IMPROVED COUNTING AND LOCALIZATION FROM DENSITY MAPS FOR OBJECT DETECTION IN 2D AND 3D MICROSCOPY IMAGING

Shijie Li, Guido Gerig
New York University
Computer Science and Engineering
Brooklyn, NY, USA

Thomas Ach
University Hospital Bonn
Ophthalmology
Bonn, Germany

ABSTRACT

Crowd object counting and localization are key steps for quantitative analysis in large-scale microscopy applications. Previous methods producing density maps based on deep learning have reached a high level of accuracy for object counting by assuming that object counting is equivalent to the integration of the density map. However, this model fails when objects overlap significantly and/or present fuzzy boundaries regarding accurate localization. To overcome this limitation, we propose an alternative method to count and localize objects from the density map. Our procedure includes the following three key aspects: 1) Proposing a new connected-component analysis method based on the statistical properties of the density map, 2) optimizing the counting results for those objects which are well-detected based on the proposed counting method, and 3) improving localization of poorly detected objects using the proposed counting method as prior information. Validation includes processing microscopy data with known ground truth and comparison with other models that use conventional processing of the density map. Our results show improved performance in counting and localization of objects in 2D and 3D microscopy data. Furthermore, the proposed method is generic, considering various applications that rely on the density map approach. Our code is publicly available.

Index Terms — object counting, object localization, density map

1. INTRODUCTION

Object counting and localization techniques are essential processing steps for highly automated quantitative analysis of large-scale microscopy images driven by biological and medical applications. The main aim is to accurately predict the number and location of target objects, such as cells, nuclei, and granules, and to identify these objects for subsequent semantic segmentation. Such tasks are often based on simple point annotations, which take significantly less effort compared to precise outlining of sought objects. The use of point annotation becomes particularly important when objects are densely clustered, but also for object analysis from 3D imagery where manual outlining may not be feasible.

Supervised object counting and localization methods fall into two categories: counting [1, 2, 3, 4, 5, 6, 7] and detection emphasized [8, 9, 10]. The former faces limitations with overlapping objects regarding counting performance. Modern crowd-counting methods are based on the training of a regression method that maps images to density maps, the integration of which is shown to be equivalent to the number of crowd objects. Placing a Gaussian kernel for each annotated dot is the most common approach to generate the density map. Lempitsky et al. [2] train a linear transform from the feature space to the density map space. Others [7, 11, 6] proposed convolutional neural networks as feature extractors. Cohen et al. [4] generate the density map with a square kernel instead of a Gaussian kernel. Improving the objective function for optimization was proposed by [3, 5]. Recent work by Guo et al. [1] shows superior counting performance on various cell datasets. They use the connected components after thresholding or take local maxima to localize the objects. However, the
evaluation of localization performance in [1] indicates that these postprocessing steps fail to predict the object’s center and tend to produce false positives when occlusion occurs, and do not seem to be robust enough to background noise. Furthermore, those methods are sensitive to hyperparameters such as threshold values.

This paper proposes a counting method that automatically determines the threshold to estimate the number of objects in a connected component. Furthermore, we use Gaussian mixture fitting to decode this information from the predicted density map. These two methods result in a postprocessing approach to the density map without needing additional hyperparameters. We validate our approach on multiple 2D and 3D datasets and demonstrate that the proposed method provides improved density maps and increases counting and localization performance compared to previous baselines.

2. DENSITY MAP ANALYSIS

In this section, we briefly summarize the basic idea of the density map (DM) approach. We then highlight the defects of connected-component labeling of the thresholded density map to localize objects.

2.1. Density Map Overview

Density map methods assume that N images \( I_1, I_2, \ldots, I_N \) \( \in \mathbb{R}^{H \times W \times C} \) and their corresponding annotations which are N sets of 2D/3D points \( P_1, P_2, \ldots, P_N \) are provided. For each point annotation set \( P_i \), we generate a density map \( D_i \) \( \in \mathbb{R}^{H \times W} \) as follows,

\[
D_i = \sum_{j=1}^{C_i} N(p_i^j, \Sigma)
\]

where \( C_i \in \mathbb{Z} \) is the number of objects in the image \( I_i \), \( p_i^j \in P_i \) and \( \Sigma \) is the covariance matrix often related to the object size and the image isotropy. The learning step of the DM method is a training of a regressor mapping from the image \( I_i \) to the density map \( D_i \). During testing, the counting prediction \( \hat{C}_i = \sum_{x} \hat{D}_i \) is computed, where \( \hat{D}_i \) is the predicted density.

