International Journal of Research in Engineering, Science and Management
Volume-3, Issue-8, August-2020
journals.resaim.com/ijresm | ISSN (Online): 2581-5792 | RESAIM Publishers

Crowd Counting Using Deep Learning

J. Vimala Devi[1], L. Chandni[2*], Mithun Madhav Prabhu[3], S. Chethana[4], Sampath Kumar[5]
[1]*Associate Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*
[2,3,4,5]*Student, Dept. of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*
*Corresponding author: chandni2426@gmail.com

***Abstract*: Crowd counting is a challenging task, due mainly to the severe occlusions between dense crowds. This work aims to take a broader view to address crowd counting from the semantic modelling perspective. Crowd counting is essentially a function of pedestrian semantic analysis involving three key factors: pedestrians, heads, and their context structure. Different body parts information is an important clue to help us judge whether a person exists at a given position. Existing methods usually perform crowd counting from the perspective of directly modeling the entire body or the heads only, without explicitly capturing the composite body part information on semantic structure that is critical to crowd counting. In our approach, we first formulate the key factors of crowd counting as models of semantic scene. Then we convert the problem of crowd counting into a problem of multi-task learning, so that the models of the semantic scene are turned into different subtasks. Lastly, in a unified scheme the deep convolution neural networks (CNNs) are used to learn the subtasks. In terms of pedestrian semantic analysis, our approach encodes the semantic nature of crowd counting and provides a new solution. Our methodology outperforms state of the art approaches in experiments on four benchmark crowd counting datasets: 1. Sets of synthetic data 2. Databases-that's right Real-time data sets. The knowledge about semantic structure is proved to be an effective cue in crowd counting scene.**

***Keywords*: Crowd counting, Crowd detection, Convolution Neural Network, Object detection.**

1. Introduction

Crowd counting is a difficult task in the exact numbering of crowds in dense scenes. A person getting hidden by another is extreme, and the change in appearance of perspective in areas vary considerably. Therefore, proportions of crowds are physically different. The theory makes use of pedestrian semantics featuring three during the process of counting the crowd main matters: the audience, the heads and their body composition. Many of the current approaches concentrate on modeling the characteristic appearance of ground features of entire crowds or even of head alone. Although overlooking the context composition of various sections of the body which is also essential for crowd counting. In this work we present the problem of counting the crowd from the perspective of semantic modeling. The three key factors of crowd counting are formulated as two types of semantic scene models.

In this project, the first semantic scene model is denoted as the map of the body part. This models the functional appearance of pedestrian body parts and their conceptual structure. A separate semantic category is incarnated for each pedestrian's body part in the body part map whereas, the meaning description of the multiple spaces in different sections is also developed in the plot. Figure (a) offers you an insightful view of our strategy. The part of the body map is produced using a single pedestrian parsing algorithm, a pre - trained CNN network model that calculates the segmentation mask of even an input pedestrian picture. This is also a strong reason where we combine all the masks of pedestrian segmentation to create a body part diagram. This project denotes an ordered map of density as the second model of the semantic scene. In existing works, density distributions of crowds are modeled by using conventional density maps while ignoring individual pedestrians body structure. A standardized density map is created based on unique pedestrian shapes given by the component map of the body.

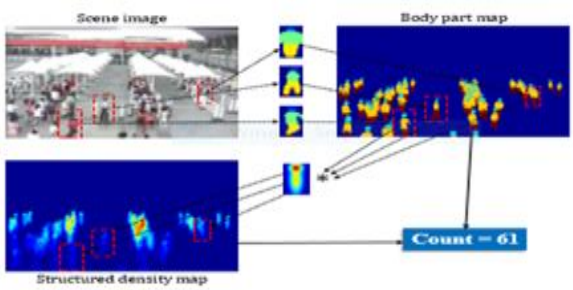


Fig. 1.

