate among black and foundation lesions. Interestingly, pre-handling isn't about edge preparing, but about improving the difference. A coordinated filter with a two-dimensional Gaussian piece has high responsiveness to dark notwithstanding brilliant sores that can be shaped individually as step borders and Gaussian. Along these lines, LoG filters together. Coordinated filters are used to recognize the transient brilliant injuries and the Gaussian power like dull sores. Morphological terminations are conveyed performed on the picture viable to keep up the brilliant locales. It smooths the substrate smothers and parts the thin (need out) vascular organizations so that light lesions stay all things considered, unaltered. This cycle, in any case, diminishes the image contrast. The picture is moved through an ideal wideband band pass channel construction to improve the differentiation of the Exudates. The bend change is productive to characterize a flat, inclining point and upward, directional data, shapes, missing, curves, and off base limit information, and so on. The image is first disintegrated into a few sub groups with a bend change. The estimated sub band is stifled and some enhancement factor expands the leftover (detailed) sub strip. This builds up the fringe of the dim injuries that upgrades their partition from the foundation.

Prediction

Image Data Generator class allows allow rotation of up to 90 degrees, horizontal flip, horizontal and vertical shift of the data. We need to apply the training standardization over the test set. Image Data Generator will generate a stream of augmented images during training. We will define Exponential Linear Unit (ELU) activation functions A single fully-connected layer after the last max pooling. The padding='same' parameter. This simply means that the output volume slices will have the same dimensions as the input ones. Batch normalization provides a way to apply data processing, similar to the standard score, for the hidden layers of the network. It normalizes output of the discrete layers for each one of the mini-batch (hence the name) in a way, which maintains its mean activation value is close to Zero, and its standard deviation close to One. We can use it with both convolutional and fully connected layers. Networks with batch normalization train faster and can use higher learning rates.

Result

Thus, by employing the pre trained YOLO model one can give the people count in a given crowd much faster than the CNN method. The proposed system is fast, light weight and efficient method. The lesser number of epochs cuts down the training time drastically with that we can achieve higher learning curves, Hence the Pre-Trained YOLO Model has mAP of 56.7 which is higher compared to the CNN with a 44.5. Thus, the Proposed system is much efficient.

CONCLUSION

Our system is a Y-O-L-O based current time dynamic people checking procedure that which makes us of cut-off assurance. Y-O-L-O-PC surpasses Y-O-L-O as its re-trains -Y-O-L-O association, which enables it to perceive essentially more boxes and show up at higher typical

sureness rate. The cut-off assurance improves appraisal zeroed in on and consequently the result is more accurate and speedier. In light of everything, this methodology is doable and it is besides set up to see inconsequential individuals and excusal them in the checking cycle. Y-O-L-O-PC has a wide degree of occupations as it can help the progress of different metropolitan zones and help in various features of present day adroit metropolitan cities.

References

- [1] H. Idrees, K. Soomro and M. Shah, "Detecting Humans in Dense Crowds Using Locally-Consistent Scale Prior and Global Occlusion Reasoning," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 1986-1998, 1 Oct. 2015, doi: 10.1109/TPAMI.2015.2396051.
- [2] M. A. Hassan, I. Pardiansyah, A. S. Malik, I. Faye and W. Rasheed, "Enhanced people counting system-based head-shoulder detection in dense crowd scenario," 2016 6th International Conference on Intelligent and Advanced Systems (ICIAS), 2016, pp. 1-6, doi: 10.1109/ICIAS.2016.7824053.
- [3] M. Küchhold, M. Simon, V. Eiselein and T. Sikora, "Scale-Adaptive Real-Time Crowd Detection and Counting for Drone Images," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 943-947, doi: 10.1109/ICIP.2018.8451289.
- [4] S. Lamba and N. Nain, "A Large-Scale Crowd Density Classification Using Spatio-Temporal Local Binary Pattern," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2017, pp. 296-302, doi: 10.1109/SITIS.2017.57.
- [5] S. Takayama, T. Suzuki, Y. Aoki, S. Isobe and M. Masuda, "Tracking People in Dense Crowds Using Supervoxels," 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2016, pp. 532-537, doi: 10.1109/SITIS.2016.90.