2.2. Connected-Component Labeling of the Thresholded Density Map

Localization of objects from the predicted density map \( \hat{D}_i \) is based on decomposing it by solving,

\[
\min_{p_i^j} \| D_i - \sum_{j=1}^{C_i} \hat{D}_i^j \|, \quad \text{s.t.} \quad \hat{D}_i^j = N(p_i^j, \Sigma).
\]

Applying connected-component labeling (CCL) to the thresholded density maps is one simple and effective way to estimate the optimal solution to this problem. However, it ignores the constraint in Eq. 2. It thus has the risk of undercounting several objects into one or overcounting by assigning background noise to an object, which would cause inconsistency between the counting and the localization from the density map. Furthermore, a particular threshold value has to be tuned manually during the training stage for different datasets. Therefore, we propose a robust approach to determine the threshold automatically and localize the objects with more substantial consistency with the counting prediction from the density map.

3. METHODOLOGY

Fig. 1 shows the entire pipeline. We first train a deep network [1] to produce density maps from the input images. Then after binarizing the density map with an automatically calculated threshold and assigning connected-component labels, the proposed connected-component analysis step predicts the number of objects in each connected component. Finally, the consequent center estimation module localizes the object centers for connected components containing one or more objects. This section will focus on explaining the connected components analysis and object center localization.

3.1. Connected-Component Analysis

If only one object is in the image and the predicted density map is perfect, the predicted density map \( \hat{D}_i = N(p_i^1, \Sigma) \). The thresholded density map can be written as \( 1_{D_i > T} \hat{D}_i = 1_{D_i > T} N(p_i^1, \Sigma) \), where \( 1(.) \) is an indicator function. Following the property of gaussian distribution, we can get

\[
\sum_{x} 1_{D_i > T} \hat{D}_i = F(r_T), \quad \text{where} \ x \ is \ the \ pixel/voxel \ location, \ F(.) \ is \ the \ Mahalanobis \ cumulative \ function, \ r_T \ is \ the \ Mahalanobis \ distance \ satisfying \ N(r_T; 0, \Sigma) = T. \ We \ can \ conclude \ that, \ in \ an \ ideal \ situation \ if \ a \ connected \ component \ in \ a \ thresholded \ density \ map \ contains \ only \ one \ object, \ it \ must \ satisfy
\]

\[
\sum_{x} 1_{D_i > T} \hat{D}_i = F(r_T) = 1
\]

It is obvious that if the connected component contains more than one object, the right-hand part of Eq. 3 would be larger than 1. Therefore the value of the right-hand part of the equation can be used as the criteria to decide the number of objects in a connected component. In real applications, some areas of the predicted density map would show very small peaks, making the criteria value very small. Our proposed method will identify these areas as regression noise and discard them. According to [12], the threshold \( T \) can be determined by \( N(r_{T_0}; 0, \Sigma) \), where \( F(r_{T_0}) \approx 0.99 \). To be noticed, it is applicable in this specific case where the gaussian kernel is used.
3.2. Object Center Localization

One important observation we can leverage for localization is that centers of objects will most likely show peaks in the density map. Hence, we apply a similar method as in [13], where we first normalize the density within the connected component area to treat it as a probability distribution. We then fit a Gaussian mixture model to the distribution. The centers of objects are then the fitted means of the Gaussian mixture.

Fig. 2. Circles represent the ground truth of object locations, and red stars mark the predicted object locations. Boxes mark locations where methods differ. Histograms (bottom row) show posterior probability distributions of predicted versus ground truth counts calculated from the testing split of the datasets.