The key objective of the standardized density map is to model semantine structure details in fine-grained format as an enhancement of the traditional density map, thereby allowing more precise entity labels. To reliably quantify the amount of pedestrians we reformulate crowd counting as a question of understanding multitasks. 3 main-tasks exist: to determine different forms of semantic scene structures and to approximate the crowd count. We create deep convolution neural networks (CNNs) to understand certain sub - tasks combined. Firstly CNNs model the mappings from captured images to semanthetic scene models which include body part map and organized density map, then measure the crowds count based on them. The CNNs are able to pull effective graphical images

from images.

2. Proposed System

The papers related to our research are presented in this section. Secondly, the strategies of crowd counting suggested in existing works of literature are discussed. But then we are reviewing some relevant literature on pedestrian semantine study, while they tackle the problem on crowd counting across a pedestrian semantine analysis viewpoint throughout this project. We further incorporate the background of CNNs, as our method of crowd counting focuses mostly on deep neural networks. Crowd count: Many types of crowd control can usually be classified into three groups: (1) standard identification (2) global regression and (3) estimate of the density. The prior crowd counting diagrams indicate the techniques of predicting pedestrian semantic structure simultaneous detection. Diverse types of detection are being used to match specific pedestrian pictures. Detection-based approaches perform best in comparatively low image sequences while strong variations limit them in large crowds. Unlike them, we structure the conceptual structure of pedestrians as models of syntactic and semantic scenarios. Under the crowded set, they are much more receptive to learning, and can be more suited to the deep learning context.

Experts are exploring a better route to addressing challenges of detection-based techniques within high-density situations via integrating spatial regression-based optimization techniques that relate among both low-level features as well as pedestrian statistics. In crowded conditions all these strategies are much more efficient than solutions based sensing. Regression techniques are also widely used. This includes regression models, Bayesian regression, ridge regression and the Gaussian method of regression. International regression-based models are still using knowledge regarding pedestrian numbers, while knowledge about spatial data and physical appearance by pedestrians is overlooked.

Studies articulate that eventual global characteristics of pedestrians as an implicit data packet for predicting crowd location data, including crowd counting based on projections of density. Lempitsky and Zisserman formulate a network model focused mostly on illustrated objects with such a Two dimensional scaling factor, and develop a regression analysis function between some of the path diagram and the scene image. Such approaches exhibit good output on crowd counting. And even in that methodology, the body-part arrangements of different pedestrian are overlooked from either the viewpoint of syntactic and semantic modeling. Within this work, we concentrate on exploring the symbolic nature of object recognition. We create semantic scene structures by extracting abundant syntactic and semantic framework features from data and using these to subsequently structured accumulation rate tags.

A. Pedestrian semantic analysis

In many practical implementations in smart security systems working in real-world settings, pedestrian semantic interpretation is a significant precondition, including several typical computer vision topics such as pedestrian detection. Individual parsing and crowd segmentation. In this review, by concentrating on pedestrian semantic research, we tackle the issue of crowd counts as the visual signs of pedestrian body-part involvement may provide ample knowledge to classify crowds. This idea is also demonstrated by the success of parts based pedestrian detection methods. That idea is also demonstrated by detection, instead of detecting the holistic pedestrian directly, the part-based methods use information on the pedestrian body structure and are able to handle occlusions with greater robustness. Unlike traditional parts-based approaches, in our methodology we formulate pedestrian body-part semiconductor structure as semiconductor models which are more suited for learning under deep neural network frameworks and are more efficient and stable in dense crowded scenes.

B. Convolution Neural Networks: The CNNs build on our crowd counting framework

The CNNs are a successful and resulting visual representation tool, being able to learn effective but easy to interpret visual images. Through distinct, we should use Completely fully convolution Networks (FCNs) to understand the semantine scene definition as outlined in our approach. FCNs are end-to - end models as a kind of pixel-wise CNN architecture. We aim now at state-of-the-art results of certain set of circumstances experiments including situation sorting, segmentation of audiences, and prediction of behaviour. The classifiers relate to discrepancy in pedestrian's volume to accomplish state-of-the-art output on the datasets.

3. Implementation

A. Problem wording

Through this work we are trying to fix the topic of single crowd photo counting. Furnished with a crowded frame, our task is to estimate the number of pedestrians on the scene. From the semantic modeling point of view, we reinterpret the original problem as a semi - supervised article mentioning three subtasks: inferencing two forms of textual scene, and measuring the number of pedestrians. The very first semant scene model tends to be the body component built to model semantine body structures for pedestrians. The second is the generic density diagram, designed to construct the density distributions and pedestrian shapes. Such two models concentrate, respectively, on data encoding of specific semantine characteristics for a crowd image. To tackle the multi-task learning problem, we create the CNNs combined in a single framework to learn the three subtasks.

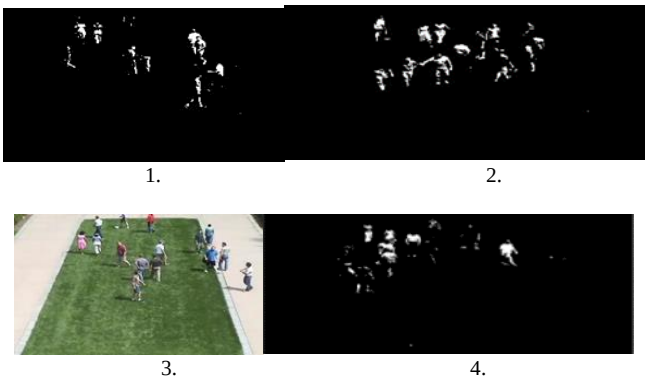
B. Body Part Map

The body part map is suggested as one of the semantic scene models to model the body-part semantic systems of individual pedestrians, which may serve as an integral basis for deciding whether a person occurs at a particular location. Throughout our system, we incorporate it as a novel controlled mark for solving crowd counting issues.

Body part map is created using the picture, perspective map and head point positions of the provided scene. Next, we will get single pictures of the pedestrians. Instead of the viewpoint differences, we use the viewpoint diagram to normalize the dimensions of the pedestrians. The pixel value denotes the number of pixels in the shot, reflecting one meter at the actual scene spot. With a person's head position, the top-left corner and bottom-right corner of the person's boundary are estimated as,

$$P_{tl} = (P_h^x - \alpha_1 \mathcal{M}(P_h), P_h^y - \beta_1 \mathcal{M}(P_h))$$
$$P_{br} = (P_h^x + \alpha_2 \mathcal{M}(P_h), P_h^y + \beta_2 \mathcal{M}(P_h)) \quad [6]$$

Where the parameters are manually set in the experiments to best approximate the actual situations. After the pedestrian images have been obtained, we normalize them to the same size, then input them into the single pedestrian parsing model to calculate their semantic segmentation masks. The pedestrian parsing model uses a deep neural network to parse a single pedestrian image into several semantics regions, including hair, head, body, legs and feet. The model is pre-trained that we use it solely to generate pedestrian semantic segmentation. We merge the hair regions into the head, and fuse the feet regions into legs as well. Finally, to create the body part map we redimension individual pedestrian to their original sizes. Illustrations, shows a chaotic scene shot, and a body part diagram referring to the scene picture. The body part maps which include named four-category pixels' model in the images of the scene the semantic structure of each individual pedestrian.



1. 2. 3. 4.

Fig. 2. Illustrations: 1. Annotations, 2. BodypartMap, 3. Conventional Density Map, 4. Structured Density Map

C. Structured density map

So it is suggested to represent all the mass distributions and pedestrian forms in the standardized mass diagram. Data-dependent in our approach is that it is created according to individual pedestrian shapes. The traditional density map proposed in current works is discussed first. This is created usually from a total of 2D Gaussian kernels centered on pedestrian locations:

$$\mathcal{D}_{\mathcal{N}}(p) = \sum_{i=1}^{C} \frac{1}{\|Z\|} \left(\mathcal{N}_h^i(p; P_h^i, \sigma_h^i) + \mathcal{N}_b^i(p; P_b^i, \sigma_b^i) \right), \quad [6]$$

Where (NH) is a standard 2D Gaussian kernel to model a pedestrian's head, and (Nb) is an abbreviated 2D Gaussian kernel to model a pedestrian's part of the body. Here Example. 2(c) show a sight pictorial following Density Map DN. The uniform density map D is further introduced because DN could not adequately represent several different forms of individual pedestrians.