4. EXPERIMENTS

Here, we introduce datasets used for verification, describe experimental configuration and evaluation metrics, and show qualitative and quantitative results.

4.1. Datasets

For a full comparison, all the 2D datasets as also used in [1] are included for evaluation. Furthermore, we generate and use synthetic data where we can control occlusions, applying the tool from [14]. We will also apply procedures to a volumetric Structural Illumination Microscopy (SIM) dataset to evaluate the method’s performance in a clinical 3D application.

VGG Simulated bacterial cells in fluorescence-light microscopy image using [14]. There are 200 256 × 256 images with 176(±61) cells in each image.

Synthetic Ellipse Dataset By adding an eccentricity parameter to [14], we generate a dataset containing object shape with eccentricities from 0 to 0.85. For each image, we allow a proportion $p$ of objects to overlap, where $p \sim U[0, 0.5]$.

MBM The dataset was first introduced by [15] containing eleven 1200 × 1200 images of healthy human bone marrow from eight different patients. The center point of the cell nuclei has been annotated. [4] further crop it into 44 600 × 600 images with 126 ± 33 cell nuclei in each image.

DCC The Dublin Cell Counting datasets is quite challenging. It contains 177 images of stem cells from various tissues and species shot by different devices. Data therefore shows large variability in regard to the density, morphology, size and artifacts. The number of cells varies from 0 to 101.

ADI We use images of human subcutaneous adipose tissue from [16]. [4] cropped 200 regions of size 1700 × 1700 and downsampled them to 150 × 150. There are 165 ± 44 cells on average in each image.

BBBC027 We used image set with low SNR in the BBBC027 [17] dataset for 3D validation. It synthesized 3D digital phantoms in human colon tissue. The weight center of the instance label is calculated as the point annotation.

SIM RPE cells were imaged with super-resolution structured illumination microscopy (SIM) as part of a recent ophthalmological study [18]. Our biomedical partners localized each organelle per cell with point annotations. There are 420 images in total, with 424 objects on average in each image.

4.2. Experimental setup

We compare our proposed pipeline with those commonly used methods to count and localize: 1) integration of density map (IoDM) and 2) connected-component labeling with the threshold (CCL-T). We train with similar steps as proposed in [1] and select the best models with the smallest counting mean square error. We then compare the counting performance with IoDM, CCL-T, and our proposed method. We find the best threshold of CCL-T during the validation step for each dataset. The image size, batch size, data augmentation, selection of the $\Sigma$ in Eq. 1 and the scaling coefficient of the loss function are the same as [1]. To rule out the randomness of results, we run ten different data splits for each dataset and average the evaluation metrics. All experiments are carried out on a desktop computer with an AMD Ryzen 9 5950X desktop processor and a single NVIDIA GeForce RTX 3080 GPU. Each experiment takes about one hour to complete.

2The original tensorflow version is available at https://github.com/mzlr/sau-net.git. We reproduce results with pytorch.
4.2. Evaluation Metrics

We use Precision, Recall, and F-measure to evaluate localization performance of the models and MAE for the counting ability. A proposed object center is True Positive (TP) if and only if its distance to the nearest annotation points is less than a threshold (8 for MBM and 5 for the rest following [8]). Each proposed point can only be assigned to one annotation point. Those annotation points without a corresponding proposal will be counted as False Negatives (FN). Proposals that are not True Positives are labeled as False Positives (FP).

4.3. Results and Analysis

Figure 2 shows results for qualitative evaluation of our method on localization performance, highlighting that our method can distinguish strongly overlapping objects while also eliminating regression noise in those areas where the CCL-T method fails. Table 1 lists quantitative results. All methods present relatively high precision regarding localization among different datasets. But the proposed method generally has lower mean absolute counting error and higher recall and $F_1$ score, especially in those datasets with severe object overlaps (VGG, Ellipse). This indicates that our method performs better in counting and produces fewer false positives of object localization in areas of clustered objects.