The form of each pedestrian is defined by the pedestrian mask. It is obtained by finalizing part map of the body, where the pixel values of foregrounds and backgrounds are set to 1 and 0, respectively. The structured density map D is computed by the multiplication of DN and BM by element followed by normalization. Illustration. 2(d) shows the structured density map our approach generates. Compared with conventional map of density in Fig. 2(c) We can see that the structured density map not only denotes latent crowd density distributions but also preserves specific shapes of each pedestrian.

D. Multi-task crowd counting framework

To approximate the overall number of pedestrians, we reformulate the original question of group measurement as a multitask teaching concern with three subtasks: the prediction of semanthetic environment models, and the evaluation of the quantity of pedestrians. We propose a single multi-task learning system, based mostly on CNN models, for learning all 3 sub - tasks together. They take a patch-wise approach, in which network data is a picture patch taken from the context of the scene.

As demonstrated in table, in deep based on deep convolution neural networking there's only the one sort of projection from sight map to video frames in standard learning algorithm based on density estimation process. In our multi-task learning system, they introduce more background to remember replacement scene part of the body model. Two stream outputs are concatenated together and mapped into fully - connected models into another count of the pedestrians.

Vibe is a pixel-based background subtraction technique that innovated by introducing several new mechanisms: (I) segmentation by comparing a pixel value to a small number of samples previously collected; (2) memory less updating policy; and (3) spatial diffusion. This project sets out the underlying ideas that motivated us to build Vibe. One of the main keys to building and updating the background models is the absence of any notion of time. Additionally, vibe introduces a spatial diffusion mechanism, which consists of modifying a neighboring pixel model while updating a pixel model. We

have shown that using spatial diffusion, increases performance at all times and that, in for the Change Detection dataset in particular, it is much easier to disperse a pixel value in the vicinity than to use it to update its own layout. In addition, the crossing of background boundaries in the segmentation map to adapt foreground pixel models is also beneficial for suppressing ghosts or static objects.

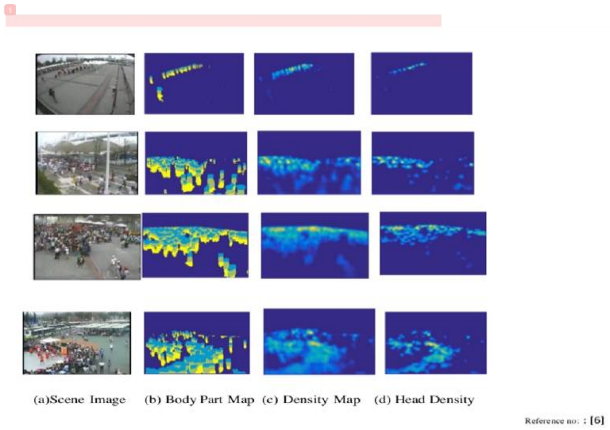


Fig. 3.

4. Results

Framework for selection of video and background extraction of frames for a given video and processing the sample video.

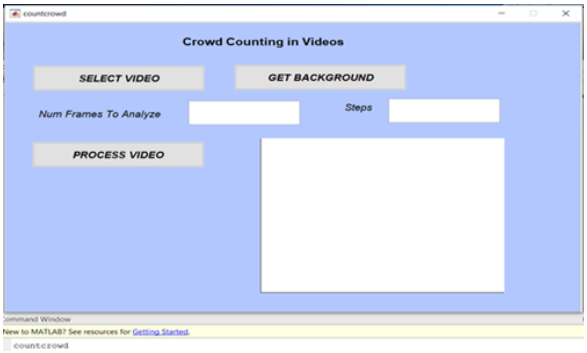


Fig. 4. Selection of video

The number of steps for each frame count and background extraction of the sample video will be achieved once the video is processed.