5. CONCLUSION

Crowd object counting and localization are essential for quantitative analysis in large-scale microscopy applications. The consistency between counting and localization is crucial for manual correction in applications without ground truth. In this work, we propose a counting method applied to deep network predicted density maps leveraging the property of multivariate normal distributions. Validation and comparison to previously proposed methods demonstrate the improvement in localization - an aspect that may also be important for subsequent semantic segmentation, and in 3D data, where there is a rapidly growing interest in automated analysis. Furthermore, our comparisons also reveal an improvement in object counting results. The proposed processing is generic given crowd object counting applications based on density maps. Future work will tackle the still-existing problem of incorrect detection produced by the regression model.
6. REFERENCES

[1] Yue Guo, Jason Stein, Guorong Wu, and Ashok Krishnamurthy, “Sau-net: A universal deep network for cell counting,” in Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics, 2019, pp. 299–306.

[2] Victor Lempitsky and Andrew Zisserman, “Learning to count objects in images,” Advances in neural information processing systems, vol. 23, 2010.

[3] Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong, “Bayesian loss for crowd count estimation with point supervision,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6142–6151.

[4] Joseph Paul Cohen, Genevieve Boucher, Craig A Glastonbury, Henry Z Lo, and Yoshua Bengio, “Countception: Counting by fully convolutional redundant counting,” in Proceedings of the IEEE International conference on computer vision workshops, 2017, pp. 18–26.

[5] Jia Wan, Ziquan Liu, and Antoni B Chan, “A generalized loss function for crowd counting and localization,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1974–1983.

[6] Weidi Xie, J Alison Noble, and Andrew Zisserman, “Microscopy cell counting and detection with fully convolutional regression networks,” Computer methods in biomechanics and biomedical engineering: Imaging & Visualization, vol. 6, no. 3, pp. 283–292, 2018.

[7] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma, “Single-image crowd counting via multi-column convolutional neural network,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 589–597, 2016.

[8] Yujie Chen, Dingkang Liang, Xiang Bai, Yongchao Xu, and Xin Yang, “Cell localization and counting using direction field map,” IEEE Journal of Biomedical and Health Informatics, 2021.

[9] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Sneed, Ian A Cree, and Nasir M Rajpoot, “Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1196–1206, 2016.

[10] Yuanpu Xie, Fuyong Xing, Xiangfei Kong, Hai Su, and Lin Yang, “Beyond classification: structured regression for robust cell detection using convolutional neural network,” in International conference on medical image computing and computer-assisted intervention. Springer, 2015, pp. 358–365.

[11] Yuhong Li, Xiaofan Zhang, and Deming Chen, “Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1091–1100, 2018.

[12] Michaël Bensimhoun, “N-dimensional cumulative function, and other useful facts about gaussians and normal densities,” Jerusalem, Israel, Tech. Rep, pp. 1–8, 2009.

[13] Rui Wang, Ran Huang, Yuan Gao, and Haiyan Sun, “Sampling method for gmm curve fitting in multi object tracking,” in 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2017, pp. 1–5.

[14] Antti Lehmusto, Pekka Ruusuvuori, Jyrki Selinummi, Heikki Huttunen, and Olli Yli-Harja, “Computational framework for simulating fluorescence microscope images with cell populations,” IEEE transactions on medical imaging, vol. 26, no. 7, pp. 1010–1016, 2007.

[15] Philipp Kainz, Martin Urschler, Samuel Schuster, Paul Wohlhart, and Vincent Lepetit, “You should use regression to detect cells,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2015, pp. 276–283.

[16] John Lonsdale, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, Gary Walters, Fernando Garcia, Nancy Young, et al., “The genotype-tissue expression (gtex) project,” Nature genetics, vol. 45, no. 6, pp. 580–585, 2013.

[17] David Svoboda, Ondřej Homola, and Stanislav Stejskal, “Generation of 3d digital phantoms of colon tissue,” in International Conference Image Analysis and Recognition. Springer, 2011, pp. 31–39.

[18] Katharina Bermond, Christina Wobbe, Ioana-Sandra Tarau, et al., “ Autofluorescent granules of the human retinal pigment epithelium: Phenotypes, intracellular distribution, and age-related topography,” Investigative Ophthalmology & Visual Science, 2020.