Fig. 5. Processing of video with respect to frames

Processing of the crowd count happens within another window and each frames count will be calculated one at a time.

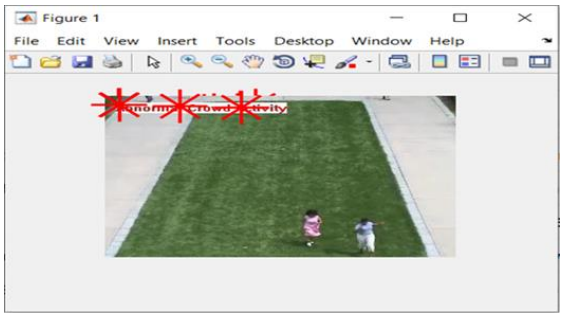


Fig. 6. Analyzing of individual frames

The end result that is obtained is the frame numbers with the crowd count for each particular frame and this will be stored a different folder.

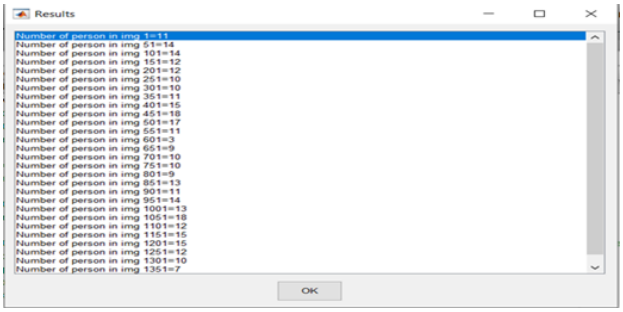


Fig. 7. Crowd Count in selected video

5. Conclusion

We've provides a qualitative procedure within that document to accurately estimate crowd count in pictures. Our methodology has centered on figuring out the semantic character of crowd counting. We developed two semantic scene models to recover features from images that is abundant with semantics. Furthermore, we have reformulated the problem of image classification as a multitask learning problem, so that the templates of the based image retrieval scene have been transformed into separate sub-tasks. We set up the CNNs in a single scheme to mutually master certain series of tasks. Our methodology has accomplished technological improvements in simulations compared with the state-of-the-art approaches on four benchmark datasets.

References

[1] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors,” in Proc. IEEE ICCV, vol. 1, 2005, pp. 90-97.
[2] Z. Lin and L. S. Davis, "Shape-Based Human Detection and Segmentation via Hierarchical Part-Template Matching," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pp. 604-618, April 2010.
[3] C. Zhang, H. Li, X. Wang, and X. Yang, “Cross scene crowd counting via deep convolutional neural networks,” in Proc. IEEE Conf. CVPR, 2015, pp. 833-841.

[4] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in Proc. IEEE Conf. C VP R, 2016, pp. 589-597.
[5] P. Luo, X. Wang, and X. Tang, "Pedestrian parsing via deep decompositional network," in Proc. IEEE ICCV, 2013, pp. 2648-2655.
[6] Siyu Huang, Xi Li, Zhongfei Zhang, Fei Wu, Shenghua Gao, Rongrong Ji, and Junwei Han, "Body Structure Aware Deep Crowd Counting."
[7] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in Proc. Adv. NIPS, 2010, pp. 1324-1332.
[8] W. Ge and R. T. Collins, "Marked point processes for crowd counting," in Proc. IEEE Conf. CVPR, 2009, pp. 2913-2920.
[9] M. Wang and X. Wang, "Automatic adaptation of a generic pedestrian detector to a specific traffic scene," in Proc. IEEE Conf. CVPR, 2011, pp. 3401-3408
[10] V. B. Subburaman, A. Descamps, and C. Carincotte, "Counting people in the crowd using a generic head detector," in Proc. IEEE Conf. AVSS, 2012, pp. 470-